UNIVERSIDAD DE GRANADA

Programa de Doctorado en Biomedicina

Centro Pfizer – Universidad de Granada – Junta de Andalucía de Genómica e Investigación Oncológica (GENYO)
Departamento de Bioquímica y Biología Molecular II

Doctoral Thesis:

Regulation of Transposable Elements by microRNAs in Cancer

PhD Candidate:
Pablo Tristán Ramos
Thesis supervisors:
Dr. Sara Rodríguez Heras
Dr. Jose Luis García Pérez

GENYO CENTRO PFIZER-UNIVERSIDAD DE GRANADA-JUNTA DE ANDALUCÍA DE GENÓMICA E INVESTIGACIÓN ONCOLÓGICA

Editor: Universidad de Granada. Tesis Doctorales
Autor: Pablo Tristán Ramos
ISBN: 978-84-1306-671-4
URI: http://hdl.handle.net/10481/63951

A mi madre,

A mi padre,

A Elvira,

"Those times when you get up early and you work hard. Those times when you stay up late and you work hard. Those times when you don't feel like working, you are too tired, you don't want to push yourself but you do it anyway.

That is actually the dream.

It's not the destination, it's the journey"

Kobe Bryant
(1978-2020)

"You have no responsibility to live up to what other people think you ought to accomplish. I have no responsibility to be like they expect me to be. It's their mistake, not my failing."

Richard P. Feynman
(1918-1988)

Para no poner la misma palabra docenas de veces, lo haré sólo una y bien grande:

G R A C I A S…

…a los y las que empezaron a allanar el camino casi un siglo atrás. Los que están, dos Antonios, y los y las que ya no están, Josefa, Emilio, Pilar y Adela. A mi segunda madre/tercera abuela Encarna aka 'La Nena'.

…al José Luis García Pérez, Cholo, por la oportunidad, la acogida, el entusiasmo contagioso, las discusiones en la oficina de verano, la confianza. Por ser un científico brillante de quien he aprendido muchísimo.

…a mis compañeras y compañeros del labo 8. Parafraseando a David Cano: "no existe tanta bondad por metro cuadrado en otro lugar del mundo". Por hacerme madurar y crecer científica y personalmente. A aquellos con los que he coincidido estos años, de quienes he aprendido y a quienes he intentado enseñar: Andrés 'Parce', Valentín, Eva, Suyapa, Martín, Raquel, César, Paula, Miguel 'nano', Jose, Pablo, Vanessa, Belén 'acha', Irene 'teta', y el siempre presente Thomas. A quienes siguen en la lucha: Marta, Alex, María, Guillermo, las tres Anas de Córdoba: 'peces', 'ea', y la prometedora 'illa'. Si escribiese unas palabras de cada uno/a tendría otra tesis, pero hay algunas personas que no puedo dejar pasar. A Laura, el pilar que sostiene el laboratorio y quien me guio en los primeros pasos con las pipetas. A Kiko por su autenticidad, por su ayuda, por compartir su vasta y aun así creciente sabiduría, y por la oportunidad de escribir un libro juntos. A Esther y a Jennifer por su esfuerzo, su energía y su alegría contagiosa: sin ellas, esta tesis sería sin duda peor. A David, al Doctor Fruit, por ser un ejemplo de pensamiento crítico, por dejarme aprender de él cada mediodía, un científico como un templo que en cualquier otro lugar civilizado estaría dedicándose a esto por méritos propios. A Meriem, el corazón más grande del grupo, una persona genial y polivalente: hace ciencia, pinta, organiza tesis, ¡y todo bien!

… to Sara Macias, a smart, nice, and tenacious scientist, for hosting me and making me feel part of the team from the beginning. To the people in the Macias lab (Lisanne, Samir, Jeroen) and in the IIIR at the University of Edinburgh: Kat, Sarah, Matilda, Yvonne, Kyriaki, Ruby, Patri 'tuice', Sujai, Elaine… for making my stay one of the greatest experiences in my life, full of curious, intelligent, hard-working and beautiful people (and beer, to be fair). Special mention to my pals Samir, for the help and wisdom, and Sarah, for the coffees and the mutual support when CLASH was not going smoothly (i.e. frequently). To Kiko (again) and Marie Jeanne for sharing their time, for the beers, the places they discovered to me, the climbing lessons, the housing, and for being an essential part of my time in Edinburgh, for which I will always be in debt.

… a la buena gente de Genyo, que hacen que el ambiente sea agradable y que venir a trabajar cueste mucho menos: Antonio, Silvia, Amador, Helena, Carlos, Alba(s), Verónica, Luis Javier, Jose, Paola, María, Víctor, Rosi, Pili, Lourdes... Imposible nombrar a todos. Mención especial a mis vecinos y vecinas de planta: María 'cousin', Paulina, Araceli, Juane, Noelia, Pedro, Iris, Marina... Y, cómo no, Fernando, Carmen, Miguel Ángel, Ana y Jorge. Y la exiliada Lucía.

… a toda la gente que intenta acercar la ciencia a la sociedad de forma honesta y clara. A Susana y Carlos por animarme a participar en el 3 Minutes Thesis y en FameLab. Al equipo que organizamos Pint of Science por la energía y el entusiasmo.

…a mis amigos y amigas de casi toda la vida, pues algunos nos conocemos desde hace más de 20 años (¡!). A los 'Pollicas' y las 'Chumis': Pablo, Silvia, los Migues, Maria Jesús, JuanPe, Carmen, Alfonso, Marisol, Sara, Rafa, Irene... Ojalá que pasen otros 20, y luego otros 20 encontrando momentos que compartir, aunque nuestras circunstancias sigan cambiando. A Alex, por una amistad que ha sobrevivido primero a 6.000, luego a 500, y ahora a casi 12.000 kilómetros de distancia. Por seguir creciendo juntos y compartiendo nuestro aprendizaje. Por

muchos más cubos de tercios filosofando, y muchos más 1-on-1 en el Sierra o donde sea. A mi hermano postizo Óscar, por todos los momentos compartidos y los que quedan (si vuelve de las garras de los Leprechauns).

…a mis amigas y amigos bioquímicos, con los que compartí cuatro intensos y memorables años, y con los que sigo (y espero seguir) disfrutando grandes momentos. A los desperdigados Sandra, Rafa, Jessica, Melanie, y al auténtico Equipo de Gobierno: Paula, Carlos, Carmen y mi hermano de otra madre Puerma.

…al deporte en general, y al baloncesto en particular, por haberme enseñado el valor del trabajo en equipo, de competir de forma honesta y saber ganar y perder, y de ser paciente. A la música, eterna acompañante inmaterial.

…a mi familia, por el continuo apoyo y cariño, tíos/as y primos/as a quienes espero poder devolver al menos algo de lo recibido.

…a mi hermano, Jorge, por haberme acompañado toda la vida, por ser ejemplo para mí en algunas cosas y porque intentar ser ejemplo para él en otras me ha hecho crecer, pese a mi poco éxito en la tarea.

…a Sara, mi mentora, directora, maestra y espero que, a partir de ahora, 'pal'. La combinación más increíble que he conocido de inteligencia, humildad y resiliencia, sin perder la autenticidad. Una de las personas más trascendentes para mí. Nunca podré agradecerle todo lo que he aprendido, no sólo en lo científico.

…a Elvira, por todo. Por compartir momentos geniales, desde aquella primera noche en el Albaicín hasta el paseo por la playa desierta de la isla más recóndita de Tailandia, o el acantilado más remoto de las Highlands. Por compartir, también, los momentos difíciles. Durante esta tesis hemos resistido separados físicamente la propia tesis, un MIR, una estancia, una residencia y, para rematar, un confinamiento (!). Ahora se abre una nueva etapa donde, por fin, borraremos la distancia. La complicidad es tanta… Ojalá que nuestras vibraciones se complementen durante mucho tiempo.

…por último, a mi madre y a mi padre y viceversa, María A. y Emilio, el principio de todo. Por enseñarme a apreciar los libros, la música, los viajes, el deporte y a las personas; por animarme a pensar por mí mismo, a ser crítico, a dar esa media vuelta a las cosas que a mí me gustaría transmitir en un futuro. Por transmitirme el valor del esfuerzo, de que la aptitud sin actitud es inútil. Por ser siempre capaces de dejar a un lado sus diferencias para buscar lo mejor para mi hermano y para mí. Por el apoyo incondicional. Como dijo Jesús Vidal cuando recogió su Goya: "A mí me gustaría tener un hijo como yo porque tendría unos padres como vosotros".

El doctorando / *The doctoral candidate* **Pablo Tristán Ramos** y los directores de la Tesis / *and the Thesis supervisors* **Dra. Sara Rodríguez Heras**, **Dr. Jose Luis García Pérez**

Garantizamos, al firmar esta Tesis Doctoral, que el trabajo ha sido realizado por el doctorando bajo la dirección de los directores de la Tesis y hasta donde nuestro conocimiento alcanza, en la realización del trabajo se han respetado los derechos de otros autores a ser citados cuando se han utilizado sus resultados o publicaciones

/

Guarantee, by signing this Doctoral Thesis, that the work has been done by the doctoral candidate under the direction of the Thesis supervisors and, as far as our knowledge reaches, in the performance of the work the rights of other authors to be cited (when their results or publications have been used) have been respected.

Lugar y fecha / *Place and date*: Granada, 2 de octubre de 2020

Directores de la tesis / *Thesis supervisors* Doctorando / *Doctoral candidate*

Dra. Sara Rodríguez Heras Pablo Tristán Ramos

Dr. Jose Luis García Pérez

Quality criteria to aim for the 'International PhD' degree from the University of Granada:

Internship in a foreign research centre.

March 2019 – June 2019. Research stay at the Institute for Immunity and Infection Research of the University of Edinburgh (Edinburgh, UK), supervised by Dr. Sara Macias, group leader, Sir Henry Dale Fellow and Chancellor's Fellow. The stay was funded by a fellowship from the Foundation for Basic Research in Medicine (Boehringer Ingelheim Fonds).

Accepted or published article in a relevant journal of the Thesis' scope, including part of the Thesis results.

- sRNA/L1-retrotransposition: using siRNAs and miRNAs to expand the applications of the cell culture-based LINE-1 retrotransposition assay. **Tristan-Ramos P**, Morell S, Sanchez L, Toledo B, Garcia-Perez JL, and Heras SR. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 2020 375: 20190346. doi: 10.1098/rstb.2019.0346
- The tumor suppressor microRNA let-7 inhibits human LINE-1 retrotransposition. **Tristan-Ramos P**, Rubio-Roldan A, Peris G, Sanchez L, Amador-Cubero S, Viollet S, Cristofari G, Heras SR. *Nature Communications*. In press

Language used for Thesis writing and defence

Following the requirements from the University of Granada to obtain an 'International PhD' degree, this Thesis has been written and will be defended in English. The sections Abstract and Conclusions have also been written and will be defended in Spanish.

Regulation of Transposable Elements by microRNAs in Cancer

INDEX

RESUMEN

Los elementos móviles o transponibles (TEs) constituyen casi la mitad del genoma humano, y siguen afectando su estructura y función. Los únicos TEs activos y autónomos en la actualidad se denominan 'Long INterspersed Elements class 1' (LINE-1s or L1s), cuyas 500,000 copias forman en torno al 17% de nuestro genoma. Sólo 80-100 de estas copias siguen siendo activas, y se denominan L1s retrocompetentes o RC-L1s. Los RC-L1s tienen una longitud de 6 kb y codifican dos proteínas imprescindibles para su movilización: L1-ORF1p, una proteína de unión a ARN con actividad chaperona, y L1-ORF2p, con actividades endonucleasa y reverso transcriptasa. Los L1s se movilizan por un mecanismo de copia y pega, generando nuevas inserciones mediante la transcripción inversa de un RNA intermediario. Por el momento, se han descrito más de un centenar de inserciones causantes de enfermedad.

En la mayoría de células somáticas humanas, la expresión de L1 está silenciada por diversos mecanismos. Sin embargo, estos elementos se expresan y movilizan en numerosos tipos de cáncer, particularmente epiteliales como los de pulmón o colon, donde pueden aumentar la tumorogénesis o incluso iniciar el tumor. Aunque se considera que una de las principales causas de la reactivación de L1 es la hipometilación de RC-L1s, siguen sin conocerse otros mecanismos post-transcripcionales de control de la movilidad de L1 que estén desregulados en cáncer.

Por otro lado, los microARNs (miARNs) son pequeños ARNs endógenos que controlan la expresión génica en eucariotas y que están involucrados en básicamente todos los procesos fisio y patológicos humanos, así como en el desarrollo. En su biogénesis resaltan dos pasos de procesamiento: uno en el núcleo mediado por el complejo 'Microprocessor' (formado por DGCR8 y Drosha), y otro en el citoplasma por Dicer. A continuación, se cargan en las proteínas Argonaute (AGO) para formar el complejo de silenciamiento inducido por RNA (RISC). Finalmente, los miARNs guían al RISC hacia su RNA mensajero

(mRNA) diana, con el que interacciona por complementariedad parcial y causa una inhibición de la traducción y/o degradación del mRNA.

Los miARNs supresores de tumores como los de la familia let-7 regulan la expresión de oncogenes y están deplecionados en cáncer, donde los L1s se movilizan. Sin embargo, se desconoce si esta depleción de miARNs supresores de tumores provoca o influencia la acumulación de inserciones de L1 en cáncer.

Comenzamos este estudio analizando datos de secuenciación de genoma completo y de miARNs en muestras de tumores de pulmón. La baja expresión de miembros de la familia de let-7 correlaciona con un aumento en las retrotransposición somática de L1. De hecho, la sobreexpresión o depleción de let-7 disminuye o aumenta, respectivamente, la retrotransposición de L1 en células en cultivos, incluyendo células de cáncer de pulmón. A nivel mecanístico, usando multitud de técnicas moleculares y bioquímicas encontramos que let-7 se une a un sitio no canónico en el ARNm de L1 y dificulta la traducción de L1-ORF2p, que es imprescindible para la retrotransposición, reduciendo su movilización.

Además, hemos generado un modelo celular de cáncer de pulmón donde podemos restaurar la expresión de let-7 de forma inducible. Este modelo permitirá el estudio de la movilización endógena de L1 al restaurar los niveles de let-7. Por último, hemos realizado la técnica de 'Cross-Linking and Sequencing of Hybrids' (CLASH) para identificar microARNs asociados a AGO2 y unidos al ARNm de L1 *in vivo*. Aunque esta técnica podría permitir la identificación de miARNs que regulen la movilidad de L1, nuestros resultados sugieren que aún requiere considerable optimización en vista del bajo número de interacciones capturadas.

En conjunto, esta Tesis contribuye a los campos de los elementos transponibles y de miARNs, así como a la biología del cáncer. De hecho, hemos descubierto una nueva función de la familia de miARNs supresores de tumores let-7: mantener la integridad del genoma somático controlando la retrotransposición de L1.

ABSTRACT

Transposable elements (TEs) comprise nearly half of the human genome, and their ongoing activity continues to impact its structure and function. The only TEs that remain autonomously active are Long INterspersed Elements class 1 (LINE-1s or L1s) retrotransposons, whose 500,000 copies account for ~17% of our genome. Only 80-100 L1s retain the ability to mobilize, and are therefore called Retrotransposition-Competent L1s (RC-L1s). RC-L1s are 6kb long and encode two proteins that are required for their mobilization: L1-ORF1p, an RNA binding protein with chaperone activity, and L1-ORF2p, with endonuclease and reverse transcriptase activities. RC-L1s mobilize via a copy-and-paste mechanism, generating new insertions by reverse transcription of an intermediate RNA. At present, over one hundred disease-causing insertions have been reported.

In most human somatic cells, L1 expression is silenced by a variety of mechanisms. However, L1s are expressed and mobilized in a wide range of human cancers, particularly in epithelial cancers such as lung and colorectal where it can drive tumorigenesis and even initiate the tumor. Although hypo-methylation of specific RC-L1s is considered one of the main causes of L1 reactivation, other post-transcriptional mechanisms that repress L1 mobilization in normal cells and are dysregulated in cancer remain to be discovered.

On the other hand, microRNAs (miRNAs) are endogenous ~22nt RNAs that function as crucial regulators of gene expression in eukaryotes and are involved in, essentially, all human developmental, physiological and pathological processes. Their biogenesis involves two major processing steps: one in the nucleus by the Microprocessor complex (formed by Drosha and DGCR8), and other in the cytoplasm by Dicer. Next, they are loaded into the Argonaute (AGO) proteins to form the RNA-Induced Silencing Complex (RISC). Lastly, miRNAs guide the RISC to a partially complementary pairing with target mRNAs, resulting in translation inhibition and/or mRNA degradation.

Among miRNAs, tumor suppressor miRNAs such as the let-7 family regulate the expression of oncogenes and are frequently downregulated in cancer, where frequent L1 retrotransposition has been reported as stated above. How-ever, whether this downregulation of tumor suppressor miRNAs causes or influences the accumulation of L1 insertions in cancer has remained largely unexplored.

We started by analysing whole genome and miRNA-sequencing data from lung tumor samples. Interestingly, downregulation of several members of the let-7 family correlates with increased somatic L1 retrotransposition. Indeed, let-7 overexpression or depletion decreases and increases, respectively, engineered L1 retrotransposition in a panel of cultured human cells, including lung cancer cells. Mechanistically, using a variety of molecular and biochemical assays we show that let-7 binds to a non-canonical site in the L1 mRNA and impairs the translation of L1-ORF2p, which is indispensable for retrotransposition, reducing its mobilization.

Additionally, we generated a lung cancer cell model in which the previously downregulated expression of let-7 can be inducibly restored. This model will enable to study endogenous L1 mobilization upon restoration of let-7 levels. Lastly, we employed Cross-Linking and Sequencing of Hybrids (CLASH) to identify miRNAs associated to AGO2 and bound to L1 mRNA *in vivo*. Although this technique could potentially allow the unbiased identification of additional miRNAs that regulate L1 mobilization, our results suggest that it still requires significant optimization as the number of captured interactions is extremely low.

Altogether, this Thesis represents a contribution to the fields of mobile genetic elements and miRNAs, as well as to cancer biology. In fact, we uncover a new role for the tumor suppressor miRNA let-7 family: maintaining somatic genome integrity by restricting L1 retrotransposition.

INTRODUCTION

I.1. Repeated sequences: the rise of 'junk' DNA

The presence of repetitive DNA in genomes was first suggested in the 60's, when conducting DNA reassociation experiments. First, it was shown that 10% of the mouse genome was comprised by highly repetitive DNA with low %GC known as satellite DNA (Waring and Britten, 1966). The presence of satellite DNA was later confirmed in many other eukaryotic genomes, from invertebrates to humans, and expanded with the discovery of various DNA fractions with different degrees of repetitiveness (Britten and Kohne, 1968). These studies partly solved an intriguing, long standing observation in eukaryotes which is known as the 'C-value* paradox': that genome size and organismal complexity are not correlated (reviewed in (Thomas, 1971)).

Almost forty years later, the sequencing of the human genome unveiled that, surprisingly, coding sequences comprised only ~3% of our genome, whereas repeated sequences accounted for at least 50% (Lander *et al.*, 2001). Briefly, and broadly, we can differentiate five classes of repeats (Krebs, Goldstein and Kilpatrick, 2014):

- Transposable Element-derived repeats.
- Processed pseudogenes: partially or completely inactive retrotransposed copies of cellular genes (protein coding or structural RNAs).
- Simple sequences repeats: repetitions or relatively short k-mers, like $(A)_n$ or $(CGG)_n$.
- Segmental duplications: large blocks of DNA, ranging from 10 to 300 kb, that have been duplicated from one region to another in the genome.
- Tandemly repeated sequences, such as centromeres, telomeres and ribosomal RNA clusters.

* C-value refers to the amount of DNA per haploid set of chromosomes.

Repeated sequences, repetitive DNA or simply 'repeats' were initially (and sometimes still are) referred to as 'junk' DNA, misleading to the conclusion that they lacked interest. Nothing could be further from the truth. We now know that repeats are key for chromosome structure and dynamics (Krebs, Goldstein and Kilpatrick, 2014), and that transposons have been fundamental in adaptive evolution (Cosby, Chang and Feschotte, 2019; Schrader and Schmitz, 2019). Unsurprisingly, alterations in repeated sequences can cause a range of mendelian disorders such as Huntington disease (Hannan, 2018), and transposon activity is associated to human disorders (Burns, 2020). Furthermore, repeats can be considered as a palaeontological record with a tremendous potential to understand genome evolution (Lander *et al.*, 2001).

This Thesis is focused on one type of human repeats: retrotransposons. Nevertheless, repeats are also important in all eukaryotes and also in prokaryotes. A notable example of prokaryotic repeats with massive impact in biology and biotechnology are Clustered Regularly Interspaced Palindromic Repeats, widely known as CRISPR (Doudna and Charpentier, 2014).

I.2. From repetitive DNA to transposable elements

Amongst the five classes of repeated sequences previously listed, the most prevalent by far is the first, as transposable element-derived repeats account for at least 45% and up to two-thirds of the human genome (Lander *et al.*, 2001; de Koning *et al.*, 2011). Besides, some of the remaining 'unique' DNA may as well derive from ancient transposable elements whose sequences have diverged too much to be identified as such (Lander *et al.*, 2001; de Koning *et al.*, 2011). This prevalence is found extensively across eukaryotic genomes: 37% of mouse DNA (Mouse Genome Sequencing Consortium *et al.*, 2002), 39% of zebrafish DNA (Howe *et al.*, 2013) and more than 80% of maize DNA (Schnable *et al.*, 2009) are derived from transposable elements.

Transposable Elements (TEs) can be defined as DNA sequences that are able to mobilize from one place to another within genomes. The first evidence for the existence of TEs was provided in the 50's by Barbara McClintock. She observed that variegation* in maize kernels was controlled by the insertion of a TE (*Ds*, Dissociation), mediated by another (*Ac*, Activator), in a specific locus of the maize genome (McClintock, 1950). This was not only the first evidence of TE's existence, but also of the potential phenotypic impact of TE mobilization. In the next decades, these elements were further characterized (Fedoroff, Wessler and Shure, 1983) and new examples of active TEs were found in other model organisms, such as *copia* in *Drosophila* (Potter *et al.*, 1979) and Ty1 in yeast (Farabaugh and Fink, 1980). Of note, despite the initial hostility to the concept of transposition, Barbara McClintock was awarded an unshared Nobel Prize in Physiology or Medicine in 1983 "for her discovery of mobile genetic elements", 33 years after her discovery.

At first, TEs were classified as 'parasitic' or 'selfish' DNAs, with the only function of self-propagating (Doolittle and Sapienza, 1980; Orgel and Crick, 1980). Indeed, examples of horizontal transfer leading to TE colonization of different genomes have been described in plants and some animals (Gilbert and Feschotte, 2018). However, nowadays it is clear that TEs interact and co-evolve with their host genome to mitigate the cost of their propagation, which in some cases has led to the repurposing of TEs for novel cellular functions (thoroughly reviewed in (Cosby, Chang and Feschotte, 2019).

* Irregular colouring, pattern of different colours

I.3. Transposable elements in the human genome

TEs can be divided in two groups based on their mobilization mechanism (Finnegan, 1989; Beck *et al.*, 2011). On one hand, DNA transposons (or class II elements) mobilize through a 'cut-and-paste' mechanism: the TE is excised from a locus and inserted in a different genomic location*. On the other hand, retrotransposons (or class I elements) mobilize through a 'copy-and-paste' mechanism that involves reverse transcription of an intermediate RNA into a new genomic location†. Retrotransposons can be further divided into autonomous or non-autonomous, depending on whether they encode the enzymatic machinery required for their mobilization or not, respectively (Finnegan, 1989; Beck *et al.*, 2011). Within autonomous TEs, an additional subdivision can be made based on the presence or absence of Long Terminal Repeats (LTR) flanking the element. An overview of the main TE types, features and examples in the human genome is provided in **Table 1**, and further details are provided in the next sections.

I.3.A. DNA Transposons

The 300,000 copies of DNA transposons comprise around 3% of the human genome (Lander *et al.*, 2001) and can be grouped in 4 superfamilies, the most abundant being hAt and Tc1/*mariner* (Pace and Feschotte, 2007). Structurally, DNA transposons are 2-3 kb long and encode a transposase gene flanked by two terminal inverted repeat (TIR) sequences **(Table 1)**. Upon transcription and translation, the transposase returns to the nucleus where it recognizes the TIRs, excises the transposon, and inserts it in a different location (reviewed in (Hickman and Dyda, 2015). This usually generates direct repeats (DR) flanking the insertions **(Table 1)**: these are not part of the transposon but a direct

* The *Ac* and *Ds* elements discovered by Barbara McClintock are DNA transposons.
† LINE-1 elements, the focus of this thesis, are retrotransposons.

consequence of the staggered* cut of the transposase. Although they were intensely active during primate radiation, there is no evidence that any DNA transposon has been active in humans in the past 37 million years (Pace and Feschotte, 2007). In general, DNA transposons are inactive in most mammals, with the notable exception of bats (Ray *et al.*, 2007, 2008)

Although not currently active in our genome, DNA transposons have had a great impact in human biology. Notably, the V(D)J-recombination proteins RAG1 and RAG2 derive from the domestication of an ancient DNA transposon from the Transib family, a probably fortuitous event that has been crucial in vertebrate evolution (Agrawal, Eastman and Schatz, 1998; Hiom, Melek and Gellert, 1998; Kapitonov and Jurka, 2005; Huang *et al.*, 2016).

I.3.B. Retrotransposons

In most mammals including humans, retrotransposons are, due to their ongoing replicative mobilization, much more abundant than DNA transposons and comprise at least a third of the human genome (Lander *et al.*, 2001). Their mobilization mechanism involves reverse transcription and integration of an intermediate RNA to generate a new retrotransposon copy. Autonomous retrotransposons encode the proteins required for their mobilization, whereas nonautonomous retrotransposons hijack the proteins of the former to mobilize (Beck *et al.*, 2011). A different division can be made based on the presence or absence of LTRs flanking the retroelement (Beck *et al.*, 2011). All currently active retrotransposons belong to the non-LTR class **(Table 1)**.

* The staggered cut leaves sticky ends in the DNA that are filled by a cellular DNA polymerase.

I.3.B.i. LTR retrotransposons

In the human genome, HERVs (Human Endogenous Retroviruses) are the most representative family of LTR retrotransposons, and account for 8% of our genome (Lander *et al.*, 2001). However, most (~85%) of the HERV-derived copies in our genome are isolated LTRs (or 'solo-LTRs'), that likely lost the internal sequence due to homologous recombination between the flanking LTRs (Lander *et al.*, 2001; Mager and Stoye, 2015). Complete HERVs have a length of 6-9 kb, and are structurally similar to retroviruses: Gag, Pro, Pol and Env* genes flanked by LTRs that contain all the required transcriptional and regulatory elements, including an RNA pol II promoter **(Table 1)**. Their mobilization occurs through a complex, multi-step mechanism involving a tRNA to prime reverse transcription in virus-like cytoplasmic particles (Wilhelm and Wilhelm, 2001), generating a dsDNA proviral genome in the cytoplasm. In the nucleus, these dsDNAs are inserted on a new genomic location in a single cut-and-paste reaction mediated by an integrase, generating target site duplications (TSDs) flanking the new insertion **(Table 1)**. To date, no active LTR retrotransposon has been described in humans (Bannert and Kurth, 2006), although polymorphic copies of HERV-K† elements have been reported, with at least one intact provirus copy (Wildschutte *et al.*, 2016), suggesting a recent peak of activity in primates.

(in the next page) **Table 1. Types of transposable elements in the human genome.** Classification, structure of a representative element, length, copies and percentage of each element in the reference human genome (HGR), and active subfamilies are shown. Black arrows represent promoter activity. DR, Direct Repeat; ITR, Inverted Terminal Repeat; TSD, Target Site Duplication; Gag, group-specific antigen; Pro, protease; Pol, polymerase; Env, envelope protein (dysfunctional); UTR, Untranslated Region; CC, Coiled Coil; RRM, RNA Recognition Motif; CTD, Carboxyl-Terminal Domain; EN, Endonuclease; RT, Reverse Transcriptase; C, Cysteine-rich domain; A/B, sequences of the RNA polymerase III promoter; A-rich, adenosine-rich segment separating the 7SL monomers; SVA, SINE-R/VNTR/Alu; VNTR, Variable Number of Tandem Repeats; SINE-R, domain derived from a HERV-K. SINE-R sequence sharing homology with HERV-K10, (envelope [ENV] and LTR); cleavage polyadenylation specific factor (CPSF) binding site (A)n, poly(A) tail. Adapted from (Beck et al., 2011).

* Gag, group-specific antigen; Pro, protease; Pol, polymerase; Env, envelope protein.
† K (or W, L…) denotes the t-RNA used by a specific element to prime its reverse transcription.

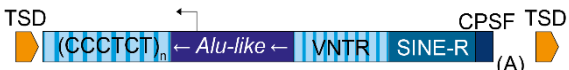
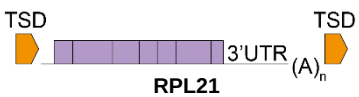
Type of transposable element	Structure (not scaled)	Length	Copy number (% HGR)
DNA Transposons			
Transposons	 **Mariner**	2-3 kb	~300,000 (~3%)
Retrotransposons			
Autonomous			
LTR retrotransposons	 **HERV-K**	6-10 kb	~450,000 (~8%)
Non-LTR retrotransposons	 **LINE-1**	~6 kb	~850,000 (~21%)
Nonautonomous			
	 **Alu**	~300 bp	>1,000,000 (~10%)
SINEs	 **SVA**	0.7- 4 kb	~2,700 (<0.2%)
Processed pseudogenes	 **RPL21**	Variable	~11,000 (<1%)

Table 1. Types of transposable elements in the human genome. *(legend in the previous page)*

Increasing evidence suggest that HERVs, like other currently inactive TEs, have impacted our genome in many ways. The first discovered case involved *syncytin*, a protein with an important role in placental development, which derives from the envelope gene of a retrovirus from the HERV-W family (Mi *et al.*, 2000). A more recent report described how several copies of the MER41 HERV have been co-opted as cis-regulatory elements to control the expression of interferon-regulated genes (Chuong, Elde and Feschotte, 2016). Lastly, some HERVs drive stem-cell specific expression of regulatory long noncoding RNAs (Kelley and Rinn, 2012), some of which are involved in pluripotency maintenance (Wang *et al.*, 2014).

I.3.B.ii. Non-LTR Retrotransposons

All the currently active TEs in the human genome belong to this class. Although their mobilization also occurs through an intermediate mRNA, these elements differ from LTR retrotransposon in their mobilization mechanism as they move by Target Primed Reverse Transcription, or TPRT[*]: a nick is generated in the genomic DNA exposing a 3'-OH that is used as a primer to reverse transcribe the intermediate RNA; a similar process is thought to occur for the other strand, leading to the generation of a new insertion (Cost *et al.*, 2002; Beck *et al.*, 2011). Long Interspersed Elements (LINEs) are autonomous (i.e. encode the proteins required for their mobilization), while Short Interspersed Elements (SINEs) require the proteins encoded by LINE elements to mobilize (Beck *et al.*, 2011; Richardson *et al.*, 2015).

[*] Interestingly, TPRT was originally described in another related non-LTR retrotransposon, R2Bm, from the genome of the silkworm *Bombyx mori* (Luan et al., 1993).

The impact of non-LTR retrotransposons in genome structure and function is unmatched, and unsurprisingly they have been related to genome evolution and human disease. This will be further detailed in section I.5.

I.3.B.ii.a. LINEs

Three distant families of Long Interspersed Elements (LINEs) account for 21% of the human genome: LINE-1, LINE-2 and LINE-3 (Lander *et al.*, 2001). However, only elements belonging to the LINE-1 (L1) family are still active nowadays (Khan, Smit and Boissinot, 2006; Beck *et al.*, 2011). LINEs are about 6 kb long and contain a 5' Untranslated Region (5'UTR) with internal RNA polymerase II promoter activity, two non-overlapping open reading frames (ORF1 and ORF2), and a short 3'UTR containing a poly(A) signal **(Table 1)**. Upon translation, the RNA is used as a template by the reverse transcriptase activity encoded in ORF2 to generate a new insertion, that will be flanked by TSDs (Beck *et al.*, 2011). Given the fact that reverse transcription usually fails to reach the 5' end of the elements, the vast majority of LINE copies in the genome are truncated and therefore inactive (Richardson *et al.*, 2015). Since LINE-1 elements are the main subject of study in this Thesis, a dedicated section will follow in I.4.

I.3.B.ii.b. SINEs

Short Interspersed Elements (SINEs) are nonautonomous retrotransposons that comprise around 13% of our genome. Three markedly different SINE families exist in our genome. The most ancient is the inactive Mammalian-wide Interspersed Repeat or MIR family, whose 400,000 copies account for 2% of the human genome (Lander *et al.*, 2001). MIRs are derived from t-RNAs, and amplified actively before mammalian radiation ~130 million years ago (Jurka, Zietkiewicz and Labuda, 1995; Smit and Riggs, 1995). They were adapted to be mobilized by LINE-2 elements. Consequently, when these became functionally

extinct, MIRs ceased to propagate (Lander *et al.*, 2001). Although it has been reported that insertion of a MIR element ~160 million years ago boosted the activity of an enhancer involved in the development of blood stem cells (Smith *et al.*, 2008), the genomic impact of MIRs is largely unknown. The other two SINE families, Alu and SVA (SINE-R/VNTR*/Alu) are still active and will be further detailed.

Alu

Alu† elements are, in terms of copy number, the most successful TEs in our genome: there are over 1 million Alu copies in the human genome resulting from their continued activity over the past ~65 million years (Batzer and Deininger, 2002). However, due to their short length (~300bp), they 'only' account for 10% of our genome (Lander *et al.*, 2001). Alu elements contain two monomers derived from the 7SL RNA gene, which is part of the signal recognition particle, or SRP‡ (Ullu and Tschudi, 1984), separated by an adenosine-rich region, and usually ends in an adenosine-rich tail (Batzer and Deininger, 2002) **(Table 1)**. The left monomer contains A and B boxes, which enable transcription by RNA polymerase III (Chu, Liu and Schmid, 1995) **(Table 1)**. Due to the lack of an RNA pol III termination signal, Alu transcription extends until that signal (usually a stretch of >4 consecutive thymines) is found in the flanking genome (Chu, Liu and Schmid, 1995). However, all Alu elements in the genome have the same length because Alu TPRT initiates at the 3' $(A)_n$ tail **(Table 1)** and because, in stark contrast with LINE-1s, they are rarely truncated.

---

* VNTR: Variable Number of Tandem Repeats

† These elements are termed 'Alu' because they contain a restriction site for the enzyme *AluI* (Houck, Rinehart and Schmid, 1979).

‡ SRP is a ribonucleoprotein involved in targeting specific proteins to the endoplasmic reticulum.

Alu elements rely on L1-encoded proteins to mobilize (Dewannieux, Esnault and Heidmann, 2003), thus they present the same insertional hallmarks such as the TSDs **(Table 1)**. Mechanistically, Alu RNAs associate with protein components of the SRP, stall the ribosome translating L1-ORF2p* from LINE-1, compete with L1 mRNA for L1-ORF2p thanks to its poly(A) tail, and finally hijack L1-ORF2p to generate a new Alu insertion via TPRT (Ahl *et al.*, 2015; Doucet, Wilusz, *et al.*, 2015). In humans, all active Alu retrotransposons belong to the AluY† family (Bennett *et al.*, 2008; Konkel *et al.*, 2015). Alu elements have significantly contributed to genome evolution (Batzer and Deininger, 2002; Cordaux and Batzer, 2009), and more than 60 disease-causing, LINE-1 mediated Alu insertions have been reported so far (Hancks and Kazazian, 2016).

SVA

SINE-R/VNTR/Alu (SVA) are the youngest TEs in humans: they arose in primate genomes ~25 million years ago and the approximately 2,700 SVA copies account for around 0.2% of the human genome (Wang *et al.*, 2005). They are composite retrotransposons that, like Alu, require LINE-1 encoded proteins to mobilize (Hancks *et al.*, 2011; Raiz *et al.*, 2012), and therefore present hallmarks of TPRT like TSDs **(Table 1)**. The presence of a 3' poly(A) tail suggest that SVAs are transcribed by RNA polymerase II, although this has not been confirmed yet and their promoter remains to be characterized (Hancks and Kazazian, 2010). SVAs can range from 700 to 4,000 bp, although a typical SVA is ~2,000 bp and is structurally complex: it contains hexameric repeats (CCCTCT), a variable number of GC-rich tandem repeats (VNTR), a SINE-R sequence sharing homology

* The protein encoded by LINE-1 ORF2.
† Y for 'young' Alu. AluY expansion has occurred over the past 35 million years (Batzer and Deininger, 2002).

with HERV-K10 (env gene and LTR), and cleavage polyadenylation specific factor (CPSF) binding site **(Table 1)**. Because of their composite structure, SVA elements are believed to have been formed from alternative splicing and genomic rearrangement of older repeats (Hancks and Kazazian, 2010). The most polymorphic SVA subfamilies, which are presumed to be active, are SVA-E and SVA-F* (Richardson *et al.*, 2015; Burns, 2020).

The impact of SVAs in human genome evolution is not fully understood yet. However, disease causing SVA insertions have been described in humans (Hancks and Kazazian, 2016; Burns, 2020).

Processed pseudogenes

Apart from Alu and SVA, LINE-1 retrotransposons are also able to reverse-transcribe and integrate polyadenylated messenger RNAs into the genome, giving rise to processed pseudogenes that (i) lack introns, (ii) present a poly(A) in the 3'end and (iii) are flanked by TSDs (Esnault, Maestre and Heidmann, 2000; Wei *et al.*, 2001) **(Table 1)**. There are over 11,000 pseudogenes of variable length in the human genome (Zhang *et al.*, 2003), and the most frequent are derived from highly expressed ribosomal genes such as RPL21† **(Table 1)**. Additionally, other types of cellular RNAs such as U6 spliceosomal RNAs can be mobilized by the L1 machinery (Buzdin *et al.*, 2003; Doucet, Droc, *et al.*, 2015). Indeed, chimeric U6/L1 insertions can be generated by retrotransposition of previously ligated U6 and L1 RNAs (Moldovan *et al.*, 2019).

Retrotransposition of processed mRNAs can result in an evolutionary advantage. A putative example is the owl monkeys' TRIM5/CypA‡ chimeric gene,

* SVA subfamilies are termed A to F based in differences in the VNTR region (Wang *et al.*, 2005).
† 60S Ribosomal Protein L21
‡ Tripartite Motif Containing 5/ Cyclophilin A

which codifies for a TRIM5-CypA fusion protein. This gene resulted from the LINE-1 mediated insertion of CypA mRNA into the TRIM5 gene, and confers them resistance to HIV-1 (Sayah *et al.*, 2004). Alternatively, this chimeric gene could have also been generated by a severely 5' truncated insertion of a 3' transduced L1 RNA. Regardless of the mechanism, this example illustrates how TEs can drive gene innovation. Moreover, processed pseudogenes can act as DNA-mediated regulators, be a source of non-coding regulatory RNAs, and even be translated to small peptides with various functions (recently reviewed in (Cheetham, Faulkner and Dinger, 2020)).

I.4. LINE-1: the only active and autonomous TE in humans

The around 500,000 copies of LINE-1 (or L1) retrotransposons account for ~17% of the human genome (Lander *et al.*, 2001). Importantly, considering that SINEs and processed pseudogenes rely on L1-encoded machinery to retrotranspose, the activity of LINE-1 elements has generated at least a third of our genome (Lander *et al.*, 2001; Beck *et al.*, 2011). However, over 99% of LINE-1 copies in the genome have been rendered inactive due to 5' truncation or, less frequently, point mutations or internal rearrangements (Grimaldi, Skowronski and Singer, 1984; Lander *et al.*, 2001; Szak *et al.*, 2002). Indeed, although an active L1 is ~6kb long, the mean length of L1 elements in the genome is <1 kb (Lander *et al.*, 2001).

I.4.A. Brief overview of LINE-1 families and nomenclature

There are 16 LINE-1 Primate-specific Amplifying (PA) subfamilies in our genome: from the youngest L1PA1 to the ancient L1PA16 (Khan, Smit and Boissinot, 2006). Functional analyses have shown that only some elements from the L1PA1 subfamily, also known as L1Hs*, are still active in our genome

* Hs for Human Specific

(Sassaman *et al.*, 1997; Brouha *et al.*, 2003; Beck *et al.*, 2010). Active L1s are also termed Retrotransposition-Competent LINE-1s, or RC-L1s. The majority of RC-L1s elements belong to a small subset of the L1Hs subfamily named Ta* (Skowronski, Fanning and Singer, 1988; Boissinot, Chevret and Furano, 2000) **(Table 1)**. Functional assays have demonstrated than an average human genome contains ~80-100 RC-L1s, although only a few are responsible for the bulk of retrotransposition and are termed 'hot' L1s (Brouha *et al.*, 2003; Beck *et al.*, 2010). Even though L1 expression is controlled at the promoter level by DNA methylation (see section I.6), activation of individual RC-L1s can be driven by locus- and cell type-specific determinants (Philippe *et al.*, 2016).

I.4.B. Structure of an active LINE-1

Active L1s are 6kb long and contain (i) a 5'UTR, (ii) two non-overlapping ORFs (ORF1 and ORF2) separated by a 63-nucleotide intergenic region which encode two different proteins (termed L1-ORF1p and L1-ORF2p), and (iii) a 3'UTR ending in a poly(A) (Dombroski *et al.*, 1991) **(Table 1)**. As mentioned before, they are usually flanked by TSDs, a hallmark of TPRT-mediated retrotransposition **(Table 1)**.

5'UTR

The 5'UTR is approximately 900 bp long and contains an internal RNA polymerase II promoter which drives L1 expression (Swergold, 1990) **(Table 1)**. It also contains a conserved antisense promoter that can drive transcription of adjacent genes and lncRNAs† (Speek, 2001; Macia *et al.*, 2011; Criscione *et al.*, 2016).

* Ta for transcribed, subset a
† Long non-coding RNAs

Besides, it harbours binding sites for several transcription factors such as SOX* (Tchénio, Casella and Heidmann, 2000), RUNX3† (Yang *et al.*, 2003), or YY1‡, which is responsible for the correct start of transcription at nucleotide +1 (Becker *et al.*, 1993; Athanikar, Badge and Moran, 2004). Recently, a novel ORF was identified in the antisense strand of the 5'UTR of primate L1s, termed ORF0 (Denli *et al.*, 2015) **(Table 1)**. ORF0 encodes a 70 amino acid peptide that seems to enhance L1 mobilization, although its mechanism is still unknown (Denli *et al.*, 2015). Interestingly, ORF0 contains two splice donor sites, and thus can form fusion proteins with proximal exons (Denli *et al.*, 2015).

ORF1

ORF1 encodes a ~40 kDa protein (L1-ORF1p) that is translated from the bicistronic L1 mRNA via Cap-dependent translation (Dmitriev *et al.*, 2007) and is indispensable for L1 retrotransposition (Moran, Holmes and Naas, 1996). Structural and biochemical studies have revealed three domains in L1-ORF1p **(Table 1)**. First, an N-terminal coiled-coil domain (CC) that enables trimerization of L1-ORF1p molecules (Khazina *et al.*, 2011). Next, an RNA recognition motif (RRM) that enables binding of L1-ORF1p to ssRNA§ (Hohjoh, Singer and Nw, 1997). Lastly, a carboxy-terminal domain (CTD) that assists RRM domain in nucleic acid binding (Khazina and Weichenrieder, 2009). Importantly, mutations in any of these domains dramatically affect L1 retrotransposition in cultured cells (Doucet *et al.*, 2010). Nucleic acid chaperone activity has been described for

* SRY-related HMG-box
† Runt-related transcription factor 3
‡ Ying-Yang 1
§ Single-stranded RNA

mouse L1-ORF1p and is assumed for human L1-ORF1p, due to the similarity of residues in their C-terminal halves (Martin and Bushman, 2001; Martin, 2006).

ORF2

ORF2 encodes a ~150 kDa protein (L1-ORF2p) that is translated from the bicistronic L1 RNA via an unconventional termination/reinitiation mechanism (Alisch *et al.*, 2006) and is indispensable for retrotransposition (Moran, Holmes and Naas, 1996). There are three distinct domains in L1-ORF2p **(Table 1)**. First, an APE*-like endonuclease (EN) domain that nicks the DNA where the new insertion will occur (Feng *et al.*, 1996). This is followed by a Z-domain, which contains a functional PCNA†-Interaction Protein (PIP) motif involved in retrotransposition (Taylor *et al.*, 2013). Next, a reverse transcriptase (RT) domain that uses the L1 mRNA as a template to generate a new insertion (Mathias *et al.*, 1991). Lastly, there is a conserved cysteine-rich (C) domain that is essential for L1 retrotransposition although its biochemical role is not fully characterized (Moran, Holmes and Naas, 1996; Doucet *et al.*, 2010). As expected, mutations in any of the domains severely reduce L1 retrotransposition (Moran, Holmes and Naas, 1996; Doucet *et al.*, 2010; Taylor *et al.*, 2013).

3'UTR

The 3'UTR of LINE-1 is approximately 200bp long and contains a conserved guanine-rich (G-rich) polypurine tract (Howell and Usdin, 1997), a canonical poly-adenylation signal and a poly(A) tail. The G-rich tract is not required for engineered retrotransposition in cultured cells (Moran, Holmes and Naas, 1996) but has been suggested to form G-quadruplex structures that might enhance L1

* Apurinic/apyrimidinic endonuclease
† Proliferating cell nuclear antigen

mobilization (Sahakyan *et al.*, 2017). The poly-adenylation signal is relatively weak and frequently bypassed by RNA polymerase II, which continues transcription until a new one is found in the flanking genome. This leads to chimeric L1 transcripts formation, and the generation of 3' transductions upon retrotransposition of those chimeric RNAs (Moran, DeBerardinis and Kazazian, 1999; Pickeral *et al.*, 2000). Notably, genomic copies of L1 are usually flanked by a poly(A) tail **(Table 1)**. Polyadenylation of L1 mRNA is crucial for retrotransposition (Doucet, Wilusz, *et al.*, 2015).

I.4.C. LINE-1 mobilization mechanism

Retrotransposition of LINE-1 occurs via a 'copy-and-paste' mechanism termed Target Primed Reverse Transcription or TPRT (Luan et al., 1993; Cost et al., 2002) **(Figure 1)**. First, a full-length RC-L1 is transcribed by RNA polymerase II from the internal promoter located in the 5'UTR (Swergold, 1990). The bicistronic mRNA is exported to the cytoplasm (Lindtner, Felber and Kjems, 2002), where translation occurs by two different mechanisms: canonical Cap-dependent translation yields many molecules of L1-ORF1p (Dmitriev et al., 2007), and an unconventional termination/reinitiation mechanism yields only a few molecules of L1-ORF2p, perhaps one per mRNA (Alisch et al., 2006). L1-ORF1p trimers and L1-ORF2p (in a ratio of ~30:1 (Taylor *et al.*, 2013, 2018)) associate preferentially with the L1 mRNA from which they were translated, a phenomenon known as cis-preference (Kimberland *et al.*, 1999; Wei *et al.*, 2001; Kulpa and Moran, 2006), to form a ribonucleoprotein particle (RNP) that is an essential retrotransposition intermediate (Hohjoh and Singer, 1996; Kulpa and Moran, 2005; Doucet et al., 2010). These RNPs also contain several host factors that are required or that control the L1 retrotransposition cycle (Goodier, Cheung and Kazazian, 2013; Taylor *et al.*, 2013, 2018). During L1 translation, nonautonomous retrotransposons Alu and SVA can hijack the L1 machinery to

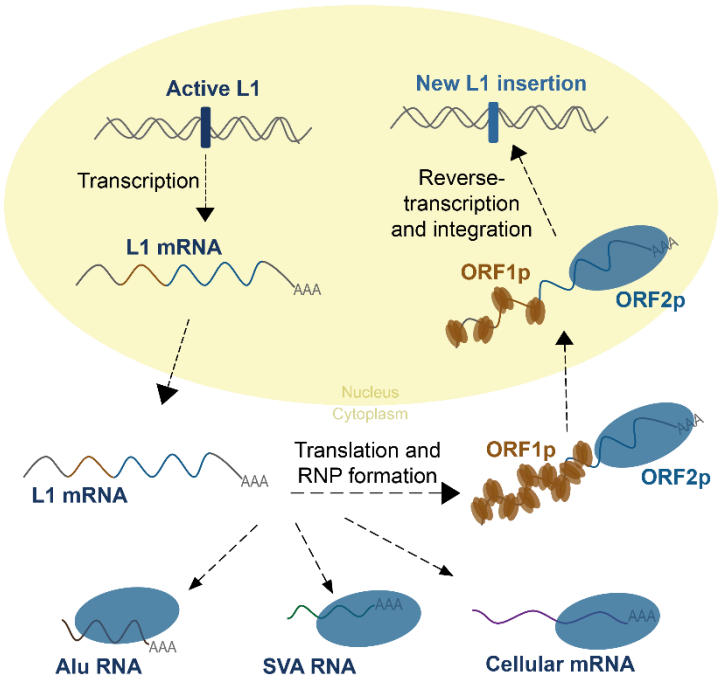


Figure 1. LINE-1 retrotransposition cycle. An active, RC-L1 is transcribed and exported to the cytoplasm. There, L1-ORF1p (brown) and L1-ORF2p (blue) are translated and form a RNP with the L1 from where they were translated. Back in the nucleus, the EN and RT domains of ORF2 catalyse reverse transcription and integration of the L1 into a new genomic location. L1-ORF2p (blue elypse) can also bind and retrotranspose other nonautonomous retrotransposons (Alu and SVA) as well as cellular messenger RNAs. Nucleus is shown in yellow. AAA denotes a poly(A) tail. Details are provided in the text.

carry out their own mobilization *in trans*, and other mRNAs can fortuitously associate with ORF2 and be retrotransposed (Dewannieux, Esnault and Heidmann, 2003; Garcia-Perez, Doucet, *et al.*, 2007; Raiz *et al.*, 2012; Ahl *et al.*, 2015).

The LINE-1 RNP then enters the nucleus, in a process that is not completely understood but does not seem to require cellular division (Kubo *et al.*, 2006; Macia *et al.*, 2017). However, several studies have suggested that cell division enhances retrotransposition (Xie *et al.*, 2013; Mita *et al.*, 2018). Besides, nuclear L1-ORF1p has remained elusive to detect, suggesting that few molecules are required for the nuclear stages of L1 mobilization (Doucet *et al.*, 2010; Taylor *et al.*, 2018).

Once in the nucleus, TPRT starts when the endonuclease domain of L1-ORF2p makes a single-strand endonucleolytic cut (or 'nick') in genomic DNA (Feng *et al.*, 1996). Notably, L1-ORF2p can target all regions of the genome, but integration is locally dictated by the presence of a degenerate consensus sequence 5'-AA/TTTT-3' (Flasch *et al.*, 2019; Sultana *et al.*, 2019). This nick exposes a 3'OH that serves as a primer for the RT domain of L1-ORF2p to reverse transcribe the L1 mRNA into the new locus (Luan *et al.*, 1993; Cost *et al.*, 2002). A poly(A) tract in the 3' end of L1 mRNA (or Alu, SVA or cellular RNAs) is strictly required for L1-ORF2p-mediated retrotransposition (Doucet, Wilusz, *et al.*, 2015). The 5' of the L1 mRNA is not frequently reached, thus most of the insertions are 5'truncated (Lander *et al.*, 2001; Beck *et al.*, 2011). Although the mechanism responsible for 5' truncation is unknown, it seems associated to DNA repair mechanisms rather than reflecting an inherent limitation of the RT to reach the 5' end of the elements during TPRT (Coufal *et al.*, 2011). The insertion is completed by nick of the complementary strand of the DNA, removal of the RNA from the RNA-DNA hybrid (perhaps via cellular RNAseH2, which has been shown to be required for L1 retrotransposition (Benitez-Guijarro *et al.*, 2018)) and synthesis of the second strand of L1 using the first L1 cDNA as a template (Richardson *et al.*, 2015). A recent report analyzed *in vitro* the insertion reaction of a LINE element from *Bombyx mori*, R2, and suggested a novel mechanism for completion of TPRT, involving a '4-way' branched DNA intermediate (Khadgi, Govindaraju and Christensen, 2019). Nevertheless, a complete understanding of TPRT requires further research and improved methods to isolate L1-ORF2p.

Remarkably, LINE-1s can also take advantage of existing DNA lesions to initiate their insertion, on a process known as EN-independent retrotransposition (Morrish *et al.*, 2002, 2007; Sen *et al.*, 2007). Indeed, a recent report suggest that ancient LINEs lacking the EN domain could have used these DNA lesions and/or 3'-OH groups present at replication forks to propagate prior to the

acquisition of the EN domain, which liberated them from this requirement and allowed L1s to be interspersed throughout the genome (Flasch *et al.*, 2019).

I.4.D. LINE-1 mobilization mechanism

To colonize almost one-fifth of our genome, L1 insertions must be heritable and therefore must occur either in germ cells or in early embryogenesis (Garcia-Perez, Widmann and Adams, 2016; Faulkner and Garcia-Perez, 2017). Indeed, pioneer work by Kazazian and col. characterized two unrelated cases of haemophilia A caused by L1 insertions in an exon of clotting factor VIII that were not present in their parents (Kazazian *et al.*, 1988).

Since then, several studies have reported L1 expression and mobilization during early embryonic development and in germ cells. Consistently, endogenous L1 RNA expression and engineered retrotransposition has been observed in human oocytes (Georgiou *et al.*, 2009); similarly, endogenous L1 expression has been found in male germ cells and testicular tumors (Ergün *et al.*, 2004). Studies of human X-linked, disease-causing, L1-mediated insertions additionally support retrotransposition occurring in germ cells and the early embryo (Brouha *et al.*, 2002; van den Hurk *et al.*, 2007; Aneichyk *et al.*, 2018). Furthermore, human embryonic teratocarcinoma cells, embryonic stem cells (hESCs) and induced pluripotent stem cells (iPSCs) support LINE-1 retrotransposition (Garcia-Perez, Marchetto, *et al.*, 2007; Garcia-Perez *et al.*, 2010; Macia *et al.*, 2011; Wissing *et al.*, 2012; Klawitter *et al.*, 2016). Importantly, human ESCs represent a physiological model to study the role of L1s in early development, and understanding the impact of transposable elements in hESCs is also fundamental for potential therapeutic applications within regenerative medicine (reviewed in (Schumann *et al.*, 2019)). At present, it is estimated that there is a new L1 insertion per ~100 births, a new Alu insertion per ~40 births, and a new SVA insertions per ~63 births (Ewing and Kazazian, 2010; Feusier *et al.*, 2019).

Of note, most of the aforementioned results have been corroborated in many studies using mouse germ and embryonic cells, transgenic mouse models, and mouse pedigrees combined with high-throughput sequencing (Branciforte and Martin, 1994; Ostertag *et al.*, 2002; Kano *et al.*, 2009; Malki *et al.*, 2014; MacLennan *et al.*, 2017; Richardson *et al.*, 2017). Interestingly, a recent report suggest that the L1 mRNA can act as an indispensable nuclear scaffold for the 2-cell mouse embryo to progress in development (Percharde *et al.*, 2018).

Curiously, even though Barbara McClintock discovered TEs in somatic tissues of maize (McClintock, 1950), it was long assumed that TE activity was restricted to germ or embryonic cells, simply because the only purpose of a 'selfish' DNA would be to accumulate copies that could be transmitted to the next generation. Fifteen years ago, this view was radically challenged. A study by Muotri and col. showed that endogenous L1s are expressed in the brain of rat and mice, and that human L1s could mobilize in rodent neural progenitor cells (NPCs) both *in vitro* and *in vivo* using a transgenic mice (Muotri *et al.*, 2005). Subsequent studies demonstrated endogenous L1 expression and engineered retrotransposition in human NPCs (Coufal *et al.*, 2009), and in hESC-derived mature, non-dividing neurons (Macia *et al.*, 2017). Simultaneously, advances in next-generation sequencing and newly developed computation tools (reviewed in (Goerner-Potvin and Bourque, 2018)), allowed the study of endogenous L1 retrotransposition in bulk brain and individual neurons. Although estimates of new L1 insertions per neuron vary greatly, it is quite clear that active retrotransposition occurs in the brain and generates somatic mosaicism (Baillie *et al.*, 2011; Evrony *et al.*, 2012, 2016; Upton *et al.*, 2015; Sanchez-Luque *et al.*, 2019). Retrotransposition in the brain may be one of the underlying causes of neurological disorders (reviewed in (Terry and Devine, 2020)), but may as well have contributed to neural complexity and to the evolution of modern humans (Guichard *et al.*, 2018). More research is warranted to explore the influence of active TEs in brain biology. Lastly, many

other adult stem cells have been tested for L1 mobilization, showing that it occurs at a very low level (reviewed in (Schumann *et al.*, 2019)), although somatic insertions have been validated in some healthy adult tissues, such as stomach (Ewing *et al.*, 2015), colon (Scott *et al.*, 2016) or esophagus (Doucet-O'Hare *et al.*, 2015).

Aside from physiological contexts, L1s have also been shown to mobilize frequently in cancer (Burns, 2017; Scott and Devine, 2017). Since this is another main point of this Thesis, it will be further discussed in section I.7.

I.5. Impact of LINE-1 retrotransposons: from evolution to disease

As a species, having active TEs in our genome can potentially confer us selective advantages against changes in the environment. However, as individuals, our genome can be 'bombarded' by retrotransposons landing randomly in our DNA, (Goodier and Kazazian, 2008; Richardson *et al.*, 2015; Muñoz-Lopez *et al.*, 2016). In fact, TEs are considered major structural variants in our genome (Huang *et al.*, 2010). New retrotransposition events mediated by LINE-1 can impact the genome through insertional mutagenesis, leading to genomic alterations such as gene disruption or splicing alterations. Besides, the accumulation of insertions (i.e. repeated sequences) increases predisposition to rearrangements, driving genomic instability. Since this has been the subject of numerous reviews (Goodier and Kazazian, 2008; Cordaux and Batzer, 2009; Konkel and Batzer, 2010; Beck *et al.*, 2011; Richardson *et al.*, 2015; Garcia-Perez, Widmann and Adams, 2016), only an overview of the potential consequences of new L1 insertions (or L1-mediated Alu/SVA insertions) will be provided in this section (several examples are illustrated in **Figure 2**):

- The first, and most obvious, is the disruption of a gene upon insertion in an exonic sequence **(Figure 2A)**. This was the case in the first disease-causing L1

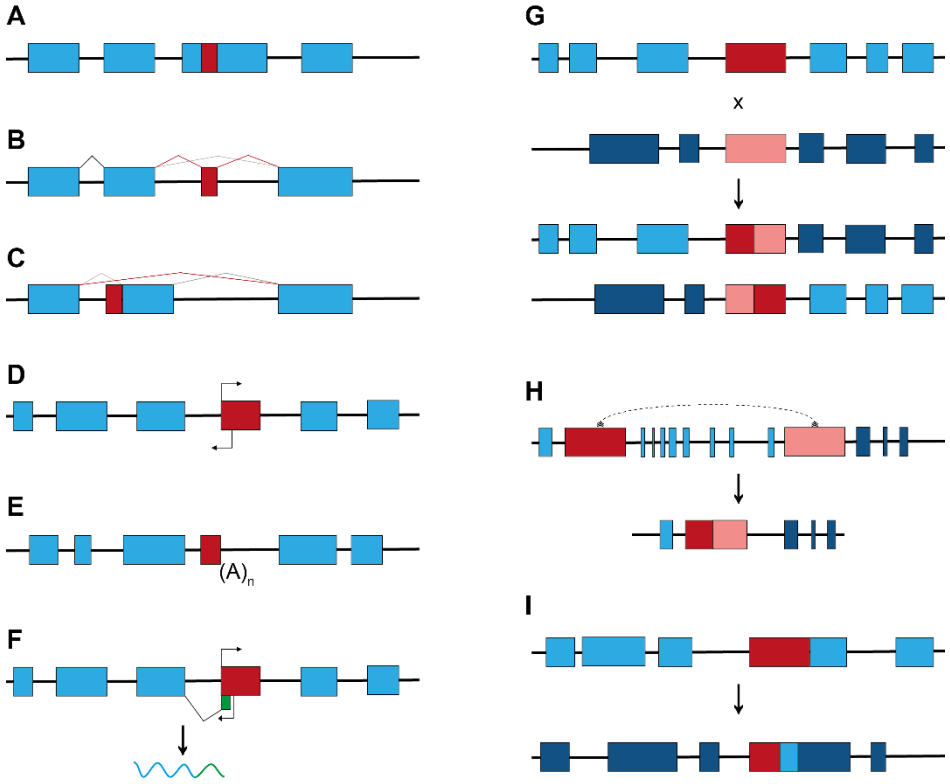


Figure 2. Several ways in which LINE-1 mobilization can impact the genome. In all cases, red boxes represent L1s, and blue boxes represent exons. (A) Exon disruption. (B) Exon inclusion. (C) Exon skipping. In (B-C) black pointed lines represent native splicing, red pointed lines represent new, altered splicing induced by L1 insertions, and pale black lines represent native splicing that the insertion alters. (D) Transcription from the sense or antisense L1 promoter. Black arrows illustrate promoter activity. (E) Premature polyadenylation. (A)n shows polyadenylation signal. (F) Nonallelic homologous recombination. (G) 3' transduction and generation of new exons.

insertion reported, where the element caused haemophilia A by disrupting exon 14 of the factor VIII gene (Kazazian *et al.*, 1988).

- Splicing sites in L1 can alter splicing of coding mRNAs (Belancio, Hedges and Deininger, 2006), generating novel isoforms **(Figure 2B)** or promoting exon skipping **(Figure 2C)**.

- The antisense promoter of L1 can drive expression of adjacent genes (Nigumann *et al.*, 2002; Criscione *et al.*, 2016) **(Figure 2D)**.

- Polyadenylation signals of L1 can induce premature polyadenylation of messenger RNAs containing intronic L1s (Han, Szak and Boeke, 2004) **(Figure 2E)**.

- The antisense ORF0 can be spliced with flanking exons to generate fusion proteins (Denli *et al.*, 2015). **(Figure 2F)**.

- L1s can mediate allelic and nonallelic homologous recombination **(Figure 2G)**, and large genomic rearrangements: inversions, deletions **(Figure 2H)** or duplications (Gilbert, Lutz-Prigge and Moran, 2002; Symer *et al.*, 2002; Rodriguez-Martin *et al.*, 2020). In fact, considering the genomic distribution of LINEs, 80% of the genome may be at risk of LINE-mediated nonallelic homologous recombination (Startek *et al.*, 2015)

- Sequences transduced by L1 can generate exon shuffling, which may lead to generation of new genes (Moran, DeBerardinis and Kazazian, 1999; Pickeral *et al.*, 2000; Sayah *et al.*, 2004) **(Figure 2I)**.

- The A/T rich nature of the L1 sequence can induce transcriptional pause of RNA polymerase II; in fact, inverse correlation has been found between gene expression and number of full length L1s contained in that gene (Han, Szak and Boeke, 2004).

While retrotransposition represents the most evident manner in which L1s can impact our genome, there are many other (perhaps even more frequent) retrotransposition-independent ways that could further influence genome function. To name a few: their transcription can alter host mRNA processing, L1-encoded proteins can mediate DNA damage, and L1 mRNA and cDNA can induce IFN response and promote sterile, age-associated inflammation (Cordaux and Batzer, 2009; Richardson *et al.*, 2015; Elbarbary, Lucas and Maquat, 2016; Bourque *et al.*, 2018; De Cecco *et al.*, 2019).

Importantly, some of these alterations have been positive from an evolutionary perspective, or co-opted to be beneficiary for the host (Cordaux and Batzer, 2009; Cosby, Chang and Feschotte, 2019). For example, LINE-mediated recruitment of RNA-binding proteins contributes to the evolution of tissue-specific RNA transcripts (Attig *et al.*, 2018) and Alu repeats control nuclear localization

of lncRNAs (Lubelsky and Ulitsky, 2018). Furthermore, it has been suggested that L1s, through mobilization of the regulatory signals they contain, can promote the generation of novel gene-regulatory networks (Friedli and Trono, 2015; Chuong, Elde and Feschotte, 2017; Imbeault, Helleboid and Trono, 2017). However, most of new L1 insertions are neutral to the host, which has enabled this elements to spread throughout the genome (Schumann *et al.*, 2019).

On the contrary, more than 120 disease-causing insertions mediated by L1 retrotransposons have been described so far in humans (Hancks and Kazazian, 2016; Burns, 2020). Some of them have occurred in exons, as the aforementioned insertion that caused haemophilia by disrupting exon 14 of clotting factor VIII gene (Kazazian *et al.*, 1988). Besides, intronic insertions involving L1 or nonautonomous retrotransposons have also been reported, such as a polymorphic Alu insertion in CD58 gene that increases the risk of developing multiple sclerosis (Payer *et al.*, 2019) and an SVA insertion in the TAF1* gene that results in intron retention and causes X-linked dystonia–parkinsonism (Aneichyk *et al.*, 2018).

I.6. Cellular mechanisms to control LINE-1 retrotransposition

While TE activity is required to confer genome plasticity, their mobilization cannot compromise the fitness of the species. Considering the tremendous impact that L1s can have in our genome (see section I.5), it does not come as a surprise that plenty of mechanisms have evolved to control human L1 mobilization. Importantly, elucidating these mechanisms is an area of intense research.

It should be first considered that L1 retrotransposons, in a way, regulate themselves. The main self-restriction mechanism is TPRT: in the majority of cases, it leads to 5' truncated insertions that are 'dead on arrival' and cannot mobilize

* TATA-box binding protein associated factor 1

again, reducing the number of RC-L1s accumulated in our genome over evolution (Beck *et al.*, 2011; Richardson *et al.*, 2015). Additionally, the L1 mRNA can be prematurely polyadenylated (Perepelitsa-Belancio and Deininger, 2003) or suffer abortive splicing (Larson *et al.*, 2018). Apart from these cases, cells have developed a large (and growing) set of restriction mechanisms involving host factors to keep retrotransposons under control. This subject has been thoroughly reviewed quite recently (Goodier, 2016; Pizarro and Cristofari, 2016), thus a brief overview will be included here, differentiating transcriptional and post-transcriptional regulatory mechanisms. An illustrative summary is provided in **Figure 3**.

I.6.A. Transcriptional.

The L1 promoter contains a prototypical CpG island with >20 CpG residues that can be methylated to repress transcription of the element (Hata and Sakaki, 1997). Notably, most of the 5-methylcytosine* in our genome is found within TEs (Yoder, Walsh and Bestor, 1997), and it has been hypothesized that gene expression regulation by methylation is a consequence of the constant battle between TEs and their hosts. Several host factors act to prevent transcription of L1s. The DNA methyl transferase DNMT1 has recently been shown to specifically act upon young, active L1 elements in human neural progenitor cells (Jönsson *et al.*, 2019). Another recent report has shown that YY1, which is required for accurate L1 transcription initiation (Athanikar, Badge and Moran, 2004), mediates methylation of young L1s, and loss of the YY1 binding site leads to hypomethylation and evasion of epigenetic repression in differentiated cells (Sanchez-Luque *et al.*, 2019). Additionally, many proteins from the KRAB-ZNF† family have expanded and evolved to target different L1 subfamilies (Castro-Diaz

* Methylation of carbon 5 in cytosine induces transcriptional repression.
† Krüppel Associated Box – Zinc Finger.

et al., 2014; Jacobs *et al.*, 2014). These proteins recognize specific sequences in DNA and recruit mediators of heterochromatin formation and DNA methylation, resulting in repression of gene expression (Ecco, Imbeault and Trono, 2017). Interestingly, it has been proposed that the arms race between TEs and KRAB-ZNFs has resulted into the evolution of new gene regulatory networks (Imbeault, Helleboid and Trono, 2017). Additionally, two studies in mouse germ cells have shown that loss of different DNA methyl transferases increases the levels of L1 mRNA (Bourc'his and Bestor, 2004; Barau *et al.*, 2016), a phenomenon that has also been observed in human cells (Castro-Diaz *et al.*, 2014). Lastly, a recent report suggest that the HUSH* complex mediates chromatin remodelling and transcriptional silencing of RC-L1s (Liu *et al.*, 2018). In sum, methylation of the promoter is considered the first line of defence against L1 mutagenic potential, although other epigenetic factors also participate in L1 silencing **(Figure 3)**.

I.6.B. Post-transcriptional.

Tens of host factors have been found to regulate LINE-1 post-transcriptionally. The first were the APOBEC3† protein family, which were shown later to regulate L1 retrotransposition by deaminating transiently exposed ssDNA‡ during integration (Bogerd *et al.*, 2006; Muckenfuss *et al.*, 2006; Richardson *et al.*, 2014). Interestingly, some APOBEC3 proteins are involved in antiviral defence, which is a common feature of many L1 restriction factors (reviewed in (Willems and Gillet, 2015)). Indeed, several antiviral proteins have also been reported to regulate L1 mobilization in cell culture: the exonuclease Trex1§ (Stetson *et al.*,

* Human Silencing Hub.
† Apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like.
‡ Single-stranded DNA
§ Three-prime-repair exonuclease

2008; Li *et al.*, 2017; Thomas *et al.*, 2017), the IFN[*]-activated endoribonuclease RNAse L[†] (Zhang *et al.*, 2014), the dNTPase[‡] SAMHD1[§] (Zhao *et al.*, 2013) or the antiviral protein ZAP[**] (Goodier *et al.*, 2015; Moldovan and Moran, 2015). Notably, other IFN-stimulated genes, such as MAVS[††] or BST2[‡‡], have been shown to inhibit L1 retrotransposition in cell culture (Goodier *et al.*, 2015), although how they achieve that regulation remains unexplored. The IFN-activated GAIT[§§] and condensin II also restrict L1 mobilization (Ward *et al.*, 2017).

On another note, RNA-processing proteins have also been involved in regulating L1 retrotransposons, such as the RNA-binding protein hnRNPL[***] (Peddigari *et al.*, 2013; Moldovan and Moran, 2015), the adenosine deaminase ADAR1[†††] (Orecchini *et al.*, 2016) or the microRNA biogenesis factor DGCR8[‡‡‡], that was shown to cleave L1 and Alu RNAs and to regulate engineered retrotransposition (Heras *et al.*, 2013). RNA helicase MOV10[§§§] has also been shown to regulate L1 mobilization in cell culture (Goodier, Cheung and Kazazian, 2012; Li *et al.*, 2013; Moldovan and Moran, 2015), and to collaborate with TUT4/7[****] to uridylate the 3'end of L1 mRNA and restrict retrotransposition (Warkocki *et al.*, 2018).

---

[*] Interferon
[†] RNAse L is activated by double-stranded RNA from a viral infection and degrades ssRNA.
[‡] Deoxynucleoside triphosphate triphosphohydrolase (an enzyme that degrades dNTPs to prevent replication of infecting viruses)
[§] SAM domain and HD domain-containing protein 1
[**] Zinc finger CCCH-type antiviral protein
[††] Mitochondrial antiviral-signaling protein
[‡‡] Bone Marrow Stromal Cell Antigen 2
[§§] Gamma-Interferon Activated Inhibitor of Translation
[***] Heterogeneous Nuclear Ribonucleoprotein L
[†††] Adenosine deaminase acting on RNA.
[‡‡‡] DiGeorge Critical Region 8
[§§§] Moloney leukemia virus 10
[****] Terminal uridyltransferases

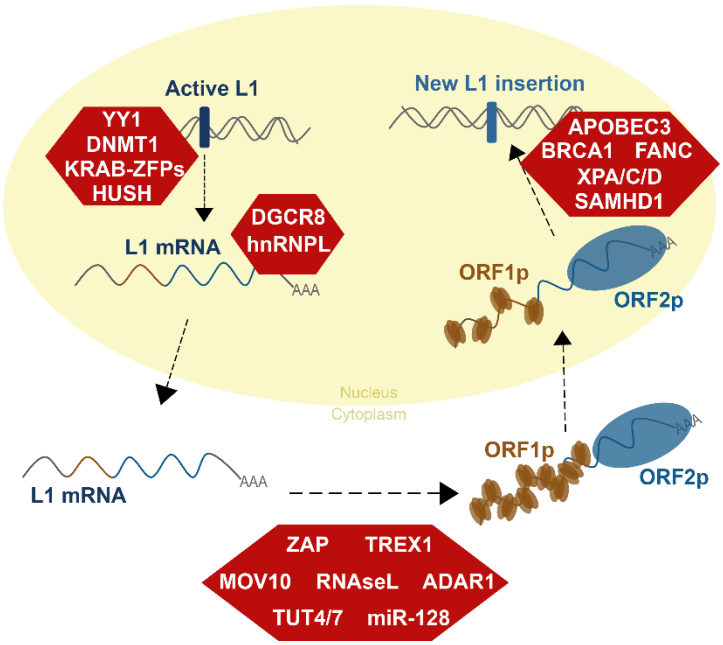


Figure 3. Cellular mechanisms that restrict L1 mobilization. Different host-factors involved in regulating different steps L1 retrotransposition are shown in red hexagons. Further details are provided within the text. Acting on L1 transcription: 5mC, 5-methylcytosine; KRAB-ZFPs, KRAB-zinc finger proteins; YY1, Ying-Yang 1; DNMT1, DNA methyl transferase 1; HUSH, Human Silencing Hub. Acting post-transcriptionally in the nucleus: DGCR8, DiGeorge Critical Region 8; hnRNPL, Heterogeneous Nuclear Ribonucleoprotein L. Acting post-transcriptionally in the cytoplasm: ZAP, Zinc-finger Antiviral Protein; TREX1, Three-prime-repair exonuclease 1; MOV10, Moloney leukemia virus 10; ADAR1, Adenosine deaminase acting on RNA; TUT4/7, Terminal uridyltransferases 4/7; miR-128, microRNA 128. Acting pre-integration in the nucleus: BRCA1, Breast cancer type 1 susceptibility protein; APOBEC3, Apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3. XPA/C/D, DNA repair protein complementing XP-A, XP-C or XP-D cells, respectively; FANC, proteins associated with Fanconi Anemia; SAMHD1, SAM domain and HD domain-containing protein 1.

Recent studies have also identified cell cycle regulators, such as the cyclin-dependent kinase inhibitor p21 and p27 (Kawano *et al.*, 2018), and DNA-repair proteins, like XPA, XPC and XPD* (Servant *et al.*, 2017) or BRCA1† and FANC‡ proteins (Liu *et al.*, 2018; Mita *et al.*, 2020) as repressors of L1 retrotransposition **(Figure 3)**. Additionally, rapid epigenetic silencing through histone deacetylation

* DNA repair protein complementing XP-A, XP-C or XP-D cells, respectively. These proteins are Involved in the nucleotide-excision repair pathway.
† Breast cancer type 1 susceptibility protein
‡ Proteins associated with Fanconi Anemia

upon insertions of engineered LINE-1s has been reported (Garcia-Perez *et al.*, 2010; Kannan *et al.*, 2017). Lastly, piRNAs* are a class of small (26-32 nt) RNAs involved in epigenetic and post-transcriptional silencing of transposons in germ cells, a role that has been thoroughly characterized in *Drosophila* and mice (Malone and Hannon, 2009; Siomi *et al.*, 2011). In humans, piRNA expression has been detected in fetal ovary and in adult testis (Williams *et al.*, 2015). Moreover, a recent report identified piRNAs mapping to the sequence of L1 in fetal testis, suggesting that active piRNAs may be functioning to repress L1s in fetal germ cells (Reznik *et al.*, 2019). Further research is needed to understand the role of piRNA in TE control in human germ cells.

Surprisingly, despite the prevalent role of microRNAs in regulation of gene expression (see section I.8 and thereafter), only one study so far has reported a microRNA involved in regulating L1 retrotransposition: miR-128 (Hamdorf *et al.*, 2015).

I.7. LINE-1 in cancer

Over the last decade, numerous reports have shown accumulation of L1 insertions in many types of cancer. In fact, around 50% of human tumors contain somatic L1 insertions. This has been thoroughly reviewed recently (Burns, 2017; Scott and Devine, 2017), and the main findings, as well as corresponding references, are summarized in **Table 2**. Briefly, the highest somatic retrotransposition has been found in epithelial cancers, mainly lung and colorectal, but also esophageal, pancreatic and ovarian cancer. Somatic insertions have also been detected, although at much lower levels, in breast, liver, kidney and prostate tumors (see **Table 2** for references).

* PIWI-interacting RNAs

Type of tumor	References	L1 activity
Lung	(Iskow *et al.*, 2010; Helman *et al.*, 2014; Tubio *et al.*, 2014)	+++
Colorrectal	(Lee *et al.*, 2012; Solyom *et al.*, 2012; Pitkänen *et al.*, 2014; Tubio *et al.*, 2014; Ewing *et al.*, 2015; Scott *et al.*, 2016)	
Esophageal	(Doucet-O'Hare *et al.*, 2015; Paterson *et al.*, 2015)	++
Pancreatic	(Rodić *et al.*, 2015)	
Ovarian	(Lee *et al.*, 2012; Tang *et al.*, 2017; Nguyen *et al.*, 2018)	
Breast	(Tubio *et al.*, 2014)	+
Liver	(Shukla *et al.*, 2013; Schauer *et al.*, 2018)	
Kidney	(Helman *et al.*, 2014)	
Prostate	(Lee *et al.*, 2012; Tubio *et al.*, 2014)	
Brain	(Carreira *et al.*, 2016)	-

Table 2. Types of cancer where L1 retrotransposition has been described. L1 activity refers to the amount of somatic insertions found in those tumors.

Importantly, these results were confirmed very recently in a massive analysis of 2,954 cancer genomes from 38 different types of tumors (Rodriguez-Martin *et al.*, 2020). In this study, Rodriguez-Martin and collaborators found that L1 insertions are the most frequent cause of structural variation in esophageal carcinoma, the second in colon and the third in lung cancer (Rodriguez-Martin *et al.*, 2020). These structural variations range from large deletions, that can deplete tumor-suppressor genes to large duplications, that can amplify oncogenes (Rodriguez-Martin *et al.*, 2020).

The evidence at the genomic level is consistent with immunohistochemistry studies that have shown expression of L1-ORF1p in many types of tumor samples. It was detected in over 90% of the breast, ovarian and pancreatic cancer, and over 50% of the esophageal, colorectal and lung (Rodić *et al.*, 2014). Strikingly, a recent, mass spectrometry-based tumor proteome profiling study was

unable to detect L1-ORF2p, suggesting that the expression of this protein is even lower than expected (Ardeljan *et al.*, 2019).

Several reports have shown that somatic L1 insertions can drive tumorigenesis (Shukla *et al.*, 2013; Helman *et al.*, 2014) and may even initiate the tumor in normal cells, although only four instances have been described so far: two insertions that interrupted the APC gene (Miki *et al.*, 1992; Scott *et al.*, 2016), one that interrupted the tumor-suppressor gene PTEN (Helman *et al.*, 2014), and one that affected a regulatory sequence within the ST18 gene and increased its expression in a hepatocellular carcinoma (Shukla *et al.*, 2013). Interestingly, ST18 was later shown to act as an oncogene mediating the progression of liver cancer (Ravà *et al.*, 2017). In sum, the evidence, and the fact that mutagenic insertions are not frequent, suggest that L1s might drive tumorigenesis more frequently than they initiate the tumor, although more high-throughput single-cell sequencing studies could challenge this assumption.

Somatic L1 insertions can also mediate considerable genomic rearrangement that may lead to oncogene amplification or removal of tumor-suppressor genes, which can result in increased survival and expansion of specific clones in the tumor (Rodriguez-Martin *et al.*, 2020). Notably, pioneer experiments in cultured cells already associated L1 with large genomic rearrangements and genome instability (Gilbert, Lutz-Prigge and Moran, 2002; Symer *et al.*, 2002). In fact, genomic instability is one of the main characteristics of human cancer, and is associated with increased risk of metastases and worse prognosis. Interestingly, a recent study reported that, while L1 expression enhances the growth advantage of p53-deficient cancer cells, it also induces DNA replication stress, making these cells more vulnerable to DNA-damaging chemotherapies (Ardeljan *et al.*, 2020).

Given the important role of L1 retrotransposition in the increased genome instability of human tumors, a fundamental question in the field is to gain insight into the causes of this L1 deregulation in cancer, to identify which controlling

mechanisms may be altered. One of the main mechanisms to control L1 activity occurs at the transcriptional level through methylation of the L1 promoter (Bourc'his and Bestor, 2004; Sanchez-Luque *et al.*, 2019), and it has been shown that the hypomethylation of specific L1s can activate retrotransposition early in tumorigenesis (Tubio *et al.*, 2014; Scott *et al.*, 2016). Consistently, hypomethylation of the L1 promoter has been associated with worse prognostic in lung cancer patients (Saito *et al.*, 2010). However, additional post-transcriptional mechanisms that control L1 mobilization and whose downregulation could contribute to the increased retrotransposition observed in tumors have not been thoroughly investigated.

I.8. Micro-introduction to microRNAs

Sixty years ago, Jacob and Monod suggested that the Lac repressor could be an "RNA fraction" that "acts at the cytoplasmic level, by controlling the activity of the messenger", as opposed to "the genetic level, controlling the production of the messenger" (Jacob and Monod, 1961). It turned out to be a protein, as were all the regulators of gene expression discovered in the next three decades. This led to the assumption that, perhaps, all regulators were proteins. However, a 1993 discovery strongly challenged this idea: two different labs reported that the lin-4 gene of the worm model *Caenorhabditis elegans* produced, instead of mRNAs, short non-coding RNAs that controlled developmental transitions by base pairing to a target mRNA, *lin-41*, with sequences of partial complementarity on its 3'UTR (Lee, Feinbaum and Ambros, 1993; Wightman, Ha and Ruvkun, 1993). A similar mode of action and expression pattern was later discovered for another critical gene for *C. elegans* development: let-7 (Reinhart *et al.*, 2000). Shortly after, let-7 sequence and expression pattern were found in many other bilaterian animals, including humans (Pasquinelli *et al.*, 2000). These studies, together with the study that first introduced the phenomenon of RNA interference (Fire *et al.*, 1998)[*], led to a model in which these small RNAs mediated transcript degradation and/or translational repression via antisense interactions with target mRNAs. Many more 'short temporal RNAs' were identified in *C. elegans*, *Drosophila* and humans, and named 'microRNAs' due to the only thing that was certain at that moment: that they were tiny (Lagos-Quintana *et al.*, 2001; Lau *et al.*, 2001; Lee and Ambros, 2001).

MicroRNAs (miRNAs) can be defined as ~22nt single stranded RNAs that, upon loading into the Argonaute (AGO) family of proteins to form the RNA-

[*] This discovery earned A. Fire and C. Mello the 2006 Nobel Prize in Physiology or Medicine.

Induced Silencing Complex (RISC), direct post-transcriptional repression of gene expression (Bartel, 2018). Nowadays, we know that hundreds of *bona fide* miRNAs exist in humans, many of which are conserved in other animals, and each of them is predicted to target several mRNAs (Friedman *et al.*, 2009). Essentially, miRNAs are involved all human developmental, physiological and pathological processes (Bartel, 2018). Indeed, the use of miRNAs for therapy, either as 'drugs' or as targets, is likely to open a new era in medicine, particularly for the treatment of cancer (reviewed in (Rupaimoole and Slack, 2017)).

I.9. Overview of miRNA genomics and nomenclature

The first discovered miRNA genes were named after their mutant phenotype, such as let-7 for 'lethal' (Pasquinelli *et al.*, 2000; Reinhart *et al.*, 2000). However, the vast majority of miRNAs are named 'miR-' followed by a number: miR-1, miR-2, etc (Fromm *et al.*, 2015). MicroRNAs named with different numbers differ in their 5' sequence, or 'seed', which is fundamental for target recognition (see section I.12). As an example, miR-20 and miR-17 contain different 'seed' sequences and will therefore pair to different target mRNAs. Additionally, when paralog miRNA genes exist in a genome, they are differentiated by adding a letter suffix (-a, -b…) to genes producing similar mature miRNAs and a number suffix (-1, -2…) to genes producing identical mature miRNAs. For example, mature miR-34a and miR-34b are very similar but not identical, whereas mature let-7a-1 and let-7a-2 are identical but transcribed from two different loci. MicroRNAs can be encoded as individual genes or as clusters containing different miRNAs that are part of a 'polycistronic' transcript (Bartel, 2018; Treiber, Treiber and Meister, 2019). When this happens, each miRNA hairpin (see section I.10) is named as deriving from an individual gene to simplify annotation. In fact, they are processed independently (Kim, Han and Siomi, 2009). Lastly, while many miRNAs are processed from long non-coding RNAs with no other function, some

miRNAs are encoded in the introns of protein-coding genes, although functional connection between a miRNA and its host gene is infrequent (Ha and Kim, 2014; Treiber, Treiber and Meister, 2019).

There are approximately 500 *bona fide* miRNAs in humans (Fromm *et al.*, 2015), although almost 2,000 annotations can be found in miRBase[*], the primary public repository for miRNAs (Griffiths-Jones *et al.*, 2006; Kozomara, Birgaoanu and Griffiths-Jones, 2019). This difference does not mean that the other annotated sequences do not have a biological function, but not every small RNA can be classified as miRNA (Fromm *et al.*, 2015). Indeed, other types of small RNAs with regulatory functions exist in humans, such as tRNA-derived fragments or tRFs (Lee *et al.*, 2009). MiRNAs that share the 'seed' region, and therefore potentially target the same mRNAs, are grouped into families such as the let-7 family or the miR-1/206 family (Bartel, 2018). However, belonging to the same 'family' does not imply common ancestry. Moreover, paralog miRNAs not necessarily belong to the same family: miR-200a and miR-200b share all but one nucleotide which in the 'seed' region, thus they have different specificities (Bartel, 2018).

This Thesis is mainly focused on let-7; therefore, a dedicated section is included in I.14. However, an overview on global miRNA biogenesis, regulation and mechanisms of action will be provided first.

I.10. Biogenesis of miRNAs

The first step in microRNA biogenesis is transcription. MicroRNAs are generally transcribed by RNA polymerase II (Cai, Hagedorn and Cullen, 2004; Lee *et al.*, 2004), although a small subset can be transcribed by RNA polymerase III as their transcription is guided by upstream Alu elements (Borchert, Lanier and

[*] http://mirbase.org/

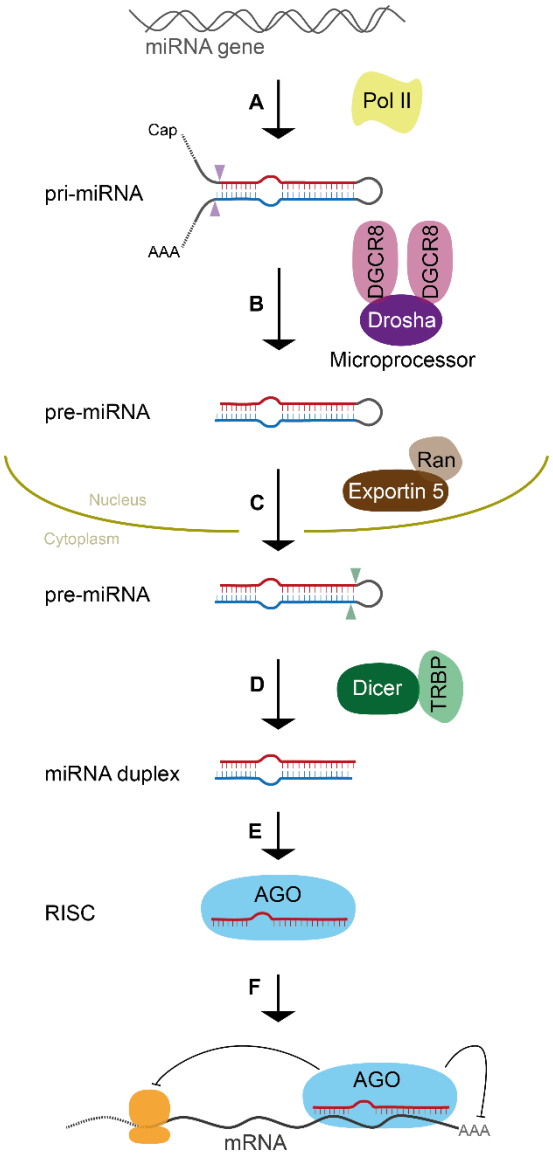


Figure 4. MicroRNA biogenesis. (A) The miRNA gene is transcribed by RNA polymerase II, shown in yellow, to yield a long pri-miRNA (primary microRNA) with a hairpin structure. **(B)** The pri-miRNA is processed at the base of the hairpin by the Microprocessor releasing the pre-miRNA (precursor microRNA). The Microprocessor complex is comprised by one molecule of Drosha (shown in purple) and two molecules of DGCR8 (shown in pink). **(C)** Exporting 5 and RanGTP mediate nuclear export of pre-miRNA to the cytoplasm. **(D)** Dicer (shown in dark green) partners with TRBP (shown in light green) and cleaves the loop of the hairpin, releasing a microRNA duplex. **(E)** One of the strands is loaded into one of the AGO proteins (shown in blue) to form the RISC, or RNA induced silencing complex. **(F)** The miRNA guides RISC to complementary sites in target mRNAs, which results in inhibition of translation (ribosomes shown in orange) or mRNA deadenylation and degradation

Davidson, 2006). Transcription yields a long, polyadenylated transcript termed primary microRNA (pri-miRNA) that harbours a hairpin structure containing the miRNA sequence **(Figure 4A)** as well as other sequence and structural motifs that are critical for pri-miRNA processing (Fang and Bartel, 2015). In the case of clustered miRNAs, usually long transcripts contain several hairpin structures that will undergo separate processing. This pri-miRNA undergoes two subsequent processing steps mediated by two RNAse III* enzymes: Drosha in the nucleus and

* A type of ribonuclease that cleaves dsRNA.

Dicer in the cytoplasm. Drosha partners with its cofactor DGCR8* to form the Microprocessor complex (Gregory *et al.*, 2004), comprised by one Drosha and two DGCR8 molecules (Nguyen *et al.*, 2015). The Microprocessor cleaves the hairpin structure of the pri-miRNA at its base, releasing a small hairpin called precursor miRNA (pre-miRNA) (Lee *et al.*, 2003) **(Figure 4B)**. The pre-miRNA is exported to the cytoplasm by exportin-5 and RanGTP† (Yi *et al.*, 2003; Bohnsack, Czaplinski and Gorlich, 2004; Lund *et al.*, 2004) **(Figure 4C)**. Once there, pre-miRNAs are cleaved by Dicer close to the hairpin, generating ~22nt miRNA duplexes (Bernstein *et al.*, 2001; Grishok *et al.*, 2001; Hutvágner *et al.*, 2001) **(Figure 4D)**. Dicer also requires a partner, TRBP‡ (Chendrimada *et al.*, 2005). However, unlike DGCR8 for Drosha, TRBP appears not to be essential for Dicer function (Ha and Kim, 2014). Finally, the miRNA duplex is loaded into the Argonaute (AGO) family of proteins **(Figure 4E)** with assistance of chaperones (Iwasaki *et al.*, 2010). Four AGO proteins exist in humans (AGO1-4) and miRNAs are incorporated indiscriminately into them (Dueck *et al.*, 2012), although only AGO2, the most highly and widely expressed, has RNA cleavage activity (Meister *et al.*, 2004; Meister, 2013). Finally, one of the strands of the duplex is discarded, a process not yet completely understood (Kim, Han and Siomi, 2009; Meister, 2013). The miRNA then guides the silencing complex to sites within mRNAs to mediate their posttranscriptional repression (Jonas and Izaurralde, 2015) **(Figure 4F)**.

Here, the canonical miRNA biogenesis has been detailed, however some cases of non-canonical (Microprocessor-independent or Dicer-independent) mechanisms of miRNA biogenesis have been described (see (Kim, Han and Siomi,

* DiGeorge Critical Region 8
† Ras-related nuclear protein-GTP
‡ Transactivation response element RNA-binding protein

2009) or (Bartel, 2018)). Indeed, in a recent study Drosha, Dicer and exportin 5 were depleted individually in a human tumor cell line and miRNA-seq* was performed (Kim, Kim and Kim, 2016). Strikingly, while the absence of Drosha abolished the expression of canonical miRNAs, several were still detected in Dicer- or exportin 5-knockout cells (Kim, Kim and Kim, 2016), suggesting that additional alternative pathways may exist for canonical miRNA biogenesis.

I.11. Regulation of miRNA expression

The majority of miRNA genes harbour their own RNA polymerase II promoter, therefore are subject to similar transcriptional regulation as protein-coding genes (Ha and Kim, 2014; Treiber, Treiber and Meister, 2019). Unsurprisingly, expression of miRNA genes can be regulated by transcription factors in embryonic stem cells (Marson *et al.*, 2008) or cancer cells (Ozsolak *et al.*, 2008). To highlight just one notable example: transcription of the miR-34 family is driven by the tumor-suppressor protein TP53 (He *et al.*, 2007). Consistently, expression of microRNA genes located in the introns of protein-coding genes is under the same transcriptional regulation as the host gene (Baskerville and Bartel, 2005). However, the vast majority of regulatory mechanisms for miRNA function work at the post-transcriptional level and are mediated by RNA-binding proteins or RBPs. This has been extensively reviewed recently (see (Ha and Kim, 2014; Michlewski and Cáceres, 2019; Treiber, Treiber and Meister, 2019)), therefore only a brief overview will be provided here.

First, it should be noted that miRNA expression does not necessarily predict function. For instance, pri-miRNA susceptibility to Microprocessor processing is a better predictor for miRNA abundance than transcription itself (Conrad *et*

* A variation of RNA-sequencing that allows isolation and identification of miRNAs

al., 2014), and effective binding to AGO proteins is a better indicator of the inhibitory potential of a miRNA than its overall abundance (Flores *et al.*, 2014). In fact, a recent study in *Drosophila* cells suggest that AGO loading represents a critical kinetic bottleneck for miRNA function (Reichholf *et al.*, 2019). Therefore, a high abundance of a miRNA does not necessarily result in comparably greater repression of its targets.

Interestingly, two recent, systematic high-throughput studies have identified many RBPs that can post-transcriptionally regulate miRNA biogenesis in a cell-type specific manner. One used pull-down of 70 pri-miRNAs followed by mass spectrometry to identify 180 potential RBPs, and functionally validated several of them (Treiber *et al.*, 2017). The other analyzed eCLIP* datasets for RBPs and found that 116 of them bind to pri- or pre-miRNA, validating several of them (Nussbacher and Yeo, 2018). Remarkably, many RBPs discovered in these two studies to be involved in miRNA biogenesis have been previously found to function in mRNA processing, suggesting that the biogenesis of mRNAs and miRNAs could be coordinated in a cell type-dependent manner (Treiber, Treiber and Meister, 2019). Consistently, hnRNP A1†, a protein involved in many aspects of mRNA processing, had been previously found to positively regulate the biogenesis of miR-18a (Guil and Cáceres, 2007).

Perhaps the most well-studied RBP controlling miRNA biogenesis is the stem cell-specific protein LIN28, which interacts with the pri- and pre-miRNA of most let-7 family members to suppress their expression (Heo *et al.*, 2008; Viswanathan, Daley and Gregory, 2008; Triboulet, Pirouz and Gregory, 2015; Ustianenko *et al.*, 2018). LIN28A recognizes specific sequences in the loop of

* Enhanced Cross-Linking and Immunoprecipitation
† Heterogeneous Nuclear Ribonucleoprotein A1

pre-let-7, and then recruits TUT4 and TUT7* which add a poly-U tail that prevents Dicer processing and marks the pre-miRNA for degradation (Heo *et al.*, 2008; Piskounova *et al.*, 2011; Thornton *et al.*, 2014). In contrast, LIN28B bind to the loop of pri-let-7 in the nucleus and blocks its processing by the Microprocessor (Viswanathan, Daley and Gregory, 2008; Piskounova *et al.*, 2011).

The recent discovery of hundreds of long intergenic noncoding RNAs (lincRNAs) and circular RNAs (circRNAs) adds another layer of complexity to the regulation of gene and miRNA expression. Some of them can act as microRNA 'sponges', which indirectly leads to increased expression of the target mRNAs that were regulated by the 'sequestered' miRNAs (Hansen *et al.*, 2013; Kallen *et al.*, 2013; Kleaveland *et al.*, 2018; Gebert and MacRae, 2019).

Notably, regulation of the proteins involved in miRNA biogenesis indirectly controls miRNA production (Ha and Kim, 2014; Treiber, Treiber and Meister, 2019). One notable example is the homeostatic autoregulation between Drosha and DGCR8: DGCR8 interacts and stabilizes Drosha, while Drosha cleaves and destabilizes DGCR8 mRNA (Han *et al.*, 2009). Additionally, post-translational modifications in DGCR8, Drosha or AGO, such as phosphorylation, ubiquitylation or sumoylation†, influence miRNA production in various ways (reviewed in (Treiber, Treiber and Meister, 2019) and (Gebert and MacRae, 2019)).

I.12. Target recognition and regulatory mechanisms

The interaction between the miRNA and its target mRNA determines how it regulates the expression of that gene and is challenging to infer or predict, often requiring experimental validation using reporter-based assays. An extensive

* Terminal Urydil Transferases 4 and 7

† Covalent attachment of a Small Ubiquitin-like Modifier (SUMO) peptide, which alters the molecular interactions of modified target proteins and induces changes in localization or stability.

pairing of a miRNA loaded into AGO2 with its target causes mRNA slicing, since AGO2 retains its endonucleolytic ability (Meister *et al.*, 2004). Although this mechanism is common in plants, in animals it has only been reported for 20 cellular transcripts (Bartel, 2018). Therefore, the most frequent repression mode by far in humans is by partial pairing of the miRNA to its target mRNA, which does not lead to mRNA slicing but to a decrease in protein synthesis either by direct translation inhibition or by mRNA destabilization (Brümmer and Hausser, 2014; Jonas and Izaurralde, 2015; Bartel, 2018; Dexheimer and Cochella, 2020).

I.12.A. MiRNA-mRNA pairing: beyond the 'seed'

Target recognition occurs primarily through Watson-Crick pairing between the binding site in the mRNA and nucleotides 2-7 of the miRNA, a region termed 'seed' (Lewis, Burge and Bartel, 2005; Bartel, 2009) **(Figure 5)**. The most frequent sites (canonical sites) involve Watson-Crick pairing of the six nucleotides in the seed, and receive different names according to the pairing of the flanking nucleotides. The 8mer site encompasses a match in nucleotide 8 of the miRNA and an adenosine (A) in the mRNA position opposite to position 1 of the miRNA **(Figure 5)**, and is the site that triggers the most potent repression (Bartel, 2018). Interestingly, the preference for an A is observed regardless of the seed sequence of the miRNA, as it promotes interactions with an adenosine-binding pocket in AGO2 that is structurally favoured (Schirle, Sheu-Gruttadauria and MacRae, 2014). Besides, there are two 7mer sites: 7mer-A1, with an A in the mRNA position opposite to position 1 of the miRNA, and 7mer-m8, which lacks that A but has a matching nucleotide to position 8 of the miRNA. Lastly, 6mer sites harbour a pairing only to the seed region of the miRNA. **(Figure 5)**.

Additional non-canonical sites have been described recently, although their effect is less potent than the canonical 8mer or 7mer sites (Kim *et al.*, 2016). Among the non-canonical sites, the most frequent are called offset 7mer and

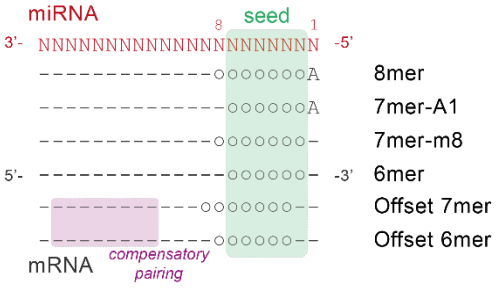


Figure 5. Types of seed found in human miRNAs. MicroRNA is shown in red, 'N' representing each nucleotide. The seed region (nucleotides 2-7 of miRNA) is highlighted in green. Circles represent nucleotides in the mRNA that match (by Watson-Crick pairing) to those in the miRNA. 'A' represents an adenosine matching nucleotide 1 of the miRNA. Compensatory pairing that occurs in the noncanonical offset 7mer and 6mer sites is marked in purple. Sites are ranked from most efficient (at the top, 8mer) to less efficient (at the bottom, offset 6mer).

offset 6mer, since they present a displaced 7 or 6 nucleotides pairing to the miRNA, respectively **(Figure 5)**. Notably, these sites usually present compensatory pairing to the 3' region of the miRNA, starting at position 13 (Bartel, 2009; Kim *et al.*, 2016) **(Figure 5).** Indeed, recent high-throughput analysis of AGO2 cross-linking and immunoprecipitation (CLIP) has uncovered pairing in the 3' end of the miRNA as a major determinant of target specificity (Moore *et al.*, 2015), which was further corroborated in *C. elegans* using a different technique (Broughton *et al.*, 2016). Functional offset seeds have also been characterized and validated experimentally other organisms such as *Anopheles gambiae* mosquitoes (Fu *et al.*, 2020).

As mentioned above, prediction of target mRNAs for any miRNA *in silico* is not always accurate. Many computational tools such as TargetScan (Agarwal *et al.*, 2015), miRanda (John *et al.*, 2004) or RNA22 (Miranda *et al.*, 2006), among others, have been optimized to precisely identify miRNA binding sites in mRNAs. However, many of these sites are not validated experimentally and therefore are not considered *bona fide* targets (Bartel, 2009; Lee and Dutta, 2009). Interestingly, in the past few years, a different approach has been taken and some techniques have been developed to unbiasedly discover new targets by analysing miRNA-mRNA interactions *in vivo* (see section I.13).

I.12.B. Effect of miRNAs on translation and mRNA stability

The understanding of how miRNAs regulate their targets has changed substantially since their discovery, when they were thought to essentially inhibit translation (Wightman, Ha and Ruvkun, 1993). At present, it is accepted that miRNA-mediated regulation of gene expression can occur in two non-exclusive ways: miRNAs can induce mRNA destabilization through deadenylation (Wu, Fan and Belasco, 2006; Baek *et al.*, 2008; Selbach *et al.*, 2008; Guo *et al.*, 2010), or translation inhibition, affecting mainly translation initiation (Pillai, 2005; Baek *et al.*, 2008; Hausser *et al.*, 2013; Zhang *et al.*, 2018). Destabilization of target mRNAs accounts for most of the observed miRNA-mediated repression (Guo et al., 2010; Eichhorn et al., 2014).

Mechanistically, the first step of miRNA-mediated repression in humans is the interaction of miRNA-loaded AGO with a protein from the GW182 family, named TNRC6A/B/C[*] in humans, to form the RISC (Liu *et al.*, 2005; Meister *et al.*, 2005; Baillat and Shiekhattar, 2009). TNRC6 then recruits the CCR4-NOT[†] and PAN2-PAN3[‡] deadenylation complexes (Braun *et al.*, 2011; Chekulaeva *et al.*, 2011; Fabian *et al.*, 2011). The CCR4/NOT complex further recruits DDX6[§], an RNA helicase that functions both as a translation repressor and as an enhancer of mRNA decapping via recruitment of DCP2[**], a component of the cellular 5′-to-3′ mRNA decay pathway (Chen *et al.*, 2014; Mathys *et al.*, 2014). Altogether, these complexes induce mRNA degradation via deadenylation and decapping (Jonas and Izaurralde, 2015).

[*] Trinucleotide Repeat-Containing gene 6A/B/C
[†] Carbon Catabolite Repression—Negative On TATA-less
[‡] Poly-A Nuclease
[§] DEAD box protein 6
[**] Decapping Protein 2

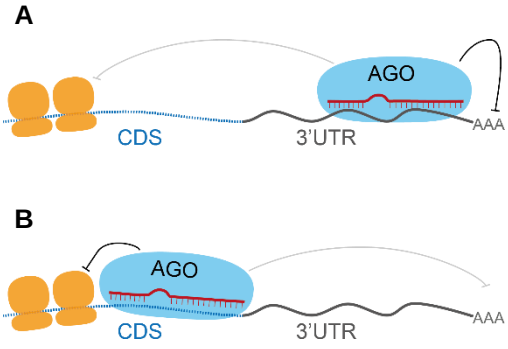


Figure 6. Effect of miRNAs on target mRNAs depends on the binding site location. (A) Binding site in the 3'UTR preferentially triggers mRNA deadenylation and destabilization, resulting in decreased mRNA levels. **(B)** Binding in CDS (Coding Sequence) preferentially results in translation inhibition without necessarily affecting mRNA levels. CDS, coding sequence; 3'UTR, 3' untranslated region; AGO, Argonaute. CDS is shown in blue and 3'UTR in grey. A miRNA is shown in red loaded into AGO, represented by blue circles. AAA denotes poly(A) tail.

The mechanisms mediating 'pure' translational repression (i.e. that cannot be explained by mRNA degradation) remain poorly understood, even though this could account for up to 26% of the repression of target mRNAs in mammalian cells (Eichhorn *et al.*, 2014). Two models have been suggested (Izaurralde, 2015). One of them suggests that RISC recruits and locks eIF4A2* on the 5'UTR of mRNAs, blocking ribosome scanning (Meijer *et al.*, 2013). Another proposes that RISC displaces eIF4A1 and eIF4A2 from the target mRNA, preventing translation initiation (Fukao *et al.*, 2014). Nevertheless, miRNA repression always results in decreased levels of the target protein, with or without reduction in the target mRNA levels.

As mentioned previously, inducing degradation of the mRNA is considered the most predominant miRNA repression mechanism rather than 'pure' translation inhibition (Guo *et al.*, 2010; Eichhorn *et al.*, 2014). Nevertheless, this is strongly influenced by the location of the miRNA binding site within the mRNA, as miRNA binding in the 3'UTR will mainly direct mRNA degradation **(Figure 6A)** whereas sites in the coding sequence will preferentially mediate translation inhibition, without necessarily affecting mRNA stability **(Figure 6B)** (Brümmer and Hausser, 2014; Hausser and Zavolan, 2014; Zhang *et al.*, 2018). Initially, it

* Eukaryotic Initiation Factor 4F2

was assumed that functional miRNA binding sites were restricted to the 3'UTR of mRNAs (Bartel, 2009; Gu *et al.*, 2009). However, in the past decades several studies have reported functional binding sites located in the coding sequence (Duursma *et al.*, 2008; Forman, Legesse-Miller and Coller, 2008; Hausser *et al.*, 2013; Zhang *et al.*, 2018), and translation repression has been reported in cultured human cells when a 3'UTR binding site for let-7a is moved to the 5'UTR of the same mRNA (Lytle, Yario and Steitz, 2007). Interestingly, the Hepatitis virus C functionally sequesters human miR-122 through binding sites located in its 5'UTR (Luna *et al.*, 2015). Since 3'UTRs are normally much larger than exons, 3'UTR sites are more frequent, which could explain that, on average, the main effect of miRNAs is to destabilize mRNAs (Guo *et al.*, 2010; Eichhorn *et al.*, 2014). Lastly, a single miRNA does not have a substantial effect in protein synthesis (Baek *et al.*, 2008; Selbach *et al.*, 2008), but most 3'UTRs contain dozens of miRNA binding sites, allowing a multi-layered, potent and fine-tuned regulation (Friedman *et al.*, 2009).

Strikingly, miRNAs have also been reported to up-regulate translation (Vasudevan, Tong and Steitz, 2007), although this phenomenon remains largely unexplored.

I.13. Methods to identify miRNA-mRNA interactions *in vivo*

Even though *in silico* tools have been developed to identify miRNA binding sites in mRNAs (Agarwal *et al.*, 2015), predictions are not accurate and often require time-consuming experimental validation, making target identification challenging (Bartel, 2009). To overcome these limitations, in the past decade several techniques have been developed to enable unbiased discovery of new miRNA targets by analysing miRNA-mRNA interactions *in vivo* (Broughton and Pasquinelli, 2016).

Over the past decade, several techniques based on ultraviolet cross-linking and immunoprecipitation (CLIP) of RNA binding proteins followed by high-throughput sequencing of the bound RNA have allowed the identification of thousands of regulatory proteins with functions in mRNA processing or location (Hentze *et al.*, 2018). Indeed, the growing number of CLIP variations (i.e. using fluorescent instead of radioactive adapters) has popularized this technique and allowed identification of specific binding sites in target mRNAs for many RNA-binding proteins (Lee and Ule, 2018). Regarding miRNAs, the first approach involved CLIP of AGO proteins and sequencing the mRNAs bound to AGO (Chi *et al.*, 2009). This approach, with some variations, has been used to discover mRNAs regulated by miRNAs in several tissues (Boudreau *et al.*, 2014; Spengler *et al.*, 2016). While useful, these approaches require the inference of which miRNA was guiding AGO to that site in the mRNA. In the past few years, these techniques have been improved to overcome the aforementioned limitation with the addition of a new step: an RNA-RNA ligation between the miRNA and its target mRNA (Broughton and Pasquinelli, 2016). Upon ligation and sequencing, chimeric reads containing the miRNA bound to its target mRNA can be found, unbiasedly identifying miRNA-target mRNA interactions. The first technique that took advantage of this additional step was CLASH*, first developed in yeast (Kudla *et al.*, 2011) and later employed successfully in cultured cells overexpressing an epitope-tagged version of AGO1 (Helwak *et al.*, 2013; Helwak and Tollervey, 2014). The development of better antibodies to immunoprecipitate AGO led to a variant of CLASH termed CLEAR-CLIP†, in which endogenous AGO2 was pulled down to obtain a more 'physiological' view of the miRNA-mRNA interaction landscape (Moore *et al.*, 2015). Interestingly, CLEAR-CLIP

* Cross-Linking and Sequencing of Hybrids
† Covalent ligation of endogenous Argonaute-bound RNAs-CLIP

has been recently employed to identify miRNA-mRNA interactions in *Anopheles gambiae* mosquitoes (Fu *et al.*, 2020).

These techniques are not exempt of limitations and technical biases, most of them based on the inherent inefficiency of the RNA-RNA ligation process and on the presence of two cross-linked sites in these chimeric RNAs, which greatly reduce the processivity of the reverse-transcriptase during sequencing library preparation (Hocq *et al.*, 2018). Consequently, it is not surprising that the number of miRNA-mRNA chimeras is usually below 1% of all sequenced reads (Broughton and Pasquinelli, 2016). Thus, some chimeras represented by just one read might represent real targeting, and can give valuable, unbiased information about which miRNAs are regulating which mRNAs *in vivo*.

I.14. The human let-7 family of microRNAs

The lethal-7 (let-7) family of miRNAs was the second miRNA ever discovered (Reinhart *et al.*, 2000) and the first in humans (Pasquinelli *et al.*, 2000). The name comes from the 'lethal' phenotype of let-7 mutants in *C. elegans* (Reinhart *et al.*, 2000). At present, it is known that mature let-7 is conserved throughout bilaterian animals (Roush and Slack, 2008; Hertel *et al.*, 2012), suggesting a major role in metazoan biology.

Genomic organization

The number of different let-7 miRNAs varies between species. For instance, while *C. elegans* and Drosophila have only one let-7 copy, Danio rerio (zebrafish) has 11 different mature sequences (Griffiths-Jones *et al.*, 2008; Kozomara, Birgaoanu and Griffiths-Jones, 2019). In humans, there are 9 different mature let-7 sequences (let-7a/b/c/d/e/f/g/i and miR-98) produced from 12 precursor sequences (Roush and Slack, 2008; Kozomara, Birgaoanu and Griffiths-Jones, 2019). This apparent discrepancy comes from the fact that some members are expressed from different precursor miRNAs located in different genomic loci:

three different precursors (let-7a-1, let-7a-2 and let-7a-3) give rise to the same mature miRNA (let-7a). The same occurs for let-7f-1 and let-7f-2. Additionally, miR-98 is considered a member of the let-7 family as well since it shares the seed region with the rest of them (Roush and Slack, 2008).

Biological functions

Broadly, let-7 miRNAs are involved in terminal differentiation of cells during development (Zhao *et al.*, 2010; Kuppusamy *et al.*, 2015; Pobezinsky *et al.*, 2015; Shenoy, Danial and Blelloch, 2015) and in control of cell proliferation (Johnson *et al.*, 2005, 2007; Yu *et al.*, 2007). In fact, let-7 miRNAs function as tumor suppressors by regulating oncogenes such as MYC, RAS or HMGA2* (Roush and Slack, 2008) (See section I.15). Downregulation of let-7 was first reported in human lung cancer samples and cell lines (Takamizawa *et al.*, 2004), although is also frequent in many other tumor types and correlates with poor prognosis (Nair, Maeda and Ioannidis, 2012). Interestingly, let-7 also down-regulates Dicer, a crucial factor involved in miRNA production (see section I.10), suggesting that global miRNA biogenesis might be controlled by let-7 miRNAs (Tokumaru *et al.*, 2008). Finally, let-7 miRNAs' role in the control of immune response has positioned them as potential targets for immunotherapy (reviewed in (Gilles and Slack, 2018)).

Regulation of let-7

Even though pri-let-7 is transcribed in nearly all cellular contexts, mature let-7 is not produced in stem cells and is downregulated in cancer cells, precisely where its main regulators, the LIN28 proteins, are expressed (Thomson *et al.*, 2006; Viswanathan and Daley, 2010; Balzeau *et al.*, 2017). LIN28A and B

* High-mobility group AT-hook 2

selectively bind to pre-let-7 in the cytoplasm or pri-let-7 in the nucleus, respectively, and induce its post-transcriptional repression (Piskounova *et al.*, 2011). Interestingly, one particular human let-7 (let-7-a3) has been reported to escape LIN28-mediated repression, because pri- and pre-let-7-a3 lack the specific sequence motifs recognized by the zinc knuckle (ZND) and cold shock (CSD) domains of LIN28 (Triboulet, Pirouz and Gregory, 2015; Ustianenko *et al.*, 2018). Nevertheless, minimal but detectable expression of mature let-7 was reported recently in cultured human embryonic stem cells, suggesting that the LIN28-let-7 axis regulation might be even more complex than estimated (Rahkonen *et al.*, 2016).

On a different note, other RNAs such as lncRNAs can act as 'sponges' and compete with the endogenous targets for the let-7 miRNAs, like the case of H19 lncRNA (Kallen *et al.*, 2013).

I.15. MicroRNAs in cancer

MicroRNAs are involved in almost all physiological and developmental processes in humans, but also in pathological contexts such as cancer (Kloosterman and Plasterk, 2006; Lee and Dutta, 2009; Ventura and Jacks, 2009; Lujambio and Lowe, 2012; Di Leva, Garofalo and Croce, 2014; Slack and Chinnaiyan, 2019). In many types of tumors, there is a clear correlation between miRNA expression and prognosis or patient survival (Yanaihara *et al.*, 2006; Nair, Maeda and Ioannidis, 2012). Furthermore, miRNAs have also been suggested to play a fundamental role in EMT* and metastases (Ma, Teruya-Feldstein and Weinberg, 2007; Nicoloso *et al.*, 2009; Ventura and Jacks, 2009). Two main groups of miRNAs can be distinguished regarding their contribution to tumor progression, and their dysregulation may be caused by different factors. These subjects will be

* Epithelial-to-mesenchymal transition

discussed in the following sections. Importantly, miRNAs, together with many other noncoding RNAs, establish complex regulatory networks whose role in the development and progression of cancer is just starting to be disentangled (Anastasiadou, Jacob and Slack, 2018).

I.15.A. Tumor-suppressor miRNAs and oncomiRs

In tumoral contexts, similarly to protein-coding genes, some miRNAs enhance while others restrict tumor growth. In this regard, miRNAs can be classified in two groups: tumor-suppressor miRNAs, that repress cell division and are frequently downregulated in cancer, and oncomiRs, that promote cell division and are often upregulated in cancer (Esquela-Kerscher and Slack, 2006; Ventura and Jacks, 2009; Lujambio and Lowe, 2012) **(Figure 7)**.

Notably, the let-7 and the miR-34 families are arguably the most representative tumor-suppressor miRNAs (Lujambio and Lowe, 2012; Inamura and Ishikawa, 2016). In the case of let-7, as it has been discussed above, it has been shown to regulate oncogenes such as Ras (Johnson *et al.*, 2005) **(Figure 7A)** or HMGA2 (Lee and Dutta, 2007). On the other hand, miR-34 (whose expression

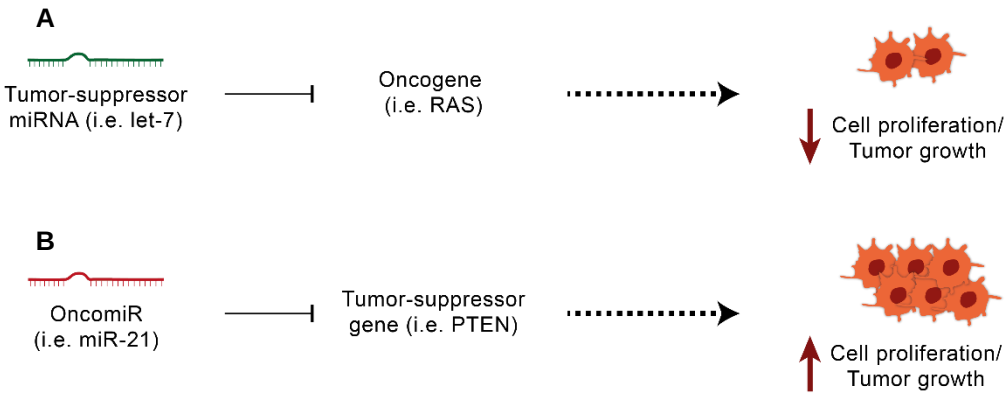


Figure 7. Effect of tumor-suppressor miRNAs and oncomiRs in tumor growth. (A) Tumor suppressor miRNAs, such as let-7, control the expression of oncogenes such as Ras, which leads to a reduction in the proliferation of the tumor. Tumor-suppressor miRNAs are frequently downregulated in cancer. **(B)** OncomiRs, such as miR-21, regulate tumor-suppressor genes such as PTEN, causing an increase in cell division and tumor growth. OncomiRs are usually upregulated in cancer.

is activated by the tumor-suppressor p53) regulates genes involved in cell proliferation such as CCNE2* or CDK4† (Chang *et al.*, 2007; He *et al.*, 2007; Tarasov *et al.*, 2007). Remarkably, miR-34a restoration has been shown to reduce tumor growth in triple-negative breast cancer cell lines (Adams *et al.*, 2016).

Some miRNAs are upregulated in tumors, as was first observed in an analysis of 540 samples from 6 different cancer types (Volinia *et al.*, 2006). These are called oncomiRs, and function similarly to protein-coding oncogenes such as Ras or Myc. (Esquela-Kerscher and Slack, 2006). One representative example is miR-21, which promotes Ras signalling (and thus cell proliferation) by targeting its negative regulators (Hatley *et al.*, 2010; Medina, Nolde and Slack, 2010) and other tumor-suppressor genes like PTEN **(Figure 7B)**. In fact, addiction of tumors to this oncomiR, and regression of the tumors upon removal of miR-21 has been reported in mouse models of pre-B-cell lymphoma (Medina, Nolde and Slack, 2010). Another important example is the mir-17–92 cluster, which includes seven miRNAs: miR-17-5p, miR-17-3p, miR-18a, miR-19a, miR-20a, miR-19b-1 and miR-92-1 (He *et al.*, 2005). Transcription of the miR-17-92 cluster in cancer is activated by MYC (O'Donnell *et al.*, 2005), and levels of these miRNAs are increased in different types of cancer (Esquela-Kerscher and Slack, 2006).

miR-20

Several studies point towards an important role for a particular miRNA within the miR-17-92 cluster in different types of human cancer: miR-20. Recently, the aberrant upregulation of miR-20a has been reported to promote colorectal cancer progression and metastasis via regulation of WTX‡ (Zhu *et al.*, 2019). Besides,

* Cyclin E2
† Cyclin-dependent kinase 4
‡ Wilms tumor gene on the X chromosome

increased miR-20 expression levels have been reported in cervical cancer patients, and depletion of miR-20 reduced tumor growth in a xenograph* model of human cervical cancer (Zhao *et al.*, 2015). Consistently, a recent meta-analysis suggest that miR-20 could be used as a biomarker for prognosis of gastrointestinal cancer patients (Huang *et al.*, 2018). However, some contradictory evidence exists for the oncogenic role of miR-20. Apart from the aforementioned examples, some reports suggest that this miRNA can also act as a tumor-suppressor by restricting cell proliferation. In fact, miR-20 regulates the transcription factor E2F1 which is required for cell progression through the cell cycle (O'Donnell *et al.*, 2005). Interestingly, mouse models suggest that miR-20a/b prevent cell proliferation and chemoresistance of breast cancer cells *in* vivo by regulating the cancer-related MAPK† signalling pathway (Si *et al.*, 2018). This apparently contradictory evidence suggest that miR-20 may work as either an oncomiR or a tumor-suppressor miRNA depending on the type of tumor.

I.15.B. Causes of miRNA dysregulation in cancer

Contrary to transposable elements, microRNAs are globally downregulated in cancer (Lu *et al.*, 2005). In fact, global miRNA depletion through genetic deletion of components of the miRNA processing machinery promotes tumorigenesis (Kumar *et al.*, 2009). One frequent cause is the deletion of miRNA genes by processes related to genome instability. In fact, the earliest evidence suggesting that miRNAs could be involved in human cancer was reported almost 20 years ago, when miR-15 and miR-16 loci were found to be deleted in more than two thirds of B cell chronic lymphocytic leukaemia (B-CLL) cases (Calin *et al.*, 2002).

* A model in which human tumor cells are transplanted into immunocompromised mice.
† Mitogen-Activated Protein Kinase

Consistently, a high-resolution aCGH* in samples from human ovarian, breast and melanoma tumors revealed that a high proportion of genomic loci containing miRNA genes exhibited DNA copy number alterations (Zhang *et al.*, 2006). Thus, deletion or amplification of miRNA genes is frequent in many types of cancer (Croce, 2009). Additionally, since miRNA promoters are subjected to a tight regulation, dysregulation of key transcription factors in cancer may as well alter miRNA expression. A relevant example is c-Myc, frequently upregulated in cancer, which activates the transcription of the oncogenic miR-17-92 cluster (O'Donnell *et al.*, 2005) and also represses transcription of tumor-suppressor miRNAs such as miR-15, miR-16 or the let-7 family (Chang *et al.*, 2008). Another important case is p53, frequently mutated and downregulated in human cancer, which activates tumor-suppressor miR-34 transcription to induce cell cycle arrest and apoptosis (He *et al.*, 2007).

Multiple proteins are involved in miRNA biogenesis, therefore any mutations or aberrant expression of them could lead to altered miRNA expression. In fact, many cancer-associated mutations occur in proteins involved in miRNA biogenesis (reviewed in (Lin and Gregory, 2015)). Notably, low expression levels of Microprocessor components Drosha and Dicer correlate with poor clinical outcome in several cancer types such as ovarian cancer, lung cancer and neuroblastoma (Karube *et al.*, 2005; Merritt *et al.*, 2008; Lin *et al.*, 2010). Furthermore, mutations in the exportin 5 (XPO5) gene can trap pre-miRNAs in the nucleus, and its restoration has tumor-suppressor features (Melo *et al.*, 2010).

Lastly, alterations in the target mRNAs can also cause misfunction of miRNAs. For example, a SNP† in the 3'UTR of the KRAS gene is located within a

* Array-based Comparative Genomic Hybridization: a molecular cytogenetic technique used to detect genome-wide chromosomal copy number changes.
† Single nucleotide polymorphism

let-7 binding site. One of the SNP variants results in higher KRAS expression due to reduced let-7 binding, suggesting that patients with this variant could be more susceptible to develop a KRAS related tumor (Chin *et al.*, 2008). Besides, in cancer cells some oncogenes present a shortening in their 3'UTRs caused by alternative cleavage and polyadenylation. Since miRNA binding sites are frequently located in the 3'UTR of mRNAs, these oncogenes can become refractory to miRNA regulation, causing an increase in cell proliferation (Mayr and Bartel, 2009).

I.16. Can miRNAs regulate active retrotransposons?

The intermediate RNA of endogenous retrotransposons is a target for multiple restriction mechanisms that limit their mobilization in the genome (see section I.6B). However, most of the mechanisms described so far involve proteins and not regulatory RNAs. Interestingly, it has been previously shown that mouse embryonic stem cells (mESCs) lacking mature miRNAs (DGCR8 or Dicer knockout) accumulate LINE-1 mRNAs (Heras *et al.*, 2013; Bodak *et al.*, 2017). While the increase in LINE-1 mRNA levels in the absence of DGCR8 was attributed to reduced non-canonical functions of the Microprocessor complex, which cleaved stem-loops present in L1 elements (Heras *et al.*, 2013), it remains possible that miRNAs regulate L1 expression levels. Additionally, many studies have shown that miRNAs, and particularly tumor-suppressor miRNAs, are downregulated in many types of cancer, where increased L1 retrotransposition has also been reported (Lujambio and Lowe, 2012; Lin and Gregory, 2015; Burns, 2017; Scott and Devine, 2017). Thus, it is reasonable to hypothesize that some miRNAs could control L1 retrotransposition and, consequently, that misexpression of miRNAs in tumors could contribute to increased LINE-1 mobilization. Indeed, a previous study demonstrated that miR-128 represses engineered L1 retrotransposition in cultured cancer cells (Hamdorf *et al.*, 2015).

OBJECTIVES

Given their mutagenic potential, several mechanisms exist in humans to control the mobilization of L1 retrotransposons in somatic cells. However, these mechanisms may be altered in cancer, leading to the increased retrotransposition that has been reported in many types of tumors.

The starting hypothesis of this Thesis is that microRNAs could regulate L1 expression, and therefore miRNA downregulation occurring in cancer could lead to increased L1 activity and, consequently, to genomic instability. The broad objective is to discover miRNAs that regulate L1 retrotransposition, whose downregulation could lead to the increased L1 mobilization observed in cancer. This could be divided in three discrete objectives:

1) Identification of miRNAs that are downregulated in tumor samples with increased somatic L1 activity using high-throughput sequencing data.
2) Characterization of the molecular mechanisms by which these miRNAs regulate L1 retrotransposition.
3) Development of a technique to unbiasedly identify miRNAs targeting L1 mRNA.

RESULTS

R.1. Downregulation of let-7 and miRNA-34a correlates with increased somatic L1 retrotransposition in human lung tumor samples

This section was performed in collaboration with bioinformaticians Guillermo Peris and Alejandro Rubio (Genyo, Granada).

To identify potential miRNAs whose deregulation could produce a change in L1 retrotransposition in epithelial tumors, we focused on Non-Small Cell Lung Cancer samples (NSCLC) where endogenous L1s are known to retrotranspose frequently (Iskow *et al.*, 2010; Helman *et al.*, 2014; Tubio *et al.*, 2014; Scott and Devine, 2017; Rodriguez-Martin *et al.*, 2020). We obtained 45 patient samples from The Cancer Genome Atlas (TCGA) for which the following was available: (i) whole genome sequencing data from tumor and matched normal lung tissue and (ii) miRNA-seq data from tumor. We computationally identified somatic L1 insertions computationally from whole-genome sequencing data using the MELT* software (Gardner *et al.*, 2017). Briefly, to detect mobile element insertions MELT searches for discordant read pairs and split reads that are enriched at genome positions containing new, non-referenced insertions (Gardner et al., 2017). First, to rule out possible biases produced by different coverage or quality of sample pairs, we analyzed the polymorphic germline L1 insertions identified by MELT. In 41/45 samples, the number of polymorphic L1 insertions found in tumor/normal DNA pairs was similar and at least 63% of them were common to both DNAs **(Table 3)**. The remaining 4 samples were discarded. After exclusion of polymorphic L1s present in the euL1db (Mir, Philippe and Cristofari, 2015), we detected 413 putative tumor-specific L1 insertions, which were absent

* Mobile Element Locator Tool

	Types of tumor	Coverages		Unfiltered Poly L1			MELT insertion calls					
Sample		PT	NT	PT	NT	Common	PT	NT	NT-PT	Poly PT	Poly NT	Poly Common
TCGA-18-3408	LUSC	70.3	34.5	138	136	98	9	0	0	48	43	43
TCGA-18-3415	LUSC	76.0	38.0	141	134	95	26	0	0	35	34	31
TCGA-18-4721	LUSC	69.5	44.6	138	141	102	16	0	1	56	57	54
TCGA-21-1076	LUSC	54.0	38.3	133	126	93	5	0	0	29	28	27
TCGA-21-1078	LUSC	44.4	45.3	118	120	78	0	1	0	14	15	14
TCGA-21-1083	LUSC	89.3	65.4	129	130	105	29	0	1	50	50	48
TCGA-22-5477	LUSC	56.0	44.2	126	131	94	17	0	0	45	47	42
TCGA-22-5485	LUSC	95.4	41.1	133	136	97	19	0	0	53	55	50
TCGA-22-5492	LUSC	81.0	32.8	136	140	103	20	0	0	45	47	42
TCGA-33-4586	LUSC	70.3	37.9	134	144	91	12	0	0	38	42	35
TCGA-34-2596	LUSC	51.8	45.6	213	136	84	55	0	1	37	41	32
TCGA-34-2600	LUSC	37.9	62.2	110	132	70	6	0	0	27	33	23
TCGA-43-3394	LUSC	47.2	46.4	138	147	108	5	0	0	26	30	23
TCGA-43-3920	LUSC	71.9	35.2	110	113	81	8	0	0	29	30	26
TCGA-56-1622	LUSC	44.3	41.9	111	124	78	3	0	0	17	25	16
TCGA-60-2698	LUSC	63.6	34.2	125	124	89	34	0	1	27	28	24
TCGA-60-2711	LUSC	47.1	52.9	143	137	99	5	0	0	29	28	27
TCGA-60-2713	LUSC	46.6	51.9	122	119	80	3	1	0	15	19	13
TCGA-60-2719	LUSC	62.5	29.0	125	123	90	4	0	0	11	12	9
TCGA-60-2724	LUSC	62.8	49.7	125	131	93	3	0	0	24	26	23
TCGA-66-2744	LUSC	66.3	51.2	130	128	99	4	0	0	38	36	36
TCGA-66-2759	LUSC	63.9	51.6	127	129	92	39	0	0	52	53	49
TCGA-66-2766	LUSC	46.8	44.1	133	125	82	9	0	0	15	11	11
TCGA-66-2789	LUSC	83.6	37.0	130	127	89	0	0	0	37	33	32
TCGA-66-2793	LUSC	73.9	35.4	129	143	91	54	0	0	42	43	41
TCGA-66-2795	LUSC	61.8	34.8	122	126	82	14	0	0	28	27	24
TCGA-38-4628	LUAD	84.5	34.6	125	121	75	0	0	0	13	12	10
TCGA-38-4630	LUAD	48.5	36.2	117	127	82	6	0	0	20	21	19
TCGA-49-4486	LUAD	34.7	35.3	106	113	70	0	0	0	10	11	10
TCGA-49-4510	LUAD	41.0	30.8	132	122	83	0	0	0	11	11	8
TCGA-49-4512	LUAD	55.8	40.5	125	126	92	0	0	0	31	28	27
TCGA-49-6742	LUAD	64.2	68.2	117	137	85	3	0	0	22	26	20
TCGA-50-5930	LUAD	46.3	45.6	127	122	87	0	0	0	24	23	23
TCGA-50-5932	LUAD	40.0	45.3	109	117	86	0	0	0	29	33	29
TCGA-50-6591	LUAD	45.1	51.2	114	131	80	0	0	0	22	26	21
TCGA-55-6972	LUAD	44.8	38.3	116	108	80	5	0	0	31	33	26
TCGA-55-6982	LUAD	64.8	33.6	139	138	92	0	0	0	21	16	16
TCGA-55-6984	LUAD	33.3	41.3	109	115	75	0	1	0	32	32	26
TCGA-55-6986	LUAD	39.7	33.4	131	130	89	0	0	0	20	20	17
TCGA-73-4659	LUAD	36.4	35.7	108	115	72	0	0	0	14	15	14
TCGA-91-6847	LUAD	46.2	42.6	101	125	68	0	0	1	22	30	19
TOTAL INSERTIONS							**413**	**3**	**5**	**1189**	**1230**	**1080**

Table 3. Summary of L1 insertions found by MELT. A description of the different columns follows. Sample: case submitter ID in GDC database; LUSC: Lung Squamous Carcinoma; LUAD: Lung Adenocarcinoma; Coverages: average depth of coverage in bam files; Unfiltered Poly L1: polymorphic L1 insertions found in MELT output before filtering; MELT insertion calls: L1 somatic insertion calls after filtering.
In coverages: PT (primary tumor sample coverage) and NT (normal tissue sample coverage). In Unfiltered poly L1: PT (number of primary tumor L1 polymorphic insertion calls), NT (number of normal tissue L1 polymorphic insertion calls), Common (Number of L1 polymorphic insertion calls found both in tumor and normal tissue). In MELT insertion calls: PT (Number of L1 insertion calls found in primary tumor), NT (Number of L1 insertion calls found in normal tissue), PT-NT (Number of L1 insertion calls found in primary tumor and normal tissue), Poly PT (Number of primary tumor L1 polymorphic insertion calls after filtering), Poly NT (Number of normal tissue L1 polymorphic insertion calls after filtering), Poly Common (number of L1 polymorphic insertion calls found both in tumor and normal tissue after filtering).
This table was generated by Guillermo Peris who performed MELT analysis.

in DNA from the same patient's normal lung tissue **(Table 3)**. Besides, we analyzed the number of insertions that were present in the normal but not in the

tumor tissue. This number is expected to be zero: somatic insertions present in the normal tissue should always be in the tumor as the latter is derived from the first. We found only 3 in the 41 samples, confirming the specificity of the method. Consistent with previous studies, 409 of the 413 tumor specific de novo L1 insertions identified here occurred in intronic and intergenic regions **(Table 3)**, likely representing passenger mutations (Burns, 2017; Scott and Devine, 2017).

To evaluate a possible correlation between L1 retrotransposition in lung cancer and miRNA expression, tumor samples were divided into two groups based on the presence (≥1) or absence (0) of tumor-specific L1 insertions **(Figure 8A)**. Using available miRNA-seq data from these TCGA samples, we analyzed the expression of 26 miRNAs that have been previously associated with the development and/or progression of lung cancer, such as the let-7 family, the miR-34 family, or the miR-17-92 cluster (Inamura and Ishikawa, 2016) **(Figure 8)**. Interestingly, we found that several members of the tumor suppressor let-7 family (let-7a, let-7e and let-7f) were significantly downregulated in the samples with ≥1 tumor-specific L1 insertions upon multiple t-testing adjusted with FDR*<0.01 **(Figure 8A)**. Notably, this correlation was also found for let-7a and let-7f using a different statistical analysis: rank-sum test. All members of the let-7 family have a similar mature sequence and therefore can bind, potentially, to the same RNA targets. However, their genomic location and timing of expression is markedly different (Roush and Slack, 2008). Interestingly, reduced expression of let-7a and let-7f has been previously observed in lung cancer samples (Takamizawa *et al.*, 2004; Yanaihara *et al.*, 2006). Additionally, miR-34a, another tumor suppressor miRNA (He *et al.*, 2007), was also significantly reduced in samples with tumor-specific L1 insertions **(Figure 8A)**. Notably, the differential expression of let-7a,

* False Discovery Rate

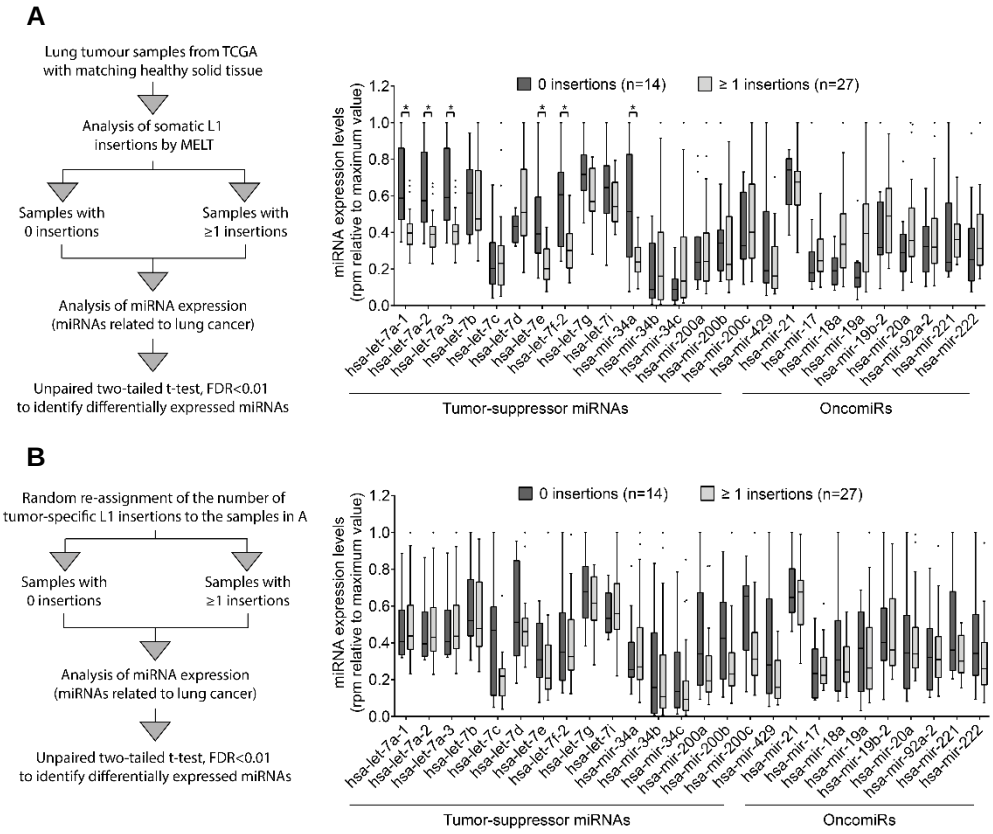


Figure 8. Downregulation of let-7 and miR-34a miRNAs correlates with accumulation of tumor-specific L1 insertions in lung tumor samples. (A) Left panel: Rationale of the bioinformatic analysis used to identify differentially expressed miRNAs in lung cancer samples with or without tumor specific L1 insertions Details are provided in the text. Right panel: A graph plot representing the expression levels of miRNAs previously associated with lung cancer in lung tumor samples without (dark grey, N=14) and with (light grey, N=27) tumor specific L1 insertions identified by MELT. Differentially expressed miRNAs are marked with * and were identified applying an unpaired two-tailed t test adjusted by FDR<0.01. To enable representation of all miRNAs in one graph, expression in reads per million (rpm) was relative to the maximum value of each miRNA in each case. Whiskers were calculated using the Tukey method. Individual black dots represent outliers. Boxes extend from 25th to 75th percentiles, and lines in the middle of the boxes represent the median. **(B)** Analysis was repeated after randomly re-assigning the value of tumor-specific L1 insertions to the samples, showing no significant correlation with miRNA levels.

let-7e, let-7f and miR-34a was also significant in a more restrictive analysis where all the miRNAs expressed in lung tumor samples (89 miRNAs) were considered. As a control, L1 insertion counts were randomly reassigned to each sample, and the same analysis was repeated. No significant correlation was found in any case (one example is shown in **Figure 8B**). Even though a possible bias in the analysis

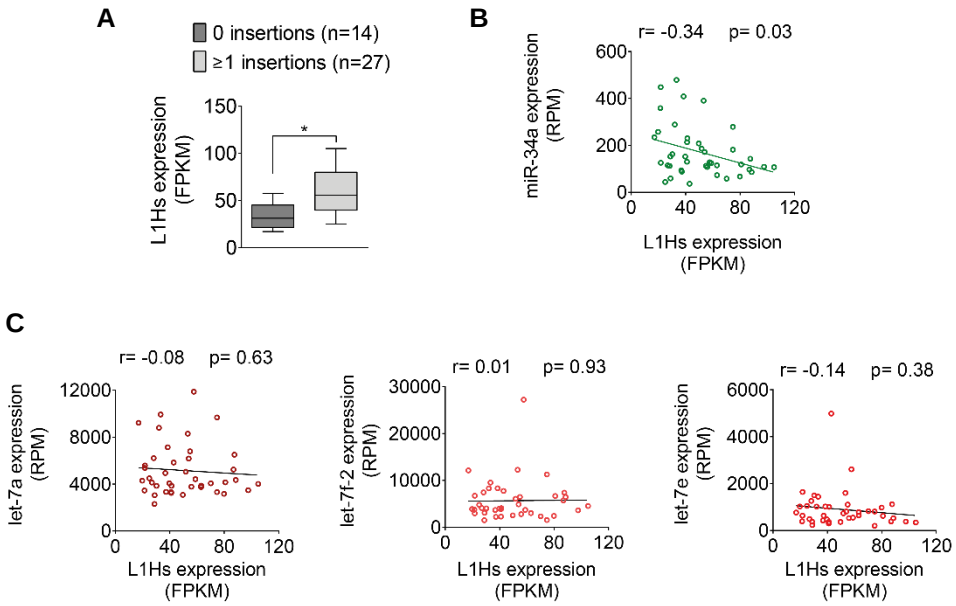


Figure 9. Increased RNA levels in samples with tumor-specific L1 insertions correlate with down-regulation of miR-34a but not let-7 miRNA levels. (A) RNA-seq analysis using SQuIRE showed that L1Hs is overexpressed in lung tumor samples with tumor-specific L1 insertions. **(B-C)** Pair-wise correlations between expression levels of L1Hs and **(B)** miR-34a or **(C)** let-7a, let-7e, let-7f-2 (Pearson' r). P-value was considered significant if <0.05.

due to sample variability and the limited number of cases available cannot be ruled out, these data suggest that let-7 and miR-34a might control the accumulation of new L1 insertions in human lung cancer samples. Next, we used SQuIRE* (Yang *et al.*, 2019) to quantify L1Hs expression in RNA-seq data from these tumor samples, available in TCGA. As expected, L1Hs RNA levels were significantly increased in samples with tumor-specific L1 insertions **(Figure 9A)**. However, L1Hs expression negatively correlates with miR-34a **(Figure 9B)** but not with let-7a/e/f expression **(Figure 9C)**.

* Software for Quantifying Interspersed Repeat Elements

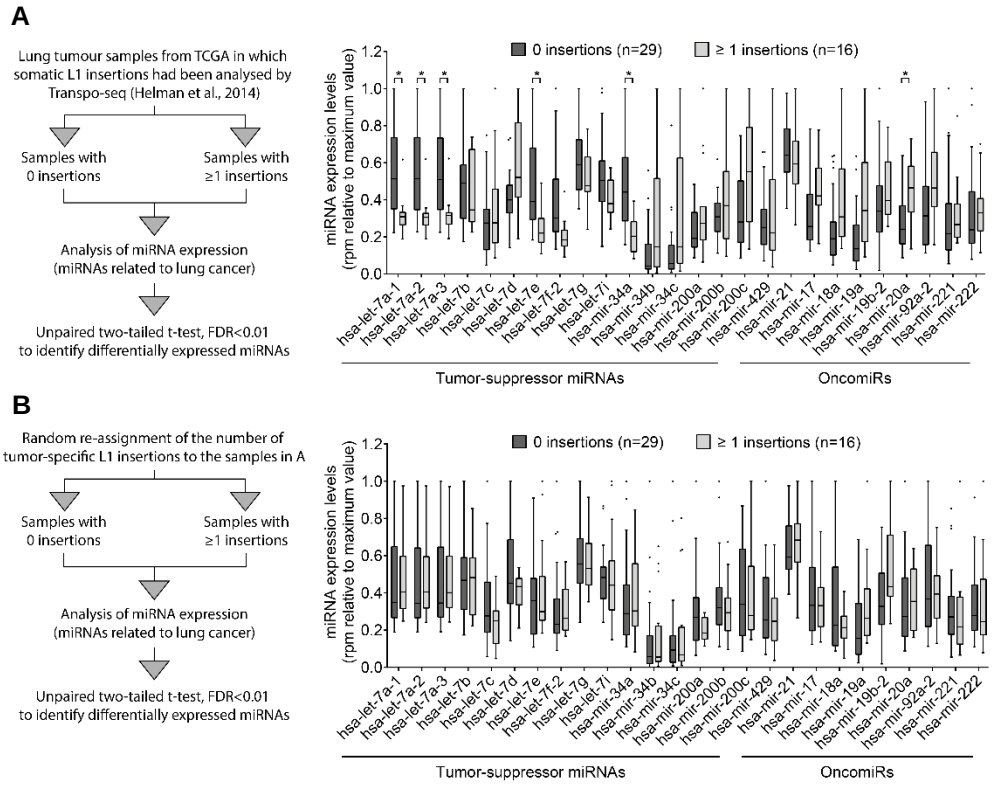


Figure 10. Downregulation of let-7 and miR-34a miRNAs correlates with accumulation of tumor-specific L1 insertions identified by Transpo-Seq in lung tumor samples. (A) Left panel: Rationale of the bioinformatic analysis used with samples analyzed by Helman and col. Correlation between miRNA expression and tumor-specific L1 insertions identified by Helman and col. using Transpo-seq was analyzed as in Figure 9. Details are provided in the text. Right panel: Correlation between miRNA expression and tumor-specific L1 insertions identified by Helman and col. using Transpo-seq was analyzed as in Figure 9. Differentially expressed miRNAs are marked with * and were identified applying an unpaired two-tailed t test adjusted by FDR<0.01. To enable representation of all miRNAs in one graph, expression in reads per million (rpm) was relative to the maximum value of each miRNA in each case. Whiskers were calculated using the Tukey method. Individual black dots represent outliers. Boxes extend from 25th to 75th percentiles, and lines in the middle of the boxes represent the median. **(B)** Analysis was repeated after randomly re-assigning the value of tumor-specific L1 insertions to the samples, showing no significant correlation with miRNA levels.

To further corroborate our results, we repeated the analysis in another set of 46 lung tumor samples in which tumor-specific L1 insertions had been identified by Helman and collaborators using a different framework: Transpo-Seq (Helman *et al.*, 2014) **(Figure 10).** Thirteen of these samples were also included in the previous analysis using MELT. Remarkably, we observed that the expression levels of let-7 family members (let-7a and let-7e) and miR-34a were significantly

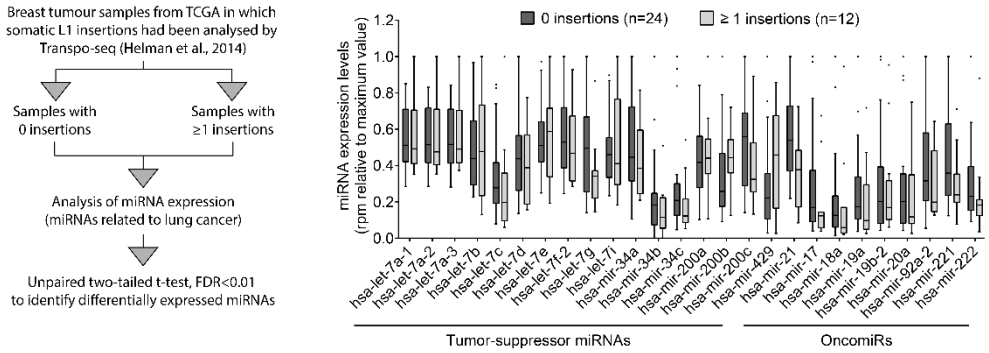


Figure 11. Absence of correlation between miRNA expression and accumulation of tumor-specific L1 insertions identified by Transpo-seq in breast cancer samples. Analysis in Figure 9 and 10 was repeated using breast samples in which tumor-specific L1 insertions had been identified by Helman and col. using Transpo-seq. A schematic representation of the bioinformatic analysis used is represented on a left panel. Differentially expressed miRNAs were identified applying an unpaired two-tailed t test adjusted by FDR<0.01. To enable representation of all miRNAs in one graph, expression (rpm) was relative to the maximum value of each miRNA in each case. Whiskers were calculated using the Tukey method. Individual black dots represent outliers. Boxes extend from 25th to 75th percentiles, and lines in the middle of the boxes represent the median.

reduced in those tumors containing tumor-specific L1 insertions identified by Transpo-Seq **(Figure 10A)**, reproducing our previous result. Interestingly, we observed another miRNA whose expression correlated with the presence of L1 insertions: miR-20a. However, in this case, expression levels of miR-20a were increased in the samples containing tumor-specific L1 insertions, contrary to what would be expected if miR-20a was a direct regulator of L1 mobilization **(Figure 10A)**. These correlations were also significant in a more restrictive analysis where all the miRNAs expressed in lung were included. Notably, the same analysis after the number of insertions had been randomly reassigned to each sample did not show any significant correlation **(Figure 10B)**. Lastly, the same analysis was performed using 36 breast cancer samples which contain a notably smaller number of tumor-specific L1 insertions per sample as determined by Transpo-Seq **(Figure 11)**. No significant correlation was found for any of the 26 miRNAs related to lung cancer **(Figure 11)** suggesting that the contribution of let-7 and mir-34a to L1 mobilization could be specific to some tumor types.

Overall, these results suggest that a downregulation of let-7 and/or miR-34 expression can influence the accumulation of tumor-specific L1 insertions in lung cancer.

R.2. Let-7 regulates human LINE-1 retrotransposition *in vitro*

To investigate whether there is a causal relationship between the accumulation of L1 insertions in tumors and the variation in let-7 and/or miR-34 expression levels, we tested the effect of these miRNAs in L1 mobilization. For that, we developed the sRNA/L1 retrotransposition assay (Tristan-Ramos *et al.*, 2020), which improves the previously described cell culture-based L1 retrotransposition assay (Moran, Holmes and Naas, 1996; Kopera *et al.*, 2016) to enable the use of microRNA mimics or inhibitors. We combined it with a panel of reporter cassettes (see Methods section for details). A scheme of the assay and the four cassettes used in this Thesis are shown in **Figure 12**. Briefly, in this assay, cells are transfected with a plasmid containing an active, full-length L1 tagged with a reporter cassette **(Figure 12)**. This cassette consists of a reporter gene (REP) in antisense orientation relative to the L1, equipped with its own promoter and polyadenylation signal, but interrupted by an intron that is in the same transcriptional orientation as the L1. Thus, a functional reporter can only be expressed after a round of successful retrotransposition that includes splicing of the intron, reverse-transcription and integration **(Figure 12)**.

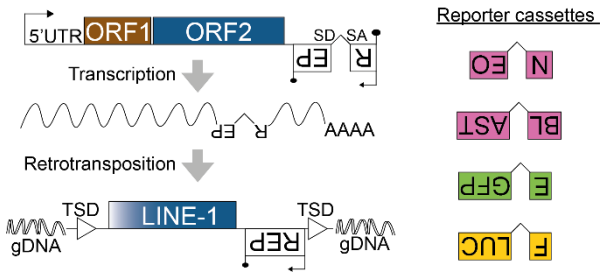


Figure 12. Rationale of the retrotransposition assay in cultured cells. Left panel, from left to right: transcription start site in the 5'UTR (black arrow), the two L1 open reading frames ORF1 (brown rectangle) and ORF2 (blue rectangle), the antisense-oriented reporter cassette (white rectangles, backward REP) interrupted by an intron, and the reporter gene promoter (inverted black arrow). Black lollipops represent poly(A) signals. TSD: Target Site Duplications. SD: Splicing Donor. SA: Splicing Acceptor. Right panel: reporter cassettes used in this study: neomycin (NEO or mneoI) and blasticidin (BLAST or mblastI) resistance, enhanced green fluorescent protein (EGFP or megfpI), and firefly luciferase (FLUC or mflucI).

As a first approach, we used cultured HeLa cells, which express high levels of let-7a and almost undetectable levels of miR-34a, as analyzed by 3'RACE RT-qPCR (**Figure 13**, see section M.15 of the methods

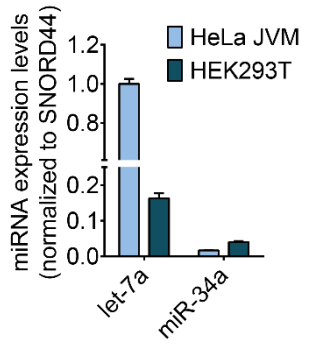
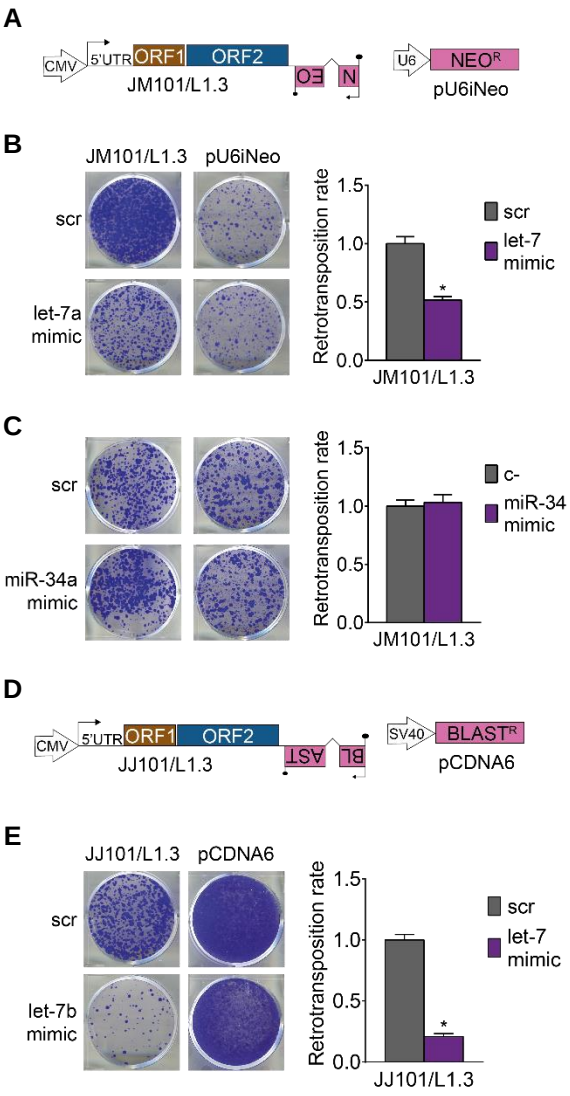


Figure 13. Quantification of mature let-7a and miR-34a levels in HeLa and HEK293T cells. 3'RACE RT-qPCR was performed to quantify mature miRNAs in HeLa and HEK293T cells. Averages of three replicates are shown, relative to the expression of let-7 in HeLa. Error bars indicate s.d.

for details). This is in agreement with previous reports suggesting that let-7 is the most expressed miRNA in HeLa cells (Khan *et al.*, 2009).

We next analyzed L1 activity upon let-7/miR34 overexpression in HeLa cells using commercially available miRNA mimics and a neomycin-resistance based retrotransposition assay (plasmid JM101/L1.3, **Figure 14A**). As a control, we performed a clonability assay co-transfecting the miRNA mimics with a plasmid encoding a constitutively expressed neomycin resistance gene (pU6iNeo) to rule out possible effects of miRNA overexpression on cell growth **(Figure 14A)**. In agreement with our observations in lung tumor samples, we reproducibly detected a significant decrease in L1 retrotransposition upon overexpression of let-7a without affecting the clonability of HeLa cells **(Figure 14B)**. Strikingly, overexpression of miR-34a did not affect L1 mobilization or cell clonability in this assay **(Figure 14C)**. Furthermore, overexpression of a different member of the let-7 family, let-7b, also reduced L1 retrotransposition using a blasticidin-resistance based assay (plasmids pJJ101/L1.3 and pCDNA6, **Figure 14D**) also reduced L1 retrotransposition **(Figure 14E)**. We next tested the effect of let-7 overexpression in a different cell line, HEK293T, with markedly different levels of mature let-7 miRNA **(Figure 13)** using a dual-luciferase reporter vector containing a different RC-L1, L1RP (plasmid pYX014, **Fig 14F**). This plasmid uses Firefly luciferase as the retrotransposition indicator and encodes a Renilla luciferase in the backbone to normalize for transfection efficiency. Notably, we observed a consistent decrease in L1 retrotransposition upon co-transfection of the let-7 mimic in HEK293T cells,



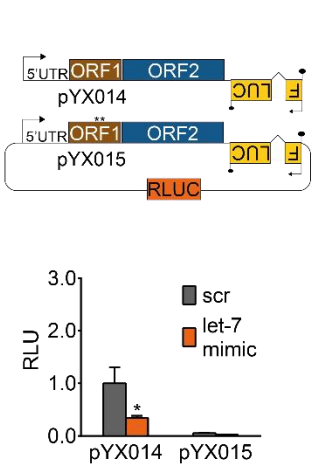


Figure 14. Let-7 overexpression decreases engineered L1 retrotransposition. (A) Structures of pJM101/L1.3 and pU6iNeo constructs are shown. **(B-C)** HeLa cells were cotransfected with one of the plasmids and let-7a/miR-34a mimic or their control (scr). **(D)** Structure of pJJ101/L1.3 and pCDNA6 are shown. **(E)** HeLa cells were cotransfected with L1 plasmid and let-7b mimic or its control (scr). **(B, C, E)** A representative well of three replicates is shown. Quantification of each experiment is shown at the right as average of three replicates. **(F)** Structures of pYX014 and pYX015 are shown. The '*' symbol in pYX015 indicates the two point-mutations in L1-ORF1p that abolish retrotransposition. **(G)** HEK293T cells were co-transfected with one of them and let-7 mimic or its control (scr). Averages of three replicates are shown. In all cases, error bars indicate s.d., and * indicates $p<0.05$ after applying a two-tailed t test. RLU: Relative Luminescence Units

measured as the relative luminescence ratio FLuc/RLuc* **(Figure 14G)**. As expected, an inactive L1RP containing two missense mutations in the ORF1-encoded protein (plasmid pYX15, **Figure 14F**) did not show luciferase activity

* Firefly Luciferase/Renilla Luciferase

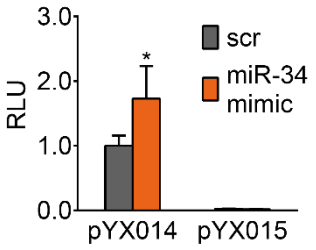


Figure 15. miR-34 overexpression increases L1 retrotransposition in HEK293T cells. HEK293T cells were cotransfected with either pYX014 or pYX015 (see Figure 7F for plasmid details) and miR-34 mimic or its control (scr). Averages of three replicates are shown. In all cases, error bars indicate s.d., and * indicates $p<0.05$ after applying a two-tailed t test. RLU: Relative Luminescence Units

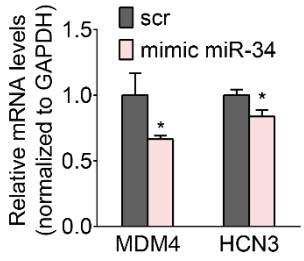


Figure 16. Mimic miR-34 transfection decreases mRNA levels of two predicted miR-34 targets. HeLa cells were transfected with mimic miR-34 or its control (scr). RT-qPCR was performed to quantify levels of MDM4 and HCN3 mRNA. Averages of three replicates are shown. Error bars indicate s.d.

(Figure 14G). Considering that miRNAs downregulate the expression of their targets, the decrease of L1 mobilization upon let-7 overexpression suggests that L1 mRNA could be a *bona fide* let-7 target. Conversely, miR-34a overexpression in HEK293T cells, where the endogenous levels are slightly higher than in HeLa **(Figure 13)** led to an increase in engineered L1 retrotransposition **(Figure 15)**. We confirmed the functionality of the mimics as we observed a decrease in the mRNA levels of two predicted miR-34 targets (MDM4* and HCN3†) by RT-qPCR **(Figure 16)**. The different effects of miR-34 overexpression observed in HeLa **(Figure 14C)** and HEK293T **(Figure 16)** cells suggest a potential indirect, cell-type specific effect of miR-34 on L1 mobilization.

To further investigate the role of let-7 on controlling L1 mobilization, we performed another panel of cell culture-based retrotransposition assays using a hairpin inhibitor to decrease intracellular let-7 levels. The inhibitor we used was designed against let-7a, however it has been shown to cross-react with other members of the let-7 family (Robertson *et al.*, 2010). Consistent with our previous results, we found that depletion of let-7 in HeLa cells caused a

* Human ortholog for Murine double minute-4
† Potassium/sodium hyperpolarization-activated cyclic nucleotide-gated channel 3

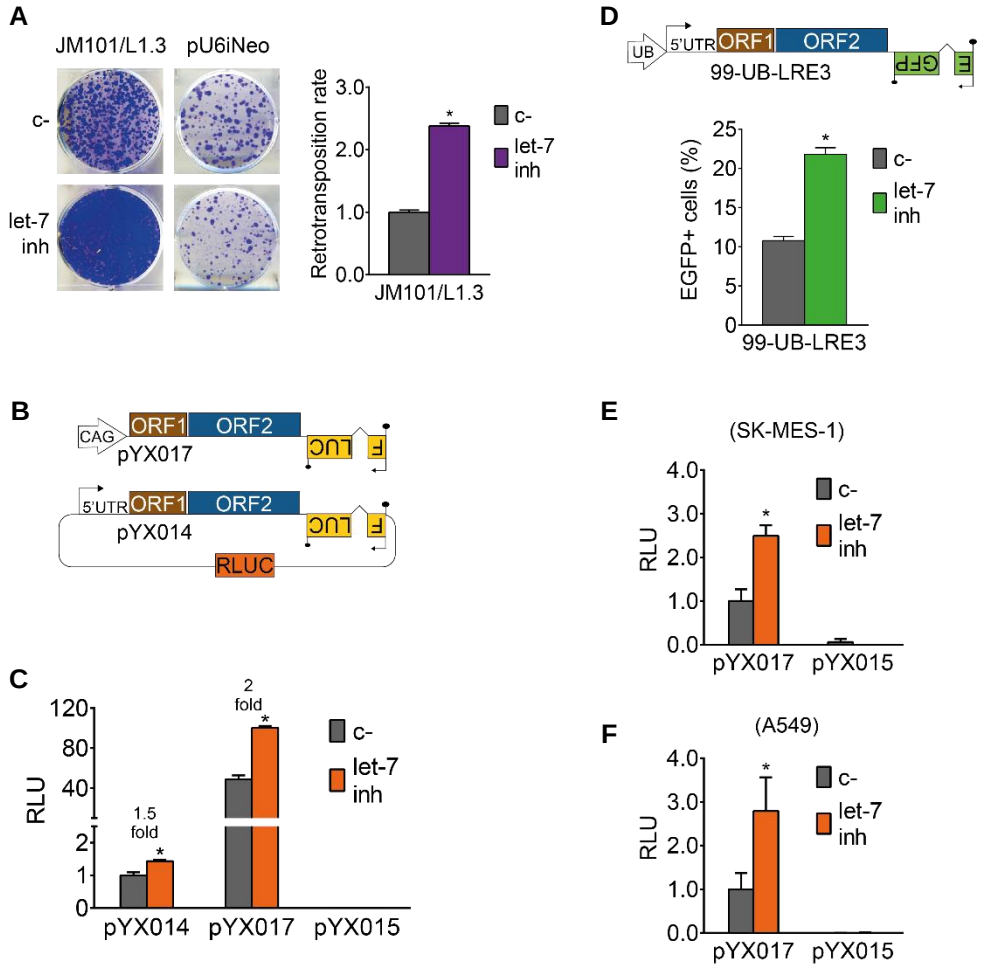


Figure 17. Let-7 inhibition increases engineered L1 retrotransposition. (A) HeLa cells were co-transfected with pJM101/L1.3 or pU6iNeo (See Figure 14A for structures) and let-7 inhibitor or its control (c-). A representative well of three replicates is shown. Quantification is shown at the right as average of three replicates. **(B)** Structures of pYX014 and pYX017 are shown (See Figure 14F for pYX015 structure). CAG: Chicken Actin Globin promoter. **(C)** HEK293T cells were cotransfected with one of the plasmids and let-7 inhibitor or its control (c-). Averages of three replicates are shown. **(D)** Structure of p99-UB-LRE3 is shown. **(E)** HEK293T cells were cotransfected with p99-UB-LRE3 and let-7 inhibitor or its control (c-). Averages of three replicates are shown. Luciferase signal is shown relative to pYX014 c-. **(F-G)** Lung cancer cell lines SK-MES-1 **(F)** or A549 **(G)** cells were cotransfected with pYX017 or pYX015 and let-7 inhibitor or its control (c-). Averages of three replicates are shown**.** In all cases, error bars indicate s.d., and * indicates $p<0.05$ after applying a two-tailed t test. RLU: Relative Luminescence Units.

two-fold increase in L1 retrotransposition without affecting the clonability of the cells using the neomycin-resistance cassette described above **(Figure 17A)**. We

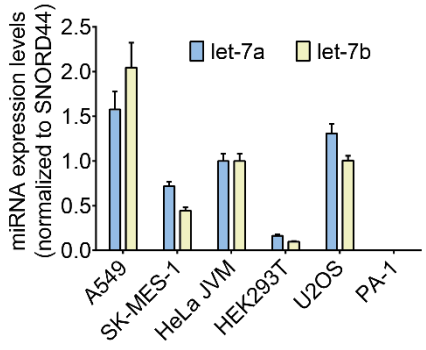


Figure 18. Quantification of mature let-7 levels in all cell lines used in this study compared to HeLa. 3'RACE RT-qPCR was performed to quantify mature let-7a and let-7b in different cell lines. Levels are shown relative to HeLa cells. Averages of three replicates are shown. Error bars indicate s.d.

reproduced that result in HEK293T cells using two different luciferase reporter vectors, pYX014 and pYX017 **(Figure 17B)**. Both contain the same active human L1, L1RP, however in pYX014 it is transcribed from the native promoter in the 5´UTR whereas in pYX017 it is highly transcribed from a CAG promoter **(Figure 17B)**. Let-7 knock-down increased L1 retrotransposition in both pYX014 and pYX017 **(Figure 17C)**. Moreover, a similar increase in L1 retrotransposition was observed in HEK293T cells upon let-7 depletion using an EGFP-based reporter cassette and a different human RC-L1, LRE3 (plasmid 99-UB-LRE3, **Figure 17D**). Lastly, since we had bioinformatically found an inverse correlation between let-7 expression and accumulation of L1 insertions in human lung tumor samples, we analyzed the effect of let-7 modulation in L1 retrotransposition in lung cancer cells. To do that, we performed the luciferase-based retrotransposition assay in two lung cancer cell lines with markedly different endogenous levels of let-7: A549 and SK-MES-1 **(Figure 18)**. Notably, we observed that depletion of let-7 increased L1 retrotransposition in both SK-MES-1 **(Figure 17E)** and A549 cells **(Figure 17F).** Altogether, these data indicate that let-7 negatively regulates human L1 mobilization in a variety of cancer cell lines.

R.3. Let-7 binds directly to the coding sequence of the L1 mRNA

The aforementioned regulation could be explained by two non-exclusive mechanisms. First, let-7 could be guiding the RISC to a direct interaction with L1 mRNAs. Second, let-7 could be regulating any host factor involved in the multiple steps of the retrotransposition cycle or in L1 control, thus regulating L1 mobilization in an indirect manner. We hypothesised that a direct effect would be sequence-dependent, therefore we performed the sRNA/L1 retrotransposition assay in HeLa cells using active LINEs from other species, which differ in sequence from the human L1 but use the same target-primed reverse transcription mechanism for mobilization. Briefly, we used mouse TGf21 ($L1G_F$ subfamily) and zebrafish L2-1 and L2-2 (L2 clade). Structures of the different LINEs are shown in **Figure 19A**, and constructs are described in section M.5 of the methods. Interestingly, we observed that only human L1 mobilization was significantly affected by either the inhibition **(Figure 19B)** or the overexpression **(Figure 19C)** of let-7. These results suggested a direct, sequence-dependent interaction between let-7 and human L1 mRNA.

During the last 20 years it has been shown that miRNAs predominantly bind to the 3'UTR of their target mRNAs, although 5'UTR and coding sequence binding sites have been described and validated (Brümmer and Hausser, 2014; Hausser and Zavolan, 2014; Zhang *et al.*, 2018). A notable example is DICER, which is targeted by let-7 in its coding sequence (Forman, Legesse-Miller and Coller, 2008). Thus, to find out where the putative let-7 binding site could be located in L1 mRNA, we performed blasticidin-resistance based retrotransposition assays using engineered human RC-L1s (L1.3) lacking either the 5´or the 3´UTR **(Figure 20A)**. Notably, the effect of let-7 depletion **(Figure 20B)** or overexpression **(Figure 20C)** in engineered L1 mobilization was not abolished

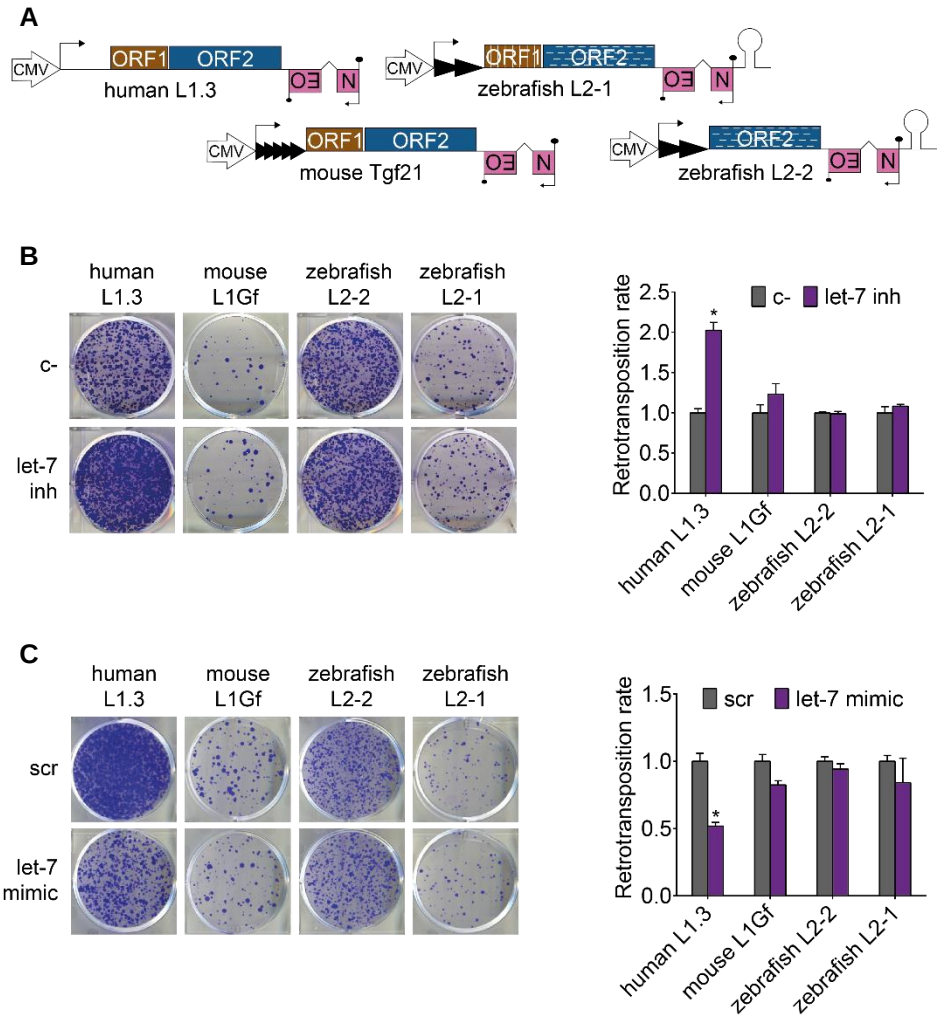


Figure 19. Human L1, but not mouse or zebrafish LINEs, is affected by let-7 depletion or over-expression. (A) Structures of the different LINEs are shown. All constructs have an exogenous CMV promoter to normalize transcription. Black arrows represent transcription start sites. Grey triangles in mouse and zebrafish LINEs illustrate the presence of monomers in the 5'UTR of these elements. Stem loop (grey) pictures the hairpin structures present in the 3' UTR of the zebrafish LINE-2s and required for retrotransposition. White stripes are included to remark the differences in sequence of zebrafish L2-2 and L2-1 with respect to the human L1.3 and mouse L1GF. **(B-C)** HeLa cells were co-transfected with LINEs from different species and **(B)** let-7 inhibitor or its control (c-), or **(C)** let-7 mimic or its control (scr). In both cases, a representative well of three biological replicates is shown. Quantification is shown on the right as average ± s.d.,'*' indicates $p<0.05$.

or reduced by the absence of either 5' or 3' UTR, suggesting that let-7 interacts with the coding sequence of human L1 mRNA **(Figure 20B-C)**.

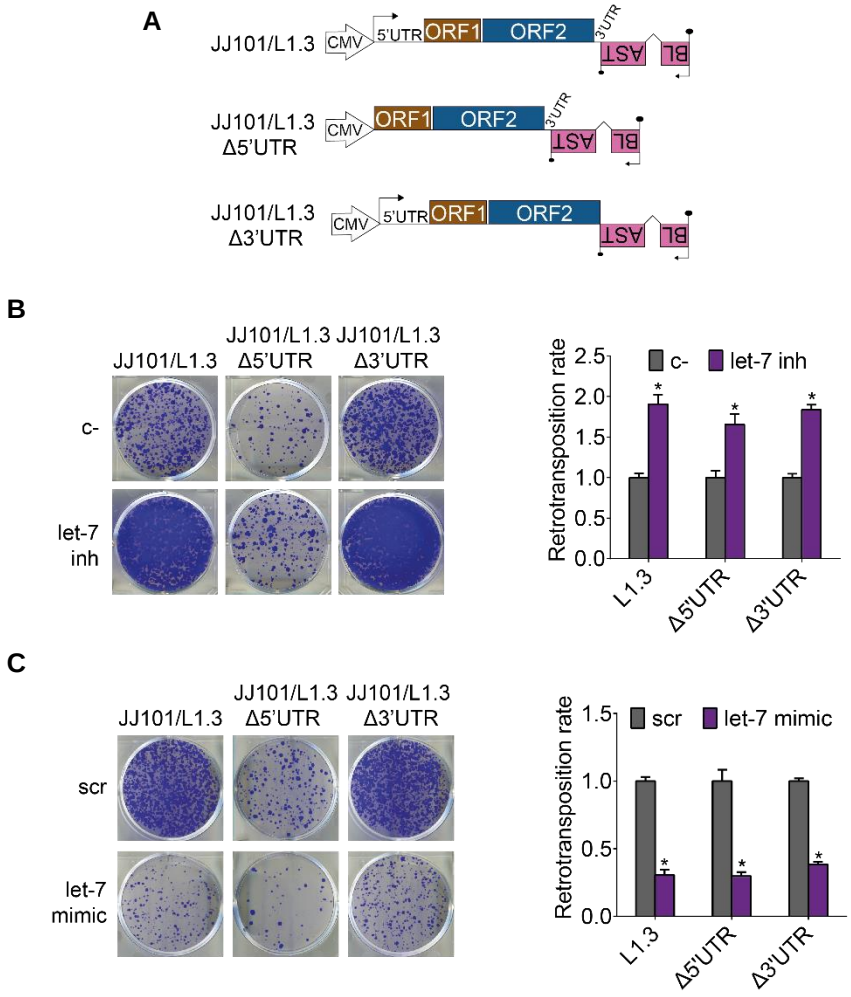


Figure 20. Absence of 3' or 5' UTR does not abolish the effect of let-7 in engineered L1 retrotransposition. (A) Structures of the different L1s lacking either 5' or 3' UTR are shown. **(B-C)** Cell culture-based retrotransposition assay with blasticidicin resistance cassette. HeLa cells were co-transfected with the L1s described above and **(B)** let-7 inhibitor or its control (c-), or **(C)** let-7 mimic or its control (scr). In both cases, a representative well of three biological replicates is shown. Quantification is shown on the right as average ± s.d.,'*' indicates $p<0.05$.

It has previously been described that L1-ORF1p often aggregates in cytoplasmic foci and colocalizes with L1 mRNA and AGO2 protein, the main component of the RISC complex (Goodier et al., 2007; Doucet et al., 2010). Thus, to further analyze whether let-7 guides the RISC complex to L1 mRNAs, we performed an RNA-Immunoprecipitation (RIP) assay. In this experiment, we used

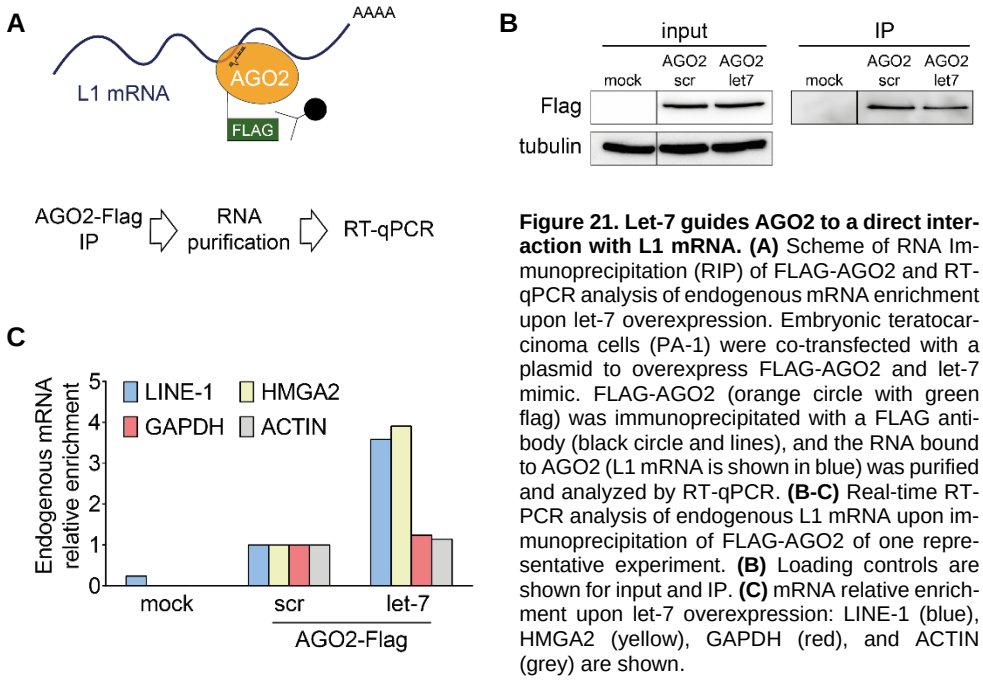


Figure 21. Let-7 guides AGO2 to a direct interaction with L1 mRNA. (A) Scheme of RNA Immunoprecipitation (RIP) of FLAG-AGO2 and RT-qPCR analysis of endogenous mRNA enrichment upon let-7 overexpression. Embryonic teratocarcinoma cells (PA-1) were co-transfected with a plasmid to overexpress FLAG-AGO2 and let-7 mimic. FLAG-AGO2 (orange circle with green flag) was immunoprecipitated with a FLAG antibody (black circle and lines), and the RNA bound to AGO2 (L1 mRNA is shown in blue) was purified and analyzed by RT-qPCR. **(B-C)** Real-time RT-PCR analysis of endogenous L1 mRNA upon immunoprecipitation of FLAG-AGO2 of one representative experiment. **(B)** Loading controls are shown for input and IP. **(C)** mRNA relative enrichment upon let-7 overexpression: LINE-1 (blue), HMGA2 (yellow), GAPDH (red), and ACTIN (grey) are shown.

a human embryonic teratocarcinoma cell line (PA-1), characterized by high levels of endogenous LINE-1 mRNA and L1-ORF1p (Garcia-Perez *et al.*, 2010) and very low levels of let-7 miRNAs **(Figure 18)**. Briefly, we overexpressed FLAG-tagged AGO2, pulled it down, purified the endogenous bound RNAs, and analyzed them by RT-qPCR **(Figure 21A)**. We reasoned that, if let-7 can bind L1 mRNA, let-7 overexpression should lead to an increase in the abundance of endogenous L1 mRNAs associated to AGO2. Strikingly, with similar enrichment in the IP **(Figure 21B)**, we observed a 3-fold enrichment in the amount of L1 mRNA bound to AGO2 upon overexpression of let-7 **(Figure 21C)**, resembling the behaviour of HMGA2 mRNA, a well-known target of let-7 (Lee and Dutta, 2007). In contrast, none of the negative controls used, GAPDH and actin mRNAs, were enriched in the immunoprecipitation **(Figure 21C)**. Thus, these data suggest that let-7 guides AGO proteins to L1 mRNA through interaction with its coding sequence.

R.4. A functional let-7 binding site is located in L1-ORF2

Next, we set out to predict and validate putative let-7 binding sites within the coding sequence of the L1 mRNA. For that, we used two different softwares: miRanda (John *et al.*, 2004) and RNA22 (Miranda *et al.*, 2006). The best predicted binding site (bs) found by each method were located, in both cases, in the ORF2 region of the consensus L1Hs sequence: positions 2650-2671 (bs (1), found by miRanda) and 4587-4608 (bs (2), found by RNA22) **(Figure 22A)**.

In order to validate them, five tandem copies of each binding site were cloned in the 3'UTR region of the Renilla luciferase (RLuc) gene in the psiCHECK-2 vector **(Figure 22B)**. This plasmid also encodes a Firefly luciferase (FLuc) gene to correct for transfection efficiency **(Figure 22B)**. Briefly, if the sequence cloned in the 3'UTR of the RLuc gene contains a let-7 binding site, we would expect a decrease in luciferase activity (measured as the ratio RLuc/FLuc) upon let-7 overexpression. As controls, an unrelated sequence of the same length ('no bs') and a sequence with perfect complementarity to let-7 ('perfect bs') were cloned. The four constructs (containing either one of the binding site candidates or one of the two controls) were co-transfected with let-7 mimic or its control (scr) in HEK293T cells. The reporter constructs containing the RNA22-predicted binding site (bs (2)) and the positive control (perfect bs), but not the one containing the miRanda-predicted binding site (bs (1)), showed a reduction of the relative luciferase ratio (RLuc/FLuc) upon let-7 overexpression, suggesting that 'bs (2)' is a *bona fide* let-7 binding site **(Figure 22C)**. Deeper analysis of the interaction of 'bs (2)' with let-7 using the RNAHybrid software (Rehmsmeier *et al.*, 2004) suggest that the functional site is located within the coding sequence of L1-ORF2 (position 4587-4610 in L1.3, a commonly used human RC-L1, accession code # L19088.1), between the RT and Cysteine-rich domains of this protein **(Figure 22D)**. Interestingly, it is predicted to form a duplex with let-7 miRNA consisting of seven Watson-Crick pairings at positions 3 to 9, preceded

A

Predicted binding site	Location (Pos in L1Hs)	Prediction method/software	dG	Validated by siCHECK?
bs (1)	ORF2 (2650-2671)	mIRanda	-17.5	NO
bs (2)	ORF2 (4587-4608)	RNA22	-19.4	**YES**

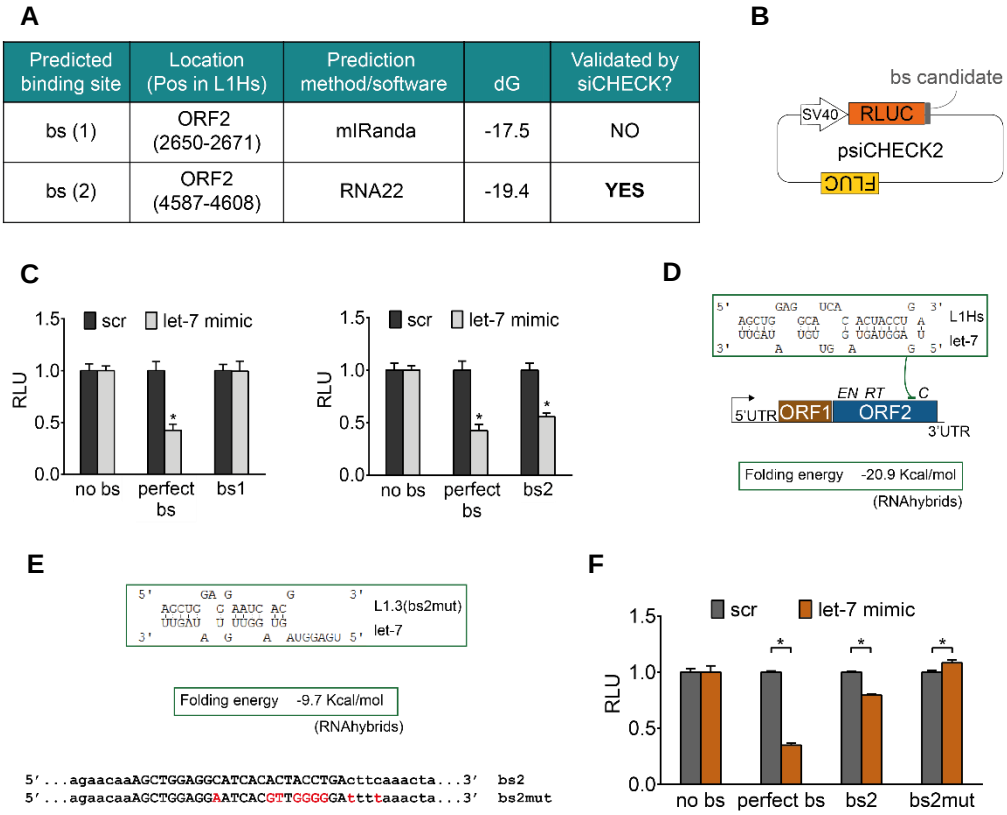


Figure 22. Validation of a predicted binding site for let-7 in the ORF2 region of L1 mRNA. (A) For each binding site, position within the consensus sequence of L1Hs, method used for its prediction and dG of its predictive binding to L1.3 are shown. **(B)** Scheme of psiCHECK2 plasmid and its main features: an SV40 promoter (white arrow), Renilla luciferase gene (RLuc, orange rectangle), the sequence we test to study whether it contains a miRNA binding sited (bs candidate, grey rectangle), and Firefly luciferase gene (FLuc, yellow rectangle) **(C)** Results of the psiCHECK2 assays with each predicted binding site. HEK293T cells were co-transfected with three different psiCHECK2 constructs and let-7 mimic (grey bars) or its control (scr, black bars). 'no bs' is a negative control (a sequence without complementarity to let-7 was cloned in the 3' UTR of the luciferase) and 'perfect bs' is a positive control (a sequence with perfect complementarity to let-7 was cloned). RLU: Relative Luciferase Units. Error bars indicate s.d. of three replicates. * indicates p<0.05 after t-testing. **(D)** RNAhybrid prediction of the validated binding site located in the L1 coding sequence. Base-pairing between this region and let-7 is shown (green rectangle). Localization of the putative binding site ('bs2') within L1 sequence is shown (green line). Structure of LINE-1 is shown: transcription start site (black arrow), 5' untranslated region (UTR), ORF1, ORF2 with its three domains endonuclease (EN), reverse transcriptase (RT), and cysteine-rich (C), and 3' UTR. Folding energy and p value of the predicted binding site are shown below. **(E)** RNAHybrid prediction of mutated bs2 site ('bs2mut'). Base pairing prediction is shown at the top, folding energy in the middle, and comparison of bs2 and bs2mut is shown below, with the nucleotides interacting with let-7 in capital letters and the different nucleotides highlighted in red. **(F)** HEK293T cells were co-transfected with the different psiCHECK2 constructs and scr (grey bars) or let-7 mimic (brown bars). Error bars represent s.d. of three replicates. RLU: Relative Luciferase Units. * indicates p<0.05 upon two-tailed t-testing.

by an adenosine at the mRNA nucleotide that pairs with the first position of the miRNA, resembling a previously described, functional, noncanonical binding site

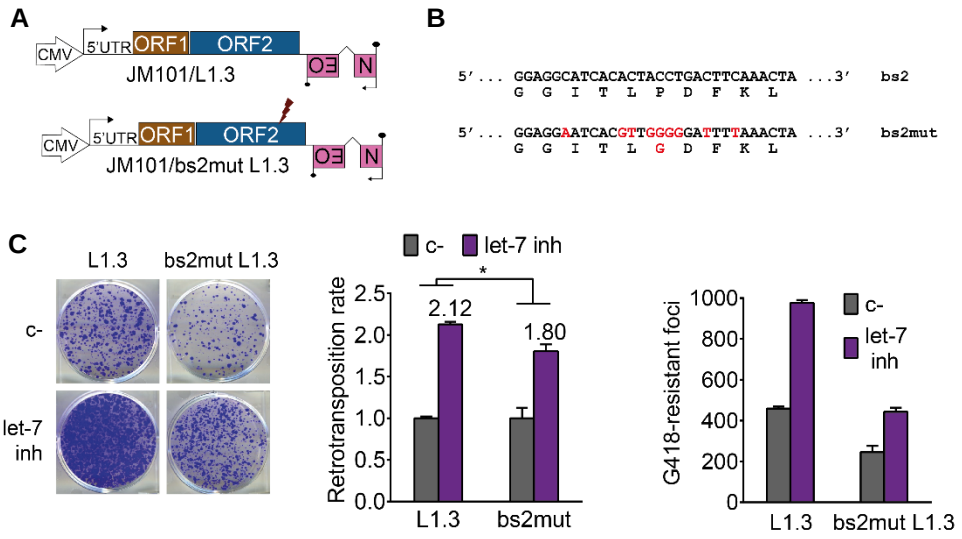


Figure 23. Mutations in binding site 'bs2' reduce the effect of let-7 depletion in L1 retrotransposition. (A) Schemes comparing the wild type L1.3 and the binding site mutant 'bs2mut L1.3' generated by site-directed mutagenesis upon cloning in JM101 vector. Red thunder indicates approximate location of the mutated binding site. **(B)** Sequence comparison between bs2 and bs2mut, showing in red the aminoacid change P to G caused by the mutations introduced. **(C)** HeLa cells were co-transfected either JM101/L1.3 or JM101/bs2mut L1.3 and let-7 inhibitor or its control (c-). A representative well of three replicates is shown in the left. Quantification is shown as retrotransposition rate (relative to c-) in the middle panel and as raw colony count in the right panel. in both cases as average of three replicates ± s.d. * indicates $p<0.05$ after a two-tailed t-test.

termed offset 7-mer (Kim *et al.*, 2016) **(Figure 22D)**. These results suggest that this refined, hereafter referred to as 'bs2', is a *bona fide* let-7 binding site. To further validate this binding site, we generated a mutant sequence to impede the predicted pairing between the mRNA and let-7 seed and therefore prevent duplex formation: 'bs2mut' **(Figure 22E)**. We cloned five tandem copies of 'bs2mut' in psiCHECK-2 vector and co-transfected it, as well as the other psiCHECK-2 constructions containing 'bs2' and the controls 'no bs' and 'perfect bs', with let-7 mimic or its control (scr) in HeLa JVM cells. As expected, the mutated sequence rescued the luciferase activity upon overexpression of let-7 **(**bs2 vs bs2mut, **Figure 22F)**.

To further corroborate the functionality of 'bs2' in the context of retrotransposition, we used site-directed mutagenesis to generate a mutated L1.3 containing

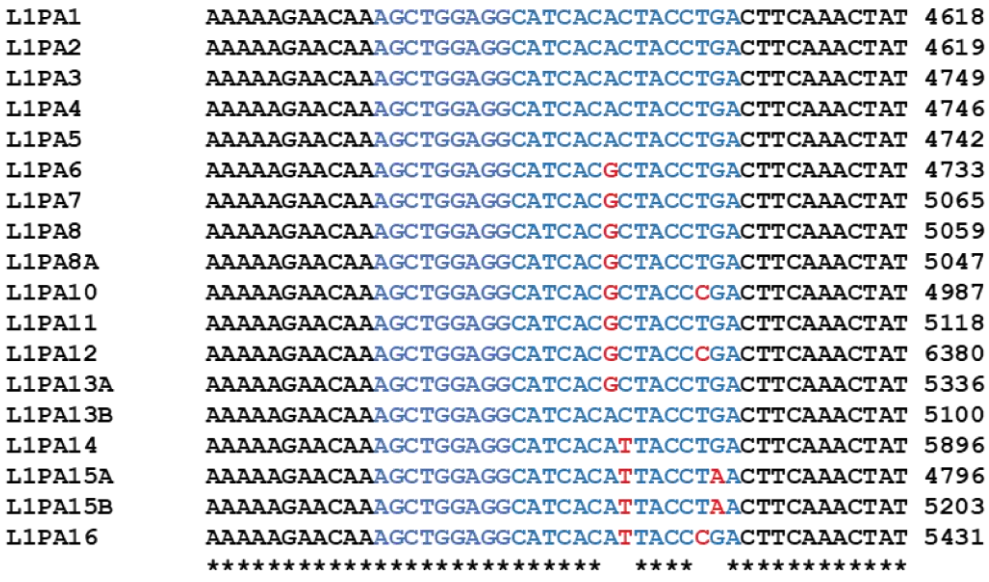


Figure 24. Conservation of bs2 across L1PA families. Alignment of the consensus sequence of L1PA1 to L1PA16 families (evolutionarily youngest, L1PA1, at the top) showing conservation of the let-7 binding site ´bs2´ (in blue). Discordant nucleotides are shown in red

the mutated site 'bs2mut' **(Figure 22E)**, and cloned it into a retrotransposition reporter vector to generate JM101/bs2mutL1.3 **(Figure 23A-B)**. We then performed a retrotransposition assay in HeLa cells upon let-7 depletion. Notably, 'bs2mutL1.3' retrotransposition was less affected by let-7 inhibition than wild-type L1.3 **(Figure 23C)**. Intriguingly, we observed a reduction in bs2mut mobility compared to wild-type L1.3 **(Figure 23C)**. This could be attributed to the mutations introduced, which entail an amino acid change (P to G) in L1-ORF2p **(Figure 23B)**. In fact, a recent report showed that a trialanine mutant in this position of ORF2 (TLP to AAA) presents a 30% reduction in retrotransposition activity (Adney *et al.*, 2019). Notably, the validated ´bs2´site is conserved through primate L1 evolution, being present in L1PA5 elements and containing a few mutations in older L1 subfamilies **(Figure 24)**. This high degree of conservation suggest that these residues are important for L1-ORF2p function, and raise the possibility of a conserved let-7 site across L1 evolution.

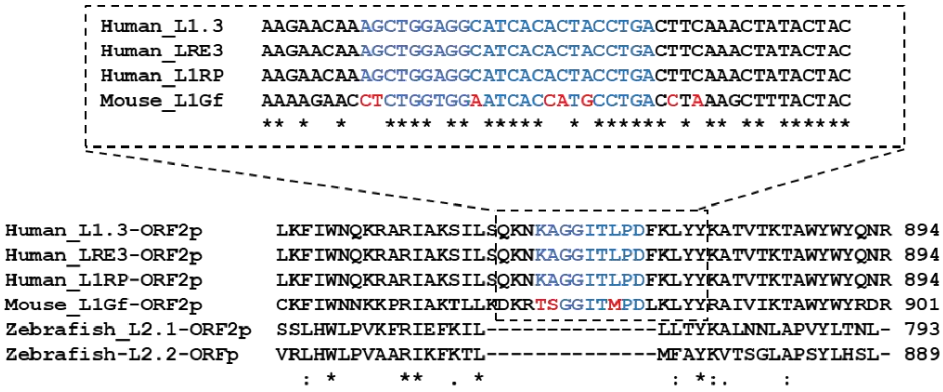


Figure 25. Alignment of the predicted binding site region in ORF2s of all L1s used in this study. Sequences of human L1.3/L1RP/LRE3, mouse $L1G_F$, and zebrafish L2.1 and L2.2 are shown. Alignment was performed with ORF2 protein sequences to localize the binding site region within each LINE (asterisks denote conserved aminoacids). Human and mouse L1s nucleotide sequences were further analyzed. Blue nucleotides represent the 'bs2' region that was validated as a binding site in psiCHECK2. Red nucleotides are those which differ in mouse $L1G_F$ respect to human L1.3, LRE3 or L1RP.

Interestingly, this binding site is present in all human RC-L1s used in this study (L1.3, LRE3 and L1RP), absent in zebrafish LINEs and relatively low conserved in mouse RC-L1s **(Figure 25)**. This is consistent with the specific effect of let-7 observed previously in section R.3: let-7 overexpression and depletion decreased or increased, respectively, retrotransposition of a human L1 but not of a mouse L1 or a zebrafish L2 **(Figure 19)**. However, the fact that mutating this binding site reduced but not abolished the effect of let-7 in L1 mobilization suggest that additional mechanisms mediated by let-7 may work to restrict human L1 retrotransposition. Overall, these results suggest that there is at least a functional let-7 binding site in the ORF2 region of human L1 mRNA.

R.5. Let-7 impairs L1-ORF2p translation

So far, our results have shown that let-7 regulates engineered L1 retrotransposition in cultured cells and identified a functional let-7 binding site in ORF2. Next, we analyzed the functional consequences of let-7 binding to L1 mRNA. MicroRNAs frequently induce mRNA degradation (Jonas and Izaurralde, 2015; Bartel, 2018), therefore we started by analyzing the levels of endogenous L1 mRNAs upon let-7 overexpression in HEK293T cells by RT-qPCR. We found a significant decrease in the mRNA levels of two canonical let-7 targets (DICER and HMGA2) but not in those of L1 mRNA at 24 and 48 hours after let-7 overexpression **(Figure 26A)**. Similarly, L1 mRNA levels were not reduced upon let-7 overexpression **(Figure 26B)**, or increased upon let-7 depletion **(Figure 26C)** when we overexpressed L1 from plasmid JM101/L1.3 in HEK293T cells. The positive control HMGA2 behave as expected **(Figure 26B-C)**. To specifically

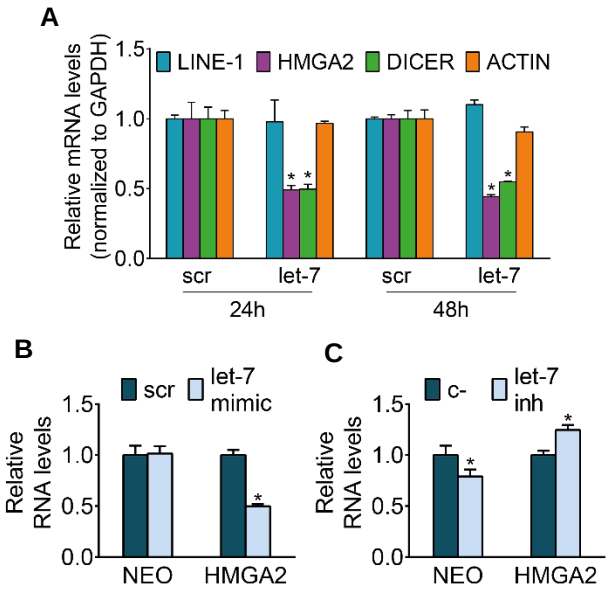


Figure 26. Let-7 does not affect endogenous or overexpressed L1 mRNA levels. (A) RT-qPCR analysis of endogenous LINE-1 (blue bar), HMGA2 (purple bar), DICER (green bar), and ACTIN (orange bar) mRNAs upon let-7 overexpression in HEK293T. Cells were transfected with let-7 mimic or its control (scr), and RNA was extracted at 24h or 48h post-transfection. GAPDH was used to normalize. **(B-C)** RT-qPCR analysis of overexpressed (transfected) L1 mRNA levels. HEK293T were transfected with JM101/L1.3 and **(B)** let-7 mimic or **(C)** let-7 inhibitor (light blue bars), as well as their respective controls 'scr' and 'c-' (dark blue bars). Overexpressed L1 mRNA was specifically detected using primers against the spliced neomycin-resistance cassette. HMGA2 was used as a positive control., EBNA-1 (constitutively expressed from the plasmid backbone) was used to normalize NEO, and GAPDH to normalize HMGA2.
In all graphs, Error bars indicate s.d. of three replicates, and + indicates $p<0.05$ after a two-tailed t-test.

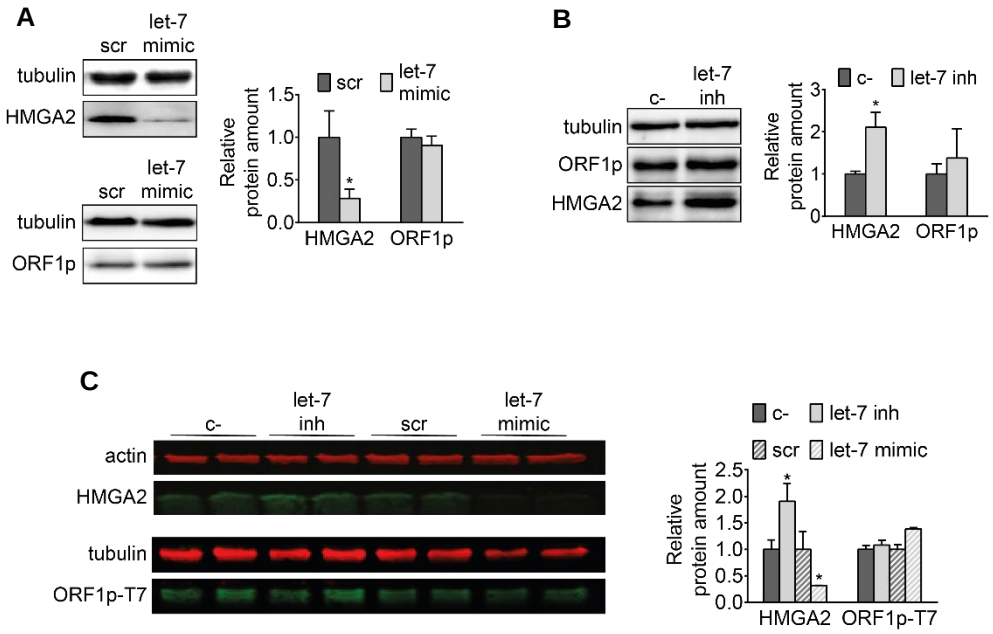


Figure 27. Let-7 does not affect endogenous L1-ORF1p or constitutively expressed T7-tagged L1-ORF1p (A-B) Western blot analyses of endogenous L1-ORF1p and HMGA2 protein levels in HEK293T cells upon let-7 **(A)** overexpression or **(B)** depletion. **(A)** Cells were transfected with let-7 mimic (light grey bars) or its control (scr, dark grey bars). Representative well and quantification of the western blot are shown. Error bars indicate s.d. of four biological replicates. **(B)** Cells were transfected with let-7 inhibitor (light grey bars) or its control (c-, dark grey bars). Representative well and quantification of the western blot are shown. Error bars indicate s.d. of two biological replicates. **(C)** Western blot analyses of stably-expressed T7-tagged L1-ORF1p upon depletion or overexpression of let-7. Stable Flp-In-293 cells expressing T7-tagged L1-ORF1p were transfected with let-7 inhibitor or mimic, or their controls (c- and scr, respectively). HMGA2 was used as a positive control. Western blot (left) and its quantification by Odyssey (right) are shown. Error bars indicate s.d. of two replicates. In the graph, dark grey bars correspond to c-, light grey bars to let-7 inhibitor, dashed dark grey bars to scr, and dashed light grey bars to let-7 mimic.
In all graphs, * indicates $p<0.05$ after a two-tailed t-test.

detect mRNA coming from the plasmid, we used primers to detect the spliced Neomycin cassette (see section M.18 of the Methods). These data suggest that let-7 expression does not trigger L1 mRNA degradation.

The main other main effect of miRNAs on target mRNAs is interference with protein translation, particularly when targeting occurs in the coding sequence (Hausser and Zavolan, 2014; Jonas and Izaurralde, 2015; Zhang *et al.*, 2018). Thus, we analyzed the levels of endogenous L1-ORF1p in HEK293T cells upon modulation of let-7 levels. We found significant changes in HMGA2 but not in

L1-ORF1p expression upon let-7 overexpression **(Figure 27A)** or depletion **(Figure 27B)**. To further corroborate this results, we overexpressed and depleted let-7 in a previously published HEK293 cell line that constitutively overexpresses a T7-tagged L1-ORF1p from a CMV promoter (MacLennan *et al.*, 2017). Consistently, we observed changes in HMGA2 but not in L1-ORF1p-T7 upon modulation of let-7 levels **(Figure 27C).**

The results shown above suggest that let-7 does not affect the levels of L1 mRNA or L1-ORF1p. Therefore, we set out to analyze changes in L1-ORF2p. The translation of L1-ORF2p occurs by a highly inefficient unconventional termination/reinitiation mechanism that produces few molecules of L1-ORF2p molecule per L1 mRNA which are nevertheless enough to support efficient retrotransposition (Alisch *et al.*, 2006; Taylor *et al.*, 2013). Indeed, a recent study failed to detect L1-ORF2p using five different antibodies in human tumor samples or cell lines by western blot, immunoprecipitation, or immunohistochemistry (Ardeljan *et al.*, 2019). In summary, it is technically challenging to detect endogenous L1-ORF2p. Therefore, to study changes in L1-ORF2p levels upon let-7 modulation, we used a monocistronic construct expressing 2xFLAG-tagged L1-ORF2p from a CMV promoter (L1-ORF2p from L1.3), pSA500 **(Figure 28A)** We co-transfected HeLa cells with pSA500 and let-7 mimic or inhibitor and their respective controls, and took three different fractions of each replicate to perform different analysis at the protein, RNA and DNA levels. Strikingly, by western-blot observed a decrease in L1-ORF2p upon let-7 overexpression **(Figure 28B)** and an increase upon let-7 depletion **(Figure 28C)**. This behaviour was very similar to that of a well described let-7 target containing several 8mer sites: DICER (Forman, Legesse-Miller and Coller, 2008) **(Figure 28B-C)**. Consistent with the data presented in **Figure 27**, we did not observe any changes in the levels of exogenous L1-ORF2-F RNA upon let-7 overexpression **(Figure 28D)** or depletion **(Figure 28E)**, as opposed to DICER whose mRNA was affected

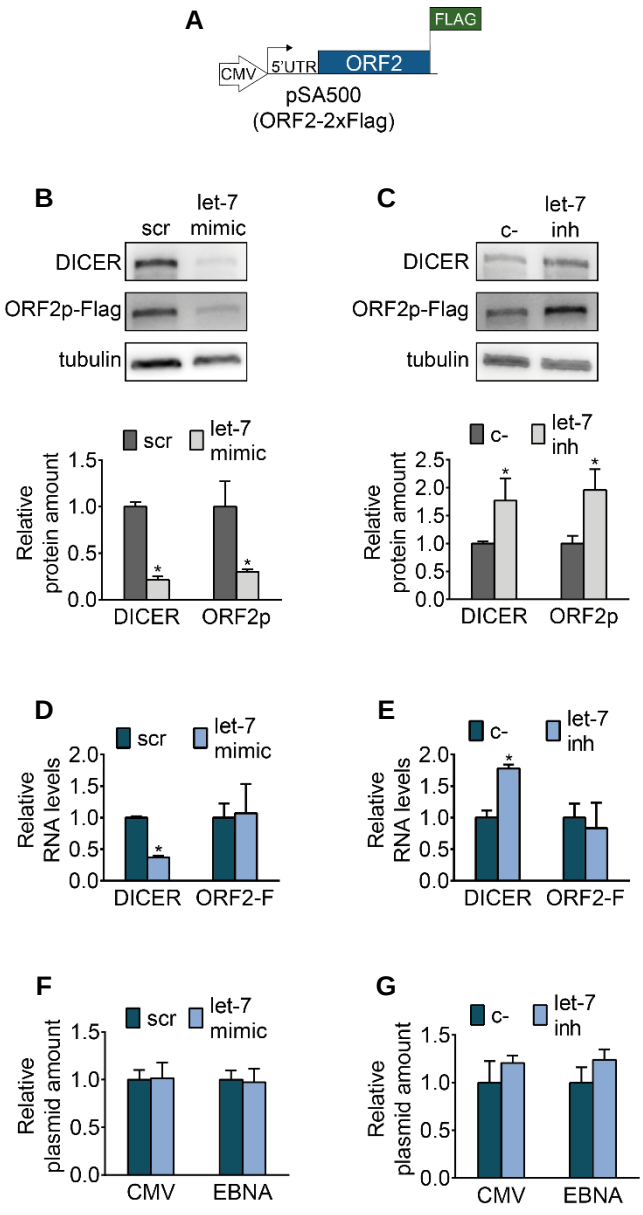


Figure 28. Let-7 impairs L1-ORF2p translation. (A) Scheme of construct pSA500, showing CMV promoter (white arrow), L1 5'UTR and its promoter (black arrow), L1.3 ORF2 (blue rectangle), and 2xFlag tag (green flag). **(B-C)** Western blot analyses of L1-ORF2p-FLAG levels upon let-7 overexpression **(B)** or depletion **(C)**. HeLa JVM cells were co-transfected with pSA500 and **(B)** let-7 mimic or its control 'scr', or **(C)** let7 inhibitor or its control 'c-'. L1-ORF2p was detected using a FLAG antibody. DICER, a known let-7 target, was used as a positive control. Representative well is shown, and quantification of the western blot is shown below. Error bars represent s.d of three replicates. * indicates $p<0.05$ after two-tailed t-test. **(D-E)** RT-qPCR analyses of the levels of DICER and L1-ORF2-Flag mRNA upon overexpression **(D)** or depletion **(E)** of let-7. Transfected RNA was specifically measured by using primers for the SV40 polyadenylation signal (see methods). GAPDH was used to normalize. Error bars represent s.d of three replicates. * indicates $p<0.05$ after two-tailed t-test. **(F-G)** qPCR analysis of the relative plasmid amount (i.e. transfection efficiency) for experiments shown in **B-E**. Plasmid levels (pSA500) upon let-7 overexpression **(F)** or depletion **(G)** were analyzed by qPCR using two different pairs of primers: CMV and EBNA. Genomic GAPDH was used to normalize. Error bars represent s.d of three replicates.
In all cases, protein, RNA or plasmids levels are shown relative to the control used (scr or c-). Of note, all analyses were done using different fractions of the same cell pools, to limit experimental variability.

by both let-7 overexpression or depletion **(Figure 28D-E)**. To specifically detect exogenous ORF2-F RNA, we used primers for the SV40 polyadenylation signal located in the plasmid (see section M.18 of the methods). Lastly, to rule out that

differences in L1-ORF2p expression were due to different transfection efficiencies, we quantified plasmid levels by qPCR using primers targeting the CMV promoter driving L1-ORF2p expression or the EBNA-1 sequence in the plasmid backbone. We did not observe significant differences in the amount of plasmid co-transfected with let-7 mimic and its control **(Figure 28F)** or let-7 inhibitor and its control **(Figure 28G)**. These data suggest that the differences in L1-ORF2p levels are not due to variations in transfection or mRNA stability but to a direct effect of let-7 in L1-ORF2p translation.

The specific translational effect shown above could be due to the structure of this offset 7-mer site or to the location within the coding sequence. To gain insight into this question, we generated two types of pSA500 mutants. On one hand, we introduced three different sequences in the 3'UTR of pSA500: a scrambled sequence ('scrb'), the binding site ('bs2') and a modified bs2 that contains a canonical 8-mer site for let-7 ('8mer') **(Figure 29A)**. On the other hand, using site directed mutagenesis we generated two point-mutations in the ORF2 coding region to transform the offset 7-mer into a canonical let-7 8-mer site, constructing pSA500-ORF2-8mer **(Figure 29B)**. The 8mer sites drive the strongest repression of gene expression by microRNAs (Bartel, 2009, 2018).

First, we co-transfected the three pSA500-3'UTR mutants and let-7 mimic or its control in HeLa cells, and divided each sample in two fractions to analyze protein and RNA levels. Western blot analyses showed that placement of 'bs2' and '8mer' sequences in the 3'UTR of pSA500 enhanced the reduction of ORF2-F upon let-7 overexpression **(Figure 29C)**. Importantly, using RT-qPCR we confirmed that similar levels of transfection (measuring constitutive EBNA expression from the plasmid backbone) and let-7 overexpression (measuring the effect on endogenous DICER) were achieved **(Figure 29D)**. Regarding the levels of ORF2-F RNA, we observed a clear tendency towards a reduction on the RNA levels upon placement of the binding site ('bs2') or the modified 8-mer binding

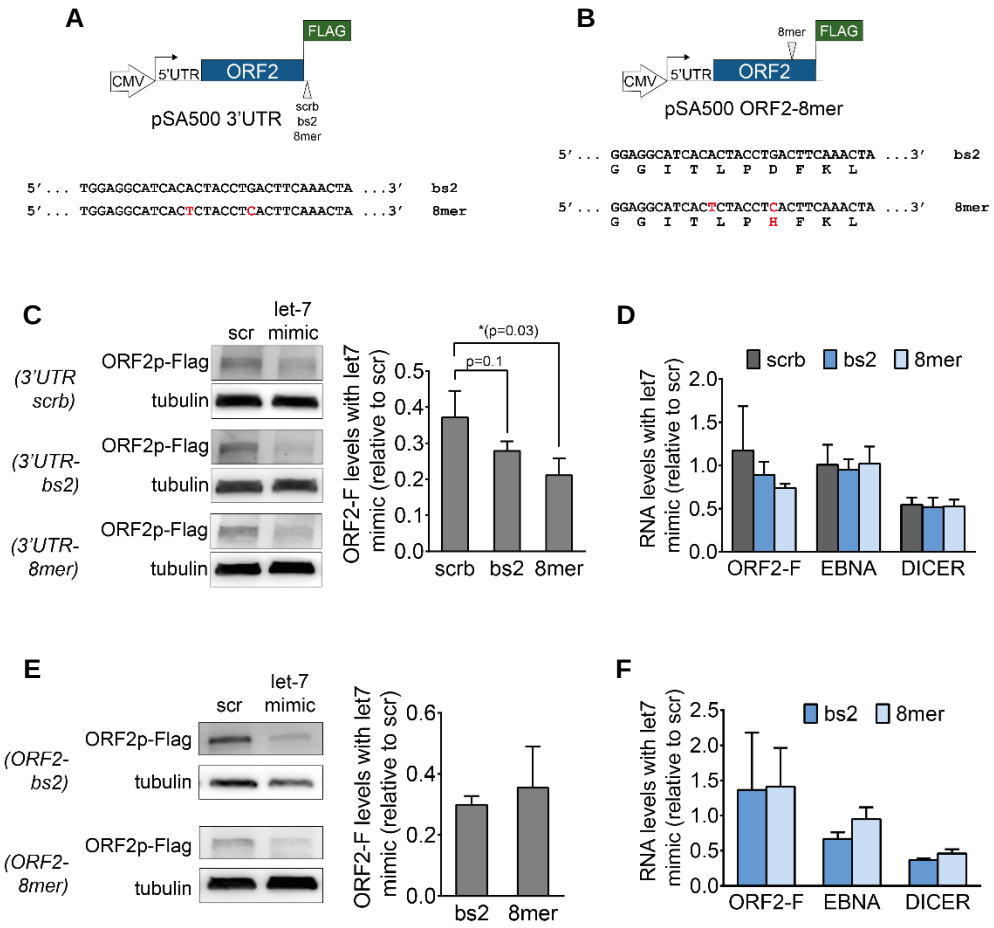


Figure 29. The specific translational effect of let-7 in ORF2 appears to be location-dependent. (A) Scheme of the three 3'UTR mutants generated in pSA500. Sequences of bs2 and 8mer are shown below, and differing nucleotides are marked in red.**(B)** Scheme of pSA-ORF2-8mer construct. Sequence of the mutated ORF2 is shown below compared to wild type ORF2, and discordant nucleotides and aminoacids are marked in red. **(C)** Western-blot analyses of the effect of let-7 overexpression in each of the pSA500-3'UTR mutants. HeLa cells were co-transfected with one of the constructs and let-7 mimic or its control 'scr'. A representative well is shown, followed by western-blot quantification represented as the relative levels of ORF2-F upon let-7overexpression relative to the control condition. Error bars represent s.d of three replicates. **(D)** RT-qPCR analysis of the levels of ORF2-F, EBNA and DICER mRNA upon let-7 overexpression. Transfected ORF2-F RNA was specifically measured by using primers for the SV40 polyadenylation signal (see methods). For ORF2-F, EBNA was used to normalize. For EBNA and DICER, GAPDH was used to normalize. Error bars represent s.d of three replicates. **(E)** Western-blot analyses of the effect of let-7 overexpression in each of the pSA500-ORF2. HeLa cells were co-transfected with one of the constructs and let-7 mimic or its control 'scr'. A representative well is shown, followed by western-blot quantification represented as in **(C)** of this figure. **(F)** RT-qPCR analysis of the levels of ORF2-F, EBNA and DICER mRNA upon let-7 overexpression. The primers and normalization used are exactly as in **(D)** of this figure.Of note, all analyses were done using different fractions of the same cell pools, to limit experimental variability.

site ('8mer') in the 3'UTR **(Figure 29D)** although the reduction was not significant (p=0.08 for '8mer').

Additionally, we co-transfected pSA500-ORF2-8mer and pSA500 in HeLa cells with let-7 mimic or its control. Western blot analyses revealed a similar effect of let-7 overexpression on wild-type L1-ORF2p or 8-mer-containing L1-ORF2p **(Figure 29E)**. We confirmed that transfection levels (measuring constitutive EBNA expression from the plasmid backbone) and let-7 overexpression (measuring the effect on endogenous DICER) were similar in both cases **(Figure 29F)**. Finally, we did not observe changes in ORF2-F RNA levels upon let-7 overexpression between wild- type ORF2 and 8-mer-containing ORF2 **(Figure 29F)**. Of note, we observed a clear reduction of L1-ORF2p levels in the scr condition when transfecting the pSA500 ORF2-8mer construct compared to pSA500 **(Figure 29E)**. Since one of these mutations causes an aminoacid change (D to H) **(Figure 29B)**, and a recent study reported that mutation of this residue and the next two to three alanines (DFK to AAA) reduced retrotransposition to 6.7% of the wild-type (Adney *et al.*, 2019), one reason for this could be that this specific aminoacid is critical for protein stability. However, we speculate that endogenous let-7 could also be binding more efficiently to this canonical site and therefore reducing L1-ORF2p-8-mer levels under control conditions. Taken together, these results suggest that the translation-specific effect of this site depends on its location within the coding sequence rather than on its structure.

Lastly, to further characterize this effect we analyzed the impact of let-7 on L1-ORF2p translation in its natural context: a full-length bicistronic L1 RNA where L1-ORF2p is translated using the aforementioned termination/reinitiation mechanism (Alisch *et al.*, 2006). For this purpose, we used a construct where, in a human L1.3 element, L1-ORF1p and L1-ORF2p are fused to EGFP and mCherry, respectively, in their C-terminus: pVan583 **(Figure 30A)**. Although the addition of both fluorescent tags reduces the efficiency of retrotransposition to ~30% of the untagged element **(Figure 30B)**, the encoded L1 is still biologically active. First, by confocal microscopy we observed a reduction of L1-ORF2p-

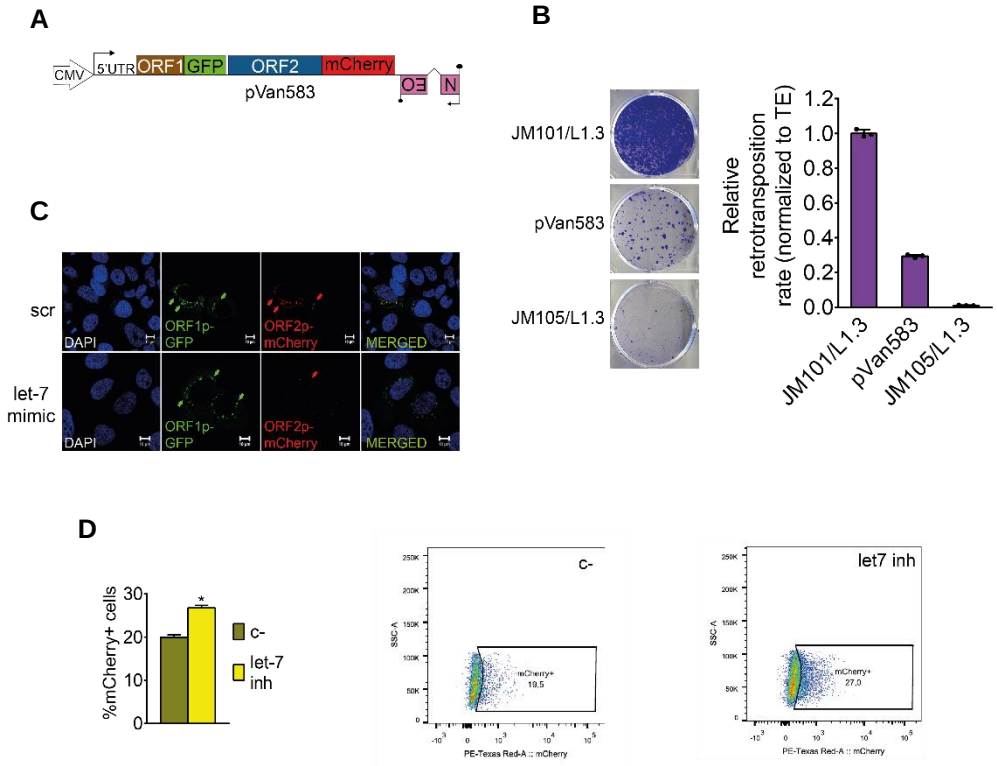


Figure 30. Let-7 control of ORF2 unconventional translation in its endogenous context. **(A)** Scheme of plasmid pVan583, a derivative of JM101 with EGFP fused to the C-terminus of ORF1 and mCherry fused to the C-terminus of ORF2. **(B)** Cell culture-based retrotransposition assay comparing retrotransposition efficiency of different constructions. HeLa cells were transfected with either JM101/L1.3, pVan583 or the negative control JM105/L1.3 (encoding a missense mutation in the RT domain, D702A, that renders the element unable to mobilize) and selected with neomycin. Quantification (corrected for transfection efficiency, TE) is shown in the right. **(C)** Confocal microscopy analysis of L1-ORF1p-GFP and L1-ORF2p-mCherry (in red) upon let-7 over-expression. U2-OS cells were co-transfected with pVan583 and let-7 mimic or its control (scr), and fluorescence was analyzed by confocal microscopy. Arrows indicate L1-ORF1p-EGFP or L1-ORF2p-mCherry foci. **(C)** Flow cytometry quantification of L1-ORF2p-mCherry levels upon let-7 depletion. HeLa cells were co-transfected with pVan583 and let-7 inhibitor or its control (c-), and fluorescence was measured by flow cytometry. Graph on the left shows the per-centage of mCherry+ cells in the EGFP+ (transfected) population. Error bars indicate s.d. of three replicates. * indicates $p<0.05$ after a two-tailed t-test. A representative cytometry plot of three replicates in each condition is shown (the percentage of ORF1p-GFP positive cells expressing ORF2p-Cherry protein).

mCherry but not of L1-ORF1p-EGFP levels upon overexpression of let-7 in U2OS cells **(Figure 30C)**. However, the reduced transfection capacity of this construct and the inefficient translation of L1-ORF2p-mCherry, prevented us from obtaining a sufficient number of double positive cells to enable a quantitative analysis by microscopy. Therefore, we turned to a more sensitive and

quantitative approach: flow cytometry. We co-transfected pVan583 with let-7 inhibitor in HeLa cells and analyzed EGFP+ cells (i.e. transfected cells, >3,500 per sample). Notably, we found that depleting let-7 led to an increase in the number of mCherry+ cells in the EGFP+ population, suggesting an increase in the synthesis of L1-ORF2p-mCherry upon let-7 inhibition **(Figure 30D)**.

Altogether, our results suggest that let-7 impairs L1-ORF2p translation, potentially altering the ratio between L1-ORF1p and L1-ORF2p, which we speculate could unbalance L1-RNP formation and, consequently, reduce L1 retrotransposition.

R.6. Generation of a lung cancer cell line with inducible expression of let-7 to study the regulation of endogenous L1 retrotransposition

Using an extensive array of bioinformatic, biochemical and molecular biology techniques we have shown that downregulation of let-7 microRNAs correlates with increased somatic insertions in lung tumor samples, and that let-7 regulates engineered L1 retrotransposition (sections R.1 to R.5). Recent advances in high-throughput sequencing and computational tools allow the identifications of endogenous somatic L1 insertions (Muñoz-Lopez *et al.*, 2016; Faulkner and Garcia-Perez, 2017; Scott and Devine, 2017; Goerner-Potvin and Bourque, 2018). Therefore, we decided to generate a cellular model in which we could stably restore let-7 levels and study whether this could control endogenous L1 retrotransposition.

Since we had found a high number of *de novo* L1 insertions in lung tumor samples, we hypothesised that a lung cancer cell line would be an ideal model for our purpose. There are two main types of lung cancer: small and non-small cell lung cancer (NSCLC), the latter one accounting for over 80% of the diagnosis[*]. The two most frequent subtypes of NSCLC are lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC). It is already described that LUSC tumors have a reduced expression of let-7 compared to LUAD (Fassina, Cappellesso and Fassan, 2011). Moreover, LUSC samples present a higher number of somatic insertions than LUAD samples as analyzed by Transpo-Seq (Helman *et al.*, 2014), and by MELT in our analyses. Evidence, therefore, suggested that a LUSC cell line would be better fitted for our purpose of restoring let-7 levels and study its effect in endogenous L1 mobilization. First, to study L1

[*] Data from cancer.net (American Society of Clinical Oncology)

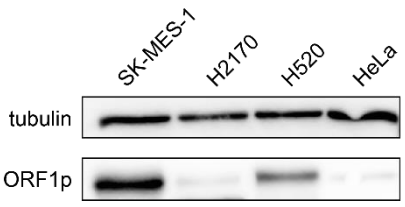


Figure 31. Western-Blot analysis of L1-ORF1p expression in several LUSC cell lines. HeLa cells are not LUSC-derived, and were used as a negative control because of their limited expression of L1-ORF1p.

expression, we analyzed the levels of L1-ORF1p by Western-Blot in a panel of LUSC cell lines provided by the group of Dr. Pedro Medina at Genyo, using HeLa cells as a control for a reduced expression of this protein. We found that SK-MES-1 cells present the highest endogenous levels of L1-ORF1p **(Figure 31)**. Notably, in previous experiments (see section R.2) we already showed that SK-MES-1 cells have reduced levels of let-7 compared to HeLa **(Figure 18)** and that they support LINE-1 retrotransposition **(Figure 17E)**. In fact, we showed that transient depletion of let-7 increased engineered L1 retrotransposition in this cell line **(Figure 17E)**. Therefore, we decided to use SK-MES-1 cells to generate an inducible model to overexpress let-7.

Lentiviral vectors are a fast and efficient way to generate stable cell lines that express a transgene in an inducible way. Thus, in collaboration with the laboratory of Dr. Francisco Martín at Genyo, we used a previously published lentiviral vector (Benabdellah *et al.*, 2016) to generate CEET-NL2-IS2 pri-let-7a-3 **(Figure 32)**. Briefly, this construction allows the inducible expression of a transgene (in this case, pri-let-7a-3) thanks to the upstream CMV promoter that contains the TetO sequence. This sequence is bound by the Tet repressor (TetR) expressed from the same construct via the EF1α* promoter **(Figure 32)**. Therefore, the transgene can only be expressed upon induction with Doxycycline (Dox), that sequesters the TetR and allows transcription of the transgene.

* Elongation factor 1-alpha

Figure 32. Structure of the lentiviral vector CEET-NL2-IS2 pri-let-7a-3. 5'LTR, 5' Long terminal repeats (brown and grey rectangle); CMV-TetO, CMV promoter (white arrow) containing a TetO sequence (green square) that will be recognized by TetR; pri-let-7a-3, the let-7a-3 gene, containing the full primary miRNA transcript (red rectangle); EF1α, human eukaryotic translocation elongation factor α 1 promoter (white arrow); TetR, tetracyclin repressor (green rectangle) with the addition of the nuclear localization signal (NL2, yellow square) of the glucocorticoid receptor; WPRE, Woodchuck Hepatitis Virus Posttranscriptional Regulatory Element (grey rectangle); 3'LTRΔ3, Self-inactivating 3' Long terminal repeat (brown and grey rectangle) containing a chimeric insulator (IS2, blue square). Orange arrows indicate the location and orientation of the primers used to quantify the number of inserted copies per cell by qPCR. Upon integration, the 3'LTR is placed upstream of the 5'LTR, therefore these primers enable specific amplification of integrated, and not plasmid, DNA.

First, to confirm the functionality of this inducible system in SK-MES-1 cells, we transduced them with lentiviruses containing EGFP as the transgene. These viruses were a gift from Dr. Francisco Martín's lab, and this experiment was performed in collaboration with María Tristan (Genyo). We incubated the transduced cells with different Dox concentrations ranging from 0.01 to 0.4μg/mL and measured transgene (EGFP) expression by flow cytometry 48 hours after Dox induction. We found that robust expression of the transgene, measured as %GFP+ cells and Median Fluorescence Intensity (MFI) of GFP, was already achieved with 0.01μg/mL Dox (**Figure 33).**

The reason for choosing this particular member of the let-7 family, let-7a-3 is that it is the only one not post-transcriptionally regulated by LIN28 (see section I.14) (Triboulet, Pirouz and Gregory, 2015; Ustianenko et al., 2018). Therefore, in the future, this construction could be used to increase let-7 levels in embryonic

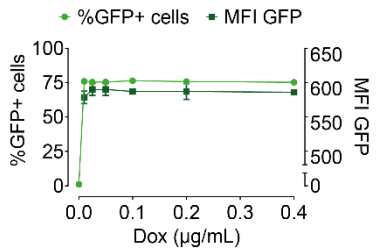


Figure 33. Optimization of Doxycycline concentration to induce transgene expression in SK-MES-1 cells. Cells were transduced with CEET-NL2-IS2 containing EGFP as the transgene, whose expression was induced with different concentrations of Dox. Analysis of %GFP+ cells (light green) and MFI (Median Fluorescence Intensity, dark green) for GFP is shown as the average of triplicates. Error bars indicate s.d.

Clone	LTR	Albumin	Copies/cell
a3 A2	**15706,546**	**4200,627**	**3,739**
a3 Tetra	4084,937	7350,271	0,555
a3 B1	**8838,616**	**5824,680**	**1,517**
a3 B4	86127,867	6587,253	13,074
a3 B6	**47092,296**	**6990,409**	**6,736**

Table 4. Quantification of the number of inserted transgene copies per cell in different clones. LTR refers to quantification of the integrated transgene. Albumin was used to normalize. Number of copies was calculated as the ratio of LTR to albumin. Clones used for the next experiments are highlighted in orange. LTR and Albumin are expressed in copies per sample, and are used to calculate the number of inserted transgene copies per cell shown in the last column.

stem cells, where LIN28 is highly expressed. Once confirmed that the system would work in our cell line, in collaboration with Araceli García and Esther Prada (Genyo) we transduced SK-MES-1 cells with lentiviruses containing CEET-NL2-IS2 pri-let-7a-3 (detailed description in Methods section), generating a pool of transduced cells hereafter called SK-MES-1(a3) cells. Integration of the lentiviral cDNA ends up with the 3'LTR positioned upstream of the 5'UTR. Therefore, we can design primers that specifically amplify inserted DNA and not plasmid DNA, as in the latter case the primers would be facing outwards **(Figure 33)**. Using qPCR, and by comparison with gDNA from a different cell line containing 1 copy per cell (a gift from Dr. Francisco Martín, Genyo), we estimated that, on average, our pool of SK-MES-1(a3) cells contained ~7 inserted copies per cell of our construction. From that assay, we estimated the MOI* to be 10. We generated clonal cell lines from that pool, and quantified the number of copies using the previously described qPCR approach. A few examples are shown in **Table 4**.

Next, we tested the effect of Dox induction in the three clones highlighted in **Table 4** (A2, B1 and B6) which have different numbers of integrated copies per cell. For that, we cultured them with 0.01 or 0.1μg/mL Dox and, 72h post-induction, measured the expression of pri-let-7a-3 RNA and HMGA2, a well-

* Multiplicity of Infection: the ratio (number of viral particles used to infect cells)/(number of cells).

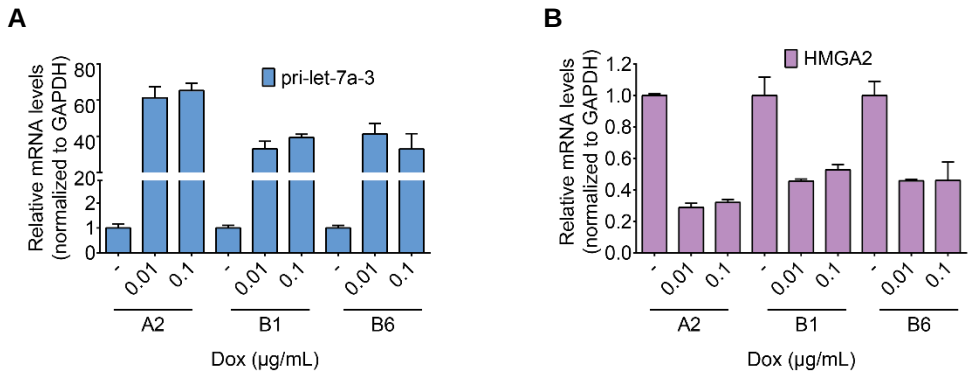


Figure 34. Dox induces transcription of pri-let-7a-3 which can be processed to generate mature let-7 that decreases HMGA2 mRNA levels. Three different SK-MES-1(a3) clones (A2, B1, B6) were cultured with no Dox (-), 0.01 or 0.1µg/mL Dox, and RNA levels of pri-let-7a-3 and HMGA2 were analysed 72h post-induction. Average of three replicates are shown. Error bars indicate s.d.

known target of let-7 (Lee and Dutta, 2007). This way, we can analyze (i) transgene expression upon Dox induction, and (ii) whether this pri-let-7a-3 is processed via Microprocessor and Dicer to mature, functional let-7. By qRT-PCR, we observed a large increase in the levels of pri-let-7a-3 upon Dox induction in all clones **(Figure 34A)** accompanied by a decrease in the mRNA levels of the target gene HMGA2 **(Figure 34B)**. We did not observe significant differences between induction with 0.01 or 0.1µg/mL Dox, which is consistent with our previous result in flow cytometry **(Figure 33)**. Considering these results, we decided to continue with clone B1 as it has the lowest copy number **(Table 4)**.

Lastly, in order to validate whether this induced let-7 was indeed able to regulate engineered L1 mobilization, we performed a luciferase-based retrotransposition assay in the presence or absence of Dox with the B1 clone. We used two plasmids that have been previously described (See section R.2). Briefly, pYX017 contains an active L1 whose expression is driven by the highly active CAG promoter, whereas pYX015 contains an inactive L1. In both cases, the L1 is followed by a firefly luciferase cassette that is activated upon retrotransposition. The backbone of the plasmids contains a constitutively expressed Renilla luciferase gene that allows normalization for transfection efficiency. We observed a reduction of

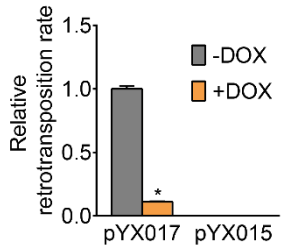


Figure 35. Dox-mediated induction of let-7 decreases engineered L1 retrotransposition in SK-MES-1(a3) B1 clone. Cells were transfected with either pYX017 or pYX015 (see Figure X for plasmid details), and incubated with Dox for 72 hours before measuring luciferase activity. Average of three replicates. Error bars indicate s.d. *, p<0.05.

engineered L1 retrotransposition upon Dox-mediated induction of let-7 expression, suggesting that our system indeed generates mature let-7 miRNA that is able to control the L1 mobilization **(Figure 35).**

Next steps of this project will include long-term culture of these cell line in the presence or absence of Doxycycline (i.e. inducing or not the expression of let-7), extraction of genomic DNA at different passages and analysis of the WGS data to identify and compare the number of tumor-specific L1 insertions by MELT as previously was described in section R.1.

R.7. miR-20 could be an L1 activator through negative regulation of an L1 retrotransposition repressor

In our bioinformatic analysis to determine misexpressed microRNAs in lung tumor samples with tumor-specific L1 insertions (section R.1), we observed an additional miRNA whose expression differed significantly between the samples with and without insertions: miR-20. However, its behaviour was the opposite of let-7 and miR-34 as it was increased in the tumors presenting de novo L1 insertions compared with tumors without them **(Figure 10)**.

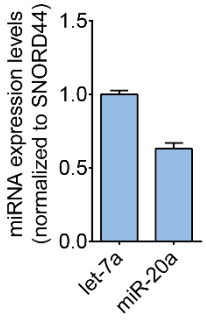


Figure 36. Quantification of mature let-7 and miR-20a levels in HeLa cells. 3'RACE RT-qPCR was performed to quantify mature miRNAs in HeLa cells. Averages of three replicates are shown. Error bars indicate s.d.

We set out to understand the potential role of miR-20 in regulating L1 mobilization. First, we analyzed mature miR-20 levels in HeLa cells by RT-qPCR, and found that this miRNA is highly expressed as its levels are approximately half of let-7a, the most expressed miRNA in HeLa (Khan *et al.*, 2009) **(Figure 36)**. Accordingly, we took advantage of our recently developed sRNA/L1 retrotransposition assay (Tristan-Ramos *et al.*, 2020) to deplete endogenous miR-20 and measure L1 mobilization.

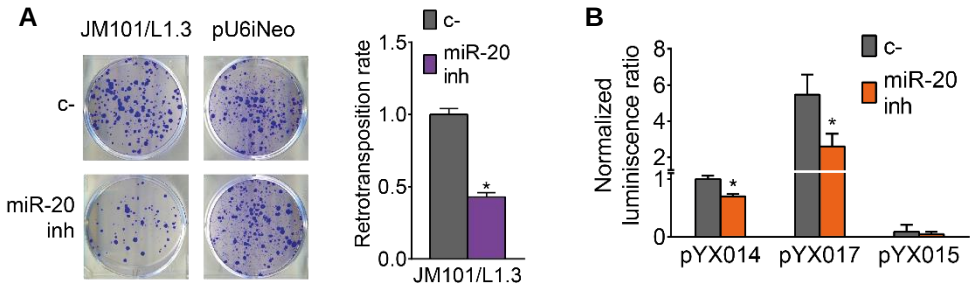


Figure 37. Depletion of miR-20 decreases engineered L1 retrotransposition. (A) HeLa cells were cotransfected with pJM101/L1.3 or pU6iNeo (See Figure 14A for structures) and miR-20 inhibitor or its control (c-). A representative well of three replicates is shown. Quantification is shown at the right as average of three replicates. **(B)** HeLa cells were cotransfected with pYX014, pYX015 or pYX017 (see Figure 14F and Figure 18B for structures) and miR-20 inhibitor or its control (c-). Averages of three replicates are shown. * indicates $p<0.05$ after two-tailed t test.

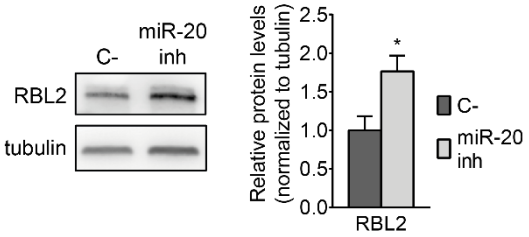


Figure 38. Depletion of miR-20 increases RBL2 protein levels. Representative of three replicates is shown. Quantification is shown in the right as average of three replicates. Error bars indicate s.d. * indicates p<0.05 after two-tailed t test.

Consistent with the correlation found *in silico*, depletion of miR-20 decreased L1 retrotransposition without affecting clonability of the cells **(Figure 37A).** For details of JM101/L1.3 and pU6iNeo structures see **Figure 14A.** Moreover, we corroborated this result using a different system in which the reporter cassette is a luciferase gene (plasmids pYX014, pYX015 and pYX017, see **Figure 14F** and **Figure 18B**). Consistently, we observed that miR-20 depletion led to a decrease in L1 retrotransposition **(Figure 37B)**. As a control, we analyzed the effect of miR-20 depletion in one of its top direct target genes, RBL2[*] (Trompeter *et al.*, 2011), by western blot. As expected, we observed an ~1.7-fold increase in RBL2 protein levels upon miR-20 depletion **(Figure 38)**. Altogether, these results suggest that miR-20 depletion decreases engineered L1 retrotransposition.

The aforementioned results point towards an indirect effect of miR-20 in L1 retrotransposition through regulation of proteins that function as repressors of L1 mobilization. Prediction of miRNA targets is challenging (see section I.12), therefore we used three independent target prediction softwares (microrna.org, miRdb.org and TargetScan.org) to identify potential miR-20 targets. Notably, 22 cellular genes were simultaneously predicted to be targeted by miR-20 by the three softwares **(Figure 39)**. Additionally, 7 different target mRNAs had been previously validated using cultured cells and luciferase constructs (Trompeter et al., 2011). To analyze whether any of these 29 genes could function as a regulator

[*] Retinoblastoma-like protein 2

of L1 retrotransposition, we investigated the results of a recent genome wide CRISPR/Cas9 screening (Liu *et al.*, 2018) that has characterized the potential role as L1 retrotransposition repressors for thousands of genes, as their depletion via CRISPR/Cas9 increased L1 mobilization in K562 or HeLa cells (Liu *et al.*, 2018). Interestingly, we found that 22 of these genes had been tested in this CRISPR/Cas9 screening. Intriguingly, 17 of those 22 genes potentially regulated by miR-20 were classified as potential L1 repressors in the genome-wide study previously mentioned: the estimated effect of their depletion is positive (>1), for L1 mobilization (in other words, their depletion increases engineered L1 retrotransposition) in K562 and HeLa cells. These 17 genes are highlighted in **Table 5**. Notably, Gene Ontology analysis* using PANTHER† (Mi *et al.*, 2019) revealed that a significant proportion of those mRNAs (at least 6/22) targeted by miR-20 encode for proteins involved in cell cycle regulation **(Table 6)**. Further study is warranted to understand the influence of miR-20 the control of L1 retrotransposition.

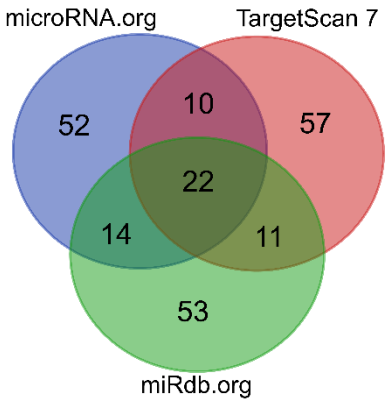


Figure 39. Twenty-two cellular genes are predicted to be targeted by miR-20 using three different miRNA target prediction softwares. MicroRNA.org, TargetScan and miRdb.org were used to predict miR-20 targets, and 22 were identified by the three of them.

* geneontology.org
† Protein Analysis Through Evolutionary Relationships

	Maximum effect estimate (95% C.I.)			Maximum effect estimate (95% C.I.)	
Gene	K562	HeLa	Gene	K562	HeLa
CCND1	1.9	1.7	EZH1	3	2.7
CCND2	2	1.1	FBXL5	4.3	1.6
CDNK1A	2.3	1.5	GPR6	5.8	3.5
EF21	1	3.8	MYT1L	1.1	2.6
PTEN	3	4.6	PKD2	2	2.6
RBL1	1	2	PLEKHA3	0.5	2.9
RBL2	4.2	2.2	RRAGD	4	2.1
CFL2	4.3	1.3	RUFY2	1.7	6.2
DYNC1LI12	1.7	3.5	VLDLR	2.7	1.8
ENPP5	0.9	3.1	ZNF800	2.1	3.9
EPHA4	-0.1	2.8	ZNFX1	5.6	4

Table 5. Human genes potentially regulated by miR-20 and the maximum estimated effect of their depletion in L1 retrotransposition in K562 or HeLa cells. Genes shown in blue have been validated experimentally as *bona fide* miR-20 targets (Trompeter *et al.*, 2011), whereas genes shown in green are found by three independent target prediction softwares (microrna.org, miRdb.org and Tar-getScan.org). The Maximum effect estimate on L1 retrotransposition rate upon depletion was obtained from (Liu *et al.*, 2018). Genes in which Maximum effect estimate was >1 in both cell lines are underlined.

GO biological process	Count	p-value	Genes
Regulation of cell cycle G1/S phase transition	6	1.25E-04	RBL1, RBL2, PTEN, PKD2, CCND1, CCND2
Regulation of G1/S transition of mitotic cell cycle	6	6.18E-05	

Table 6. Several predicted miR-20 targets that could function as repressors of L1 mobilization can are involved in cell cycle regulation. Gene Ontology analysis was performed with PANTHER software. Count denotes the number of genes involved in that particular biological process. P-value was adjusted for multiple testing using the Bonferroni correction. RBL1/2: Retinoblastoma-like protein 1/2. PTEN: Phosphatase and Tensin Homolog. PKD2: Polycystin 2. CCND1/2: Cyclin D1/D2.

R.8. CLASH: unbiased identification of miRNA-TE interactions *in vivo*

Most experiments in this section, including CLASH*, were performed during my stay at the Institute for Immunity and Infection Research of the University of Edinburgh (Edinburgh, UK).

So far, we focused on the let-7 miRNA because in our bioinformatic analysis of whole genome and miRNA-sequencing data from tumor samples we found that downregulation of several members of this miRNA family correlated with increased somatic L1 insertions (see section R.1). Similarly, we analyzed the role of miR-20 in L1 retrotransposition because it was significantly increased in samples with tumor-specific L1 insertions (see section R.6). However, we wondered what other miRNAs may be involved in directly regulating L1 mobilization.

To identify novel miRNAs that might be involved in controlling L1 mobilization, we performed CLASH in human embryonic teratocarcinoma cells (PA-1), which are characterized by a stable karyotype, expression of embryonic markers such as OCT4, NANOG or SOX2, and high endogenous expression of LINE-1 retrotransposons (Garcia-Perez *et al.*, 2010). CLASH allows the identification and precise mapping of miRNA-mRNA interactions *in vivo* (see section I.13). Notably, although the first CLASH was performed in overexpressed tagged AGO1 (Helwak *et al.*, 2013; Helwak and Tollervey, 2014), we immunoprecipitated endogenous AGO2, the most thoroughly characterized. Detailed description of the CLASH protocol is provided in section M.23 of the Methods. Here, representative results from one of the three replicates that were performed will be shown, in parallel with an overview of all the steps.

* Cross-Linking and Sequencing of Hybrids

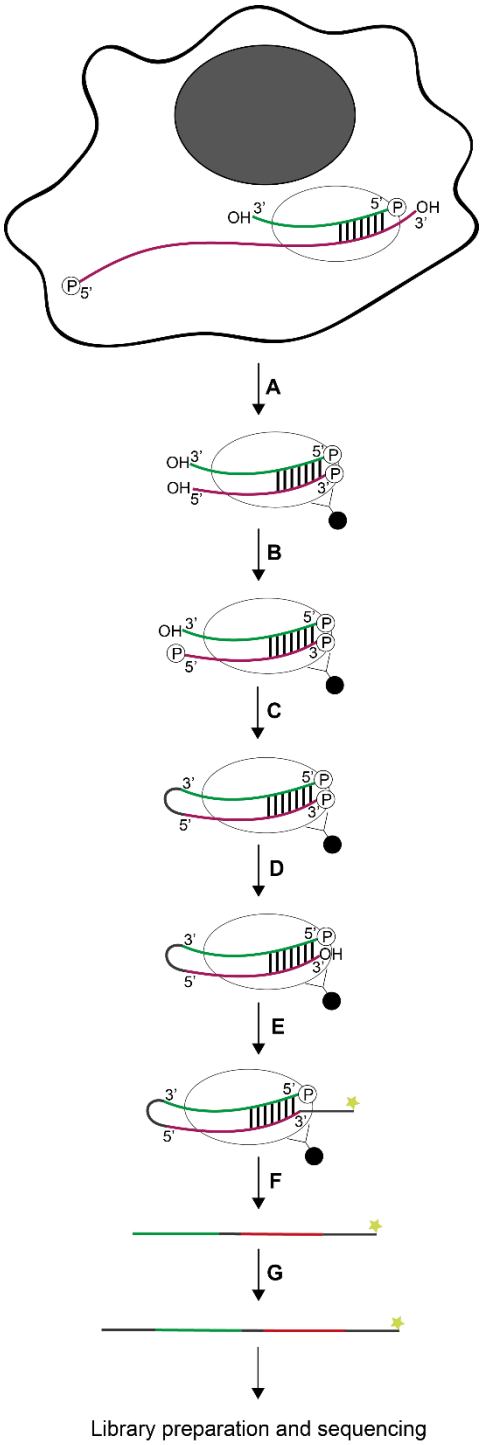


Figure 40. Schematic summary of CLASH. MicroRNA is shown in green, mRNA in red. White ellipse represents AGO2. Black circle and attached Y represent the magnetic beads and the anti-Ago2 bound to them. **(A)** After UV cross-linking of living cells, these are lysed and treated with DNase and RNase, and AGO2 is immuno-precipitated. **(B)** The 5' end of the mRNA is phosphorylated. **(C)** An RNA ligase mediates intermolecular RNA-RNA ligation between the miRNA and the mRNA. **(D)** The 3' end of the mRNA is dephosphorylated. **(E)** Fluorescent 3' adaptor is ligated (in the dark). **(F)** Elution, RNA purification. **(G)** 5'adaptor ligation. And then library preparation and sequencing. More details are provided in the methods section.

The first step of CLASH was the ultraviolet (UV) crosslinking of living cells, which generates a stable interaction between proteins and bound RNAs **(Figure 40)**. Next, after a short RNAse treatment to trim mRNAs bound to AGO2 that leaves an -OH in their 5'end, endogenous AGO2-RNA complexes were immunoprecipitated (IP), using a specific antibody and an IgG as a negative control **(Figure 40A)**. This IPs were performed in crosslinked (+UV) and non-crosslinked (-UV) PA-1 cells. We corroborated that we were able to efficiently and specifically pull down endogenous AGO2 **(Figure 41)**, however we observed a general reduction of protein levels upon UV cross-linking **(Figure 41)**. As expected, IP

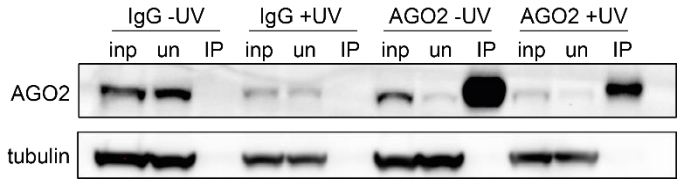


Figure 42. Efficient and specific immunoprecipitation of endogenous AGO2 from PA-1 cells. Western blot for different fractions of the four different IPs are shown. IgG, immunoprecipitation with a negative control, a non-specific antibody. UV (+/-) denotes whether or not PA-1 cells were UV cross-linked before immunoprecipitation. Inp, pre-cleared input (1%); un, unbound fraction (1% after over-night IP); IP, immunoprecipitated fraction (5% ot total IP after CLASH washes).

with IgG did not pull down any AGO2 **(Figure 41)**.

This immunoprecipitated samples were then treated with PNK to phosphorylate the 5' end of the mRNA **(Figure 40B)**, and then an intermolecular RNA-RNA ligation was performed using an RNA ligase, in which we expect the miRNAs to be ligated to the mRNA they are targeting **(Figure 40C)**. Next, the 3' end of the mRNA was dephosphorylated **(Figure 40D)** to enable ligation of the fluorescent 3' adapter **(Figure 40E)**. After ligation of the fluorescent 3' adapter, samples were run in a polyacrylamide gel and visualized. We observe a clear band (LOW) and a smear (HIGH) **(Figure 42)**. The LOW band contains mainly AGO2 bound to miRNAs only, and therefore runs at its usual size of ~100 kDa **(Figure 42)**. However, the HIGH band/s contain AGO2 bound to longer RNAs of different lengths, which cause a shift in AGO2 mobility through the gel **(Figure 42)**. In this band is where we expect to find chimeras (i.e. chimeric reads that correspond to miRNAs that have been ligated to their target mRNAs). Nevertheless, we also extracted and proceeded with the LOW band,

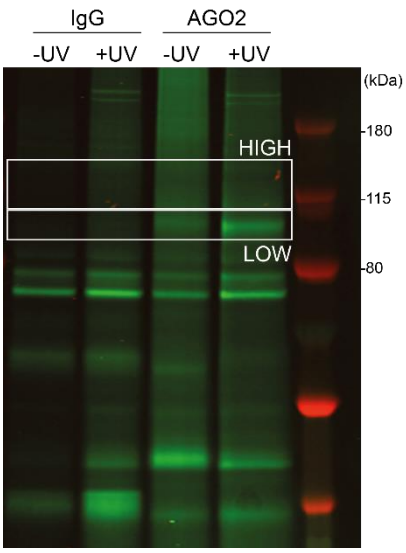


Figure 41. Visualization of the fluorescently-labelled immunoprecipitated RNAs. After ligation of the fluorescent adaptor, samples were separated by size in a polyacrylamide gel. HIGH and LOW bands were cut and processed separately for each lane.

since this will also yield us a valuable information: the landscape of functional microRNAs in embryonic teratocarcinoma cells. Needless to say, each band from each lane was processed independently. As expected, we obtained no labelled RNAs in the IgG control **(Figure 42)**.

The RNA was eluted and purified **(Figure 40F)**, and after ligation of 5' adapter **(Figure 40G)** it was used to prepare a small RNA sequencing library (details in section M.22). Importantly, we used a specific kit that prevents primer dimer formation (Shore *et al.*, 2016). Bioinformatic analysis and chimera identification was performed by Dr. Grzegorz Kudla (Institute for Genetics and Molecular Medicine, Edinburgh, UK), who first developed the technique as well as the bioinformatic analysis pipeline in yeast (Kudla *et al.*, 2011) and then in human cells (Helwak *et al.*, 2013).

On average, we obtained ~10 million mapped reads in the IgG samples, and ~25-35 million mapped reads in the AGO2 samples, and only about 0.01% of those reads corresponded to chimeras. Interestingly, although in low abundance,

Sample	Total Chimeras	microRNA:mRNA	microRNA:transposonRNA
IgG-UV_Low	1010	37	0
IgG-UV_High	8558	0	0
IgGplusUV_Low	1647	3	3
IgGplusUV_High	12774	1	0
AGO2-UV_Low	7178	487	0
AGO2-UV_High	21020	1909	119
AGO2plusUV_Low	9900	550	12
AGO2plusUV_High	9366	324	47

Table 7. MicroRNA:mRNA and microRNA:transposonRNA chimeras are enriched in AGO2 CLASH samples. Each of the three CLASH replicates contained one of the eight samples shown in the table. Numbers correspond to the sum of all replicates. Total number of chimeras, including all types shown in Figure 43, is shown in the second column. Chimeras containing a fragment of microRNA bound to a fragment of mRNA or a fragment of transposon RNA are shown in the third and fourth column, respectively.

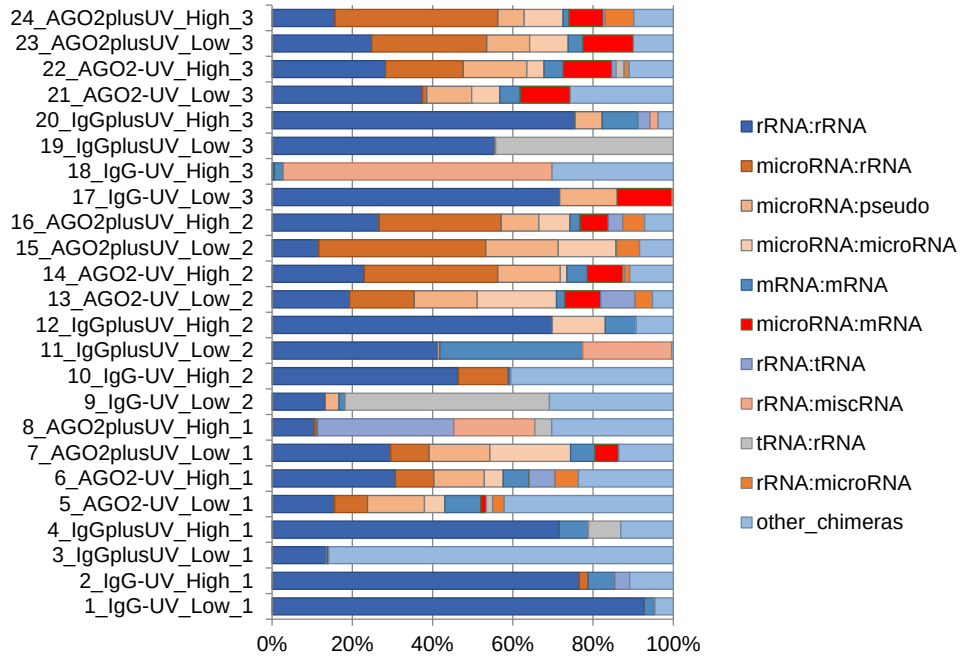


Figure 43. Overview of the abundance of different types of chimeras in all CLASH replicates. The percentage of every type of chimera respect to the total number of chimeras found in each sample is shown. MicroRNA:mRNA chimeras are shown in red. rRNA: ribosomal RNA. Pseudo: pseudogenes. tRNA: transfer RNA. Numbers 1, 2 and 3 in each row denote the different CLASH replicates.

microRNA:mRNA chimeras were found specifically in LOW and HIGH bands corresponding to AGO2 samples **(Table 7)**. Consistently, chimeras containing miRNAs were not frequent in IgG samples **(Figure 43)**. These results suggest that we were able to sequence RNAs specifically associated with AGO2, and probably real microRNA-mRNA interactions. Notably, we found that microRNA:transposonRNA chimeras were also specifically detected in AGO2 samples, although in even lower abundance **(Table 7)**. These results suggest that CLASH enables detection of AGO2-associated chimeras containing miRNAs that could represent *in vivo* targeting, although at a very low efficiency. This could be due to several limiting steps inherent to the technique such as the intermolecular RNA-RNA ligation or the reverse transcription of the chimeric RNAs, which are further discussed in section D.3.

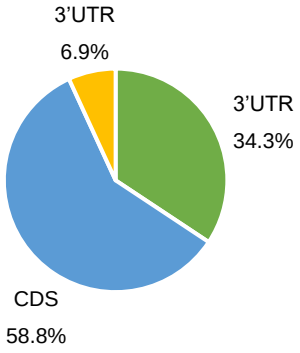


Figure 44. Messenger RNA fragments found in chimeras with miRNAs are located predominantly in the coding sequence of mRNAs. The distribution of mRNA fragments found in chimeras with miRNAs is shown, together with the percentage respect to the total mRNA fragments. Blue: CDS (coding sequence), green: 3'UTR, yellow: 5'UTR.

First, we analyzed the microRNA:mRNA chimeras based on the location of the mRNA fragment and the seed type. From the 102 mRNA fragments found in chimeras with miRNAs, 35 were mapped to 3'UTRs, 60 to coding sequences and 7 to 5'UTRs **(Figure 44)**. Regarding seed type, 11/60 CDS sites, 9/35 3'UTR sites and 0/7 5'UTR sites contained seeds previously characterized as functional (types S8, S7A, S7, S6A or S6). This distribution is similar to the one obtained in a previous CLASH experiment in HEK293T cells (Helwak *et al.*, 2013). Interestingly, 4/9

microRNA (miRBase code)	mRNA (ENST)	Seed bp	Seed type	Folding energy	Published evidence
miR-181b (MIMAT0000257)	SLC2A3 (00000075120)	6	S7	-19,5	-
miR-302a (MIMAT0000684)	ZNF3 (00000299667)	5	S6S	-17,7	-
miR-106a (MIMAT0000103)	MTMR3 (00000333027)	6	S7	-16	+
miR-182 (MIMAT0000259)	VKORC1 (00000354895)	5	S6S	-15,7	-
miR-367 (MIMAT0000719)	ACTB (00000331789)	5	S6S	-14,2	-
miR-182 (MIMAT0000259)	CDC27 (00000441612)	6	S8	-13,3	-
miR-17 (MIMAT0000070)	TMBIM6 (00000423828)	6	S6	-12,9	+
miR-17 (MIMAT0000070)	CENPQ (00000371200)	6	S6	-12,1	+
miR-454 (MIMAT0003885)	MYL12B (00000400174)	6	S7A	-11,7	-
miR-302a (MIMAT0000684)	USP37 (00000258399)	6	S7	-9,3	+

Table 8. Chimeras confidently identified by CLASH in PA-1 cells. Shown, from left to right columns: microRNA and its miRbase code, mRNA and its ENST number, number of Watson-Crick base pairs in the seed region, type of seed (Helwak *et al.*, 2013), folding energy of the predicted pairing, and whether or not there is reported evidence of interaction between these miRNAs and target mRNAs (from miRTarBase). Chimeras are ordered from top to bottom starting from the lowest folding energy, and coloured according to the type of seed (S8, green; S7, red; S6, blue).

chimeras located in the 3'UTR and containing functional seeds have been shown to interact experimentally in miRTarBase, a database for validated microRNA-target interactions (Hsu et al., 2011; Huang et al., 2019) **(Table 8)**, suggesting that a significant proportion of the identified chimeras represent *bona fide* interactions between miRNAs and target mRNAs.

microRNA (miRBase code)	Type of TE	Number of reads
miR-210 (MIMAT0000267)	L1-Ta	3
miR-367 (MIMAT0000719)	L1-Ta	1
miR-190 (MIMAT0000458)	L1-Ta	1
miR-92b (MIMAT0003218)	AluYb8	2
miR-367 (MIMAT0000719)	AluY	1
miR-367 (MIMAT0000719)	AluSq	1
miR-101 (MIMAT0000099)	AluSq	1
miR-106b (MIMAT0000680)	AluSq	1
miR-302b (MIMAT0000715)	AluSq	1

Table 9. Chimeras containing transposable elements' RNA identified by CLASH in PA-1 cells. Shown, from left to right columns: microRNA and its miRbase code, type of TE to which the RNA fragment belongs, and number of reads containing each chimera. Chimeras are coloured according to the type of TE found (L1, yellow; Alu, blue)

Lastly, we searched the sequencing data for chimeras containing mRNA from L1s or other active human TEs such as Alu or SVA. We only identified 3 different chimeras containing L1 RNA and miRNAs: miR-210 (3 reads), miR-367 and miR-190 **(Table 9)**. These L1s belong to the highly active Ta family (see section I.4.A). Additionally, we found 2 different chimeras containing AluY RNA and miRNAs (miR-92b (2 reads) and miR-367) and 4 containing AluS RNA and miRNAs (miR-367, miR-101, miR-106b and miR-302b) **(Table 9)**. As expected, given that the 3'UTR of L1 is extremely short, all L1 sequences in chimeras are located in CDS. However, as we show in sections R.1 to R.5, miRNAs can target CDS of L1, restricting their mobilization.

Next steps in this project will be to validate these interactions, and test whether these miRNAs control L1 retrotransposition taking advantage of our recently published sRNA/L1 retrotransposition assay (Tristan-Ramos *et al.*, 2020). Furthermore, we will further optimize it to test engineered Alu retrotransposition, and whether miRNAs can control mobilization of this nonautonomous

TE. However, the limited number of identified chimeras suggest that CLASH still requires significant optimization to enable comprehensive detection of miRNAs targeting transposable elements.

DISCUSSION

D.1. The tumor suppressor miRNA Let-7 inhibits L1 retrotransposition

Multiple studies have linked LINE-1 retrotransposons to cancer, specially over the past decade (reviewed in (Burns, 2017) and (Scott and Devine, 2017), and in section I.7 of this Thesis). Particularly, L1 insertions have been found to occur at high frequencies in lung cancer genomes (Iskow *et al.*, 2010), as well as in other epithelial cancers such as esophageal or colorectal (Rodriguez-Martin *et al.*, 2020). Novel L1 insertions (or L1-mediated Alu or SVA insertions) can cause alteration in splicing, disruption of exons, or mediate large genomic rearrangements (section I.5). Unsurprisingly, retrotransposition is associated with genomic instability and human disease (Kazazian and Moran, 2017; Burns, 2020). An open question subjected to active research remains in the field: how are these L1 elements silenced and derepressed in somatic human tissues? And, furthermore: how do these processes impact tumorigenesis?

It is known that DNA methylation of the L1 promoter is the first line of defence to restrict L1 activity (Bourc'his and Bestor, 2004; Sanchez-Luque *et al.*, 2019). In fact, a correlation between increased somatic L1 insertions and hypomethylation of L1 promoters has been shown, not only at a global level (Tubio *et al.*, 2014) but also at specific loci (Sanchez-Luque *et al.*, 2019). Besides, hypomethylation of specific L1s can activate retrotransposition early in tumorigenesis (Scott *et al.*, 2016). In lung cancer patients, hypomethylation of the L1 promoter has been associated with worse prognostic (Saito *et al.*, 2010). However, taking into account the high level of somatic L1 activity observed in some of these patients, it is tempting to speculate that active L1s might also escape other post-transcriptional restriction mechanisms (Goodier, 2016; Pizarro and Cristofari, 2016).

On the other hand, miRNAs are deregulated in cancer (Lu *et al.*, 2005), and can be classified into tumor suppressor miRNAs and oncomiRs (Esquela-

Kerscher and Slack, 2006; Ventura and Jacks, 2009; Lujambio and Lowe, 2012) (section I.15). The formers regulate oncogenes and are downregulated in tumors, whereas the latter promote cell proliferation by regulating tumor suppressor genes, and are frequently upregulated in cancer. Amongst all tumor suppressor miRNAs, let-7 is the most frequently downregulated miRNA in cancer, and this typically correlates with poor prognosis (Nair, Maeda and Ioannidis, 2012). On a functional level, it has been well characterized that a decrease in let-7 miRNAs causes overexpression of their target oncogenes, such as MYC, RAS or HMGA2 (Roush and Slack, 2008; Powers *et al.*, 2016).

In this Thesis, we describe a novel tumor-suppressor role for let-7: restricting L1 retrotransposition contributing to the maintenance of genome integrity.

We began our study by analysing somatic L1 insertions in human NSCLC* samples using MELT (Gardner *et al.*, 2017), and found a high frequency of retrotransposition, in line with previous and recent reports from other groups (Iskow *et al.*, 2010; Helman *et al.*, 2014; Rodriguez-Martin *et al.*, 2020). Furthermore, we found that the samples with somatic L1 insertions presented reduced expression of several members of the let-7 family of miRNAs, as well as reduced levels of miR-34a, which is, like let-7, a well-known tumor suppressor miRNA **(Section R.1)**. Notably, we reproduced these results when using the somatic L1 insertion number calculated by a different pipeline called Transpo-Seq (Helman *et al.*, 2014) **(Section R.1)**.

Next, we developed an assay to study the effect of miRNAs in engineered L1 retrotransposition in cultured human cells (the sRNA/L1 retrotransposition assay (Tristan-Ramos *et al.*, 2020)), and used it to demonstrate that let-7 can negatively regulate L1 mobilization in a wide variety of cultured human cells, including

* Non-small cell lung cancer

lung cancer cells **(section R.2)**. It is worth noting that the expression of miR-34a, another tumor suppressor miRNA, is also reduced in lung tumor samples with accumulated L1 insertions. Furthermore, it inversely correlates with L1Hs RNA levels. However, under these experimental conditions, we did not observe a consistent effect of miR-34a in L1 retrotransposition. We hypothesise that miR-34 might target a member of the epigenetic regulatory network that controls L1Hs expression, and thus could act as an indirect regulator of L1 mobilization.

The effect of let-7 in engineered L1 mobilization could be explained by two non-exclusive ways. On one hand, let-7-guided RISC could be directly interacting with L1 mRNAs. On the other hand, let-7 could be regulating any host factor required for the retrotransposition cycle (Taylor *et al.*, 2013, 2018), thus regulating L1 mobilization in an indirect manner. Since the main effectors of miRNA-mediated gene silencing are AGO proteins (Meister, 2013), we performed an AGO2 RNA immunoprecipitation as well as sRNA/L1 retrotransposition assays (Tristan-Ramos *et al.*, 2020) with constructs lacking 3' or 5'UTR. The results suggest that let-7 is actually guiding AGO2 to the coding sequence (CDS) of human L1 mRNA **(Section R.3)**. This is in agreement with previous studies reporting co-localization of L1-ORF1p/L1-ORF2p and AGO2 in various cell lines (Goodier *et al.*, 2007; Doucet *et al.*, 2010). In fact, we further found and experimentally validated a non-canonical offset 7-mer site in the CDS (specifically in ORF2) of L1 mRNA **(Section R.4)**. MiRNAs often bind to their target mRNAs in the 3'UTR, however CDS binding sites have also been described and validated experimentally (Hausser and Zavolan, 2014; Jonas and Izaurralde, 2015; Zhang *et al.*, 2018). To highlight one particular example, let-7 targets DICER within its coding sequence (Forman, Legesse-Miller and Coller, 2008). Preferential occurrence of MicroRNA Responsive Elements (MREs) in the CDS of mRNAs with short 3'UTR, such as L1 mRNA, has been described (Reczko *et al.*, 2012). We speculate that let-7 (and other microRNAs) could have evolved to target L1

within its CDS and specifically in ORF2, the largest and most conserved region in L1. Indeed, the binding site that we discovered and validated is conserved in young L1 subfamilies (L1PA1-L1PA5) and only presents subtle variations in older L1 subfamilies (L1PA6-L1PA16) **(Section R.4)**. Supporting this hypothesis, the binding site of the only other miRNA described to target L1 (miR-128) is also located in ORF2 (Hamdorf *et al.*, 2015). Regarding the characteristics of this offset 7-mer site, they have been previously described an experimentally validated in humans (Kim *et al.*, 2016) and even in mosquitoes using a variant of CLASH to analyze endogenous miRNA-mRNA interactions (Fu *et al.*, 2020). Nevertheless, it is important to mention that these non-canonical sites, while functional, are not as efficient as canonical 8-mer or 7-mer sites (Section I.12).

Importantly, even though the interaction between let-7 miRNA and L1 mRNA likely occurs in any cell that simultaneously expresses both RNAs, we did not observe any correlation between increased accumulation of tumor-specific L1 insertions and reduced let-7 levels in breast cancer **(Section R.1)**. We speculate that, in some cell types, other transcriptional or post-transcriptional regulatory layers suppressing L1 mobilization may overshadow miRNA-mediated L1 inhibition. In fact, L1 reactivation occurs less frequently in breast cancer compared to other types of tumors such as non-small cell lung cancer (Helman *et al.*, 2014; Tubio *et al.*, 2014; Rodriguez-Martin *et al.*, 2020)

Interestingly, several mRNAs can undergo alternative cleavage and polyadenylation in cancer, which results in shorter 3'UTRs that lack miRNA binding sites, making these genes more 'resistant' to repression via miRNAs (Mayr and Bartel, 2009). Besides, is has been reported that proliferating human and mouse cells express shortened 3'UTRs that contain fewer miRNA binding sites (Sandberg *et al.*, 2008). Overall, these and other publications suggest that 3'UTR targeting allows a context-dependent regulation, whereas CDS targeting enables a context-independent regulation. From an evolutionary perspective, it seems

more sensible for human cells to develop ways to control L1 mobilisation regardless of the developmental or disease context, given the potential genomic instability that L1s can generate upon retrotransposition (Richardson *et al.*, 2015; Kazazian and Moran, 2017; Burns, 2020). Accordingly, our lab previously showed that knocking down the Microprocessor increased to the same extent the retrotransposition of a LINE-1 lacking the 3'UTR and that of a full-length element (Heras *et al.*, 2013).

MicroRNAs with binding sites in the CDS of their target mRNAs tend to mainly repress translation, whereas miRNAs targeting the 3'UTR region mainly induce deadenylation and mRNA degradation (Brümmer and Hausser, 2014; Hausser and Zavolan, 2014; Jonas and Izaurralde, 2015). Indeed, previous studies have suggested that sites located in CDS preferentially inhibit translation regardless of their structure and sequence (Hausser *et al.*, 2013). Moreover, a recent report showed that CDS sites require extensive 3' pairing, such as the binding site that we validated, and induce ribosome stalling instead of mRNA degradation (Zhang *et al.*, 2018). Consistently, we have observed that let-7 impairs translation of L1-ORF2p without affecting L1 mRNA levels **(section R.5)**. This is further supported by the lack of correlation between the levels of let-7 and the expression of L1Hs RNA observed in lung tumor samples **(section R.1)**. Interestingly, the accumulation and translation of an L1 mRNA variant in which the natural ´bs2´ site was substituted by a canonical let-7 8-mer ('8mer') site is similarly affected by let-7 overexpression as wild-type molecules **(section R.5)**, suggesting that binding to the CDS region itself rather than the structure of base-pairing mediates translational repression, as previously described for other miRNA targeting CDS sites (Hausser *et al.*, 2013). Despite being essential for L1 retrotransposition (Moran, Holmes and Naas, 1996), L1-ORF2p is expressed as a very low level (Alisch *et al.*, 2006; Ardeljan *et al.*, 2019). Therefore, a small reduction in the

abundance of L1-ORF2p could have a strong impact in RNP formation and in L1 retrotransposition.

Considering that there are many cellular factors involved in the L1 retrotransposition cycle (Taylor *et al.*, 2013, 2018), and that let-7 can potentially target hundreds of human mRNAs (Friedman *et al.*, 2009; Bartel, 2018), we cannot discard additional redundant indirect effects that complement the one described in this Thesis to control L1 mobilization. It seems reasonable that, besides targeting L1 mRNA, let-7 could as well regulate other mRNAs encoding proteins that positively impact human L1 retrotransposition. This is supported by our finding that mutations of the ORF2 binding site reduce, but not abolish, the effect of let-7 in L1 mobilization **(section R.4)**. Another possibility is that additional let-7 binding sites exist within CDS of human L1 mRNA. However, we failed to experimentally validate other two binding sites: one located in a different region of ORF2 that was predicted by miRanda **(section R.4)**, and other located in in L1-ORF1 identified using local alignments for let-7 (data not shown). Therefore, alternative approaches such as CLASH might be needed to unbiasedly identify other let-7 binding sites within L1 mRNA. Nevertheless, the CLASH presented in this Thesis **(section R.8)**, discussed in section D.3, was performed in human embryonic teratocarcinoma cells that have a high endogenous expression of L1 mRNAs but almost no endogenous expression of let-7 **(section R.2)**.

There is another question that remains unanswered, and it concerns the role that this regulation may have in other non-pathological processes such as differentiation. L1s are highly expressed in embryonic stem cells and its expression decreases as differentiation advances (Garcia-Perez, Widmann and Adams, 2016; Faulkner and Garcia-Perez, 2017), the opposite pattern of let-7 miRNA expression (Pasquinelli *et al.*, 2000; Kloosterman and Plasterk, 2006; Roush and Slack, 2008). Additionally, it has been described that, as embryonic development progresses, miRNAs effect on target mRNAs switches from translational repression

to mRNA degradation (Subtelny *et al.*, 2014). Given that our results show that let-7 impairs L1-ORF2p translation without altering mRNA levels, it would be interesting to gain insight into the possible role of let-7 controlling L1 mobilization during early embryonic development differentiation until complete methylation of the L1 promoter, the main mechanism to restrict endogenous L1 mobilization in somatic cells (Goodier, 2016), is achieved.

Lastly, another question one might ask is whether there are more miRNAs that regulate LINE-1 retrotransposons. Experimental approaches such as CLASH, which allow unbiased identification of miRNA-mRNA interactions (Helwak *et al.*, 2013), should enable further research in this topic. This will be discussed in section D.3. Additionally, high-throughput studies with large sample size from different cancer types to analyze somatic L1 insertions as well as miRNA expression, similar to the study presented in **section R.1**, could also be useful to identify novel roles for these small regulators with such a big role in biology.

The interaction between let-7 and L1 mRNAs could also be viewed from a different perspective. Even if there was only one let-7 binding site in this RNA, its high abundance could make it function as a let-7 sponge (Ebert, Neilson and Sharp, 2007), helping to maintain the embryonic phenotype where L1s are highly expressed and let-7 miRNAs are not, and complementing the activity of LIN28 proteins (Viswanathan, Daley and Gregory, 2008; Viswanathan and Daley, 2010). Interestingly, this let-7 sponging effect has been described for the H19 lncRNA (Kallen *et al.*, 2013).

Let-7 was the first miRNA discovered in humans (Pasquinelli *et al.*, 2000), and is one of the most highly conserved miRNAs in metazoans, involved in multiple fundamental biological processes, including differentiation, metabolism, and cancer (Roush and Slack, 2008). The findings described in **sections R.1-5** support a model in which let-7 guides RISC to the CDS of active L1s' mRNA and impairs

the translation of L1-ORF2p. This alters the ratio between L1-ORF2p and L1-ORF1p, unbalancing RNP formation and, ultimately, reducing L1 retrotransposition **(Figure 45)**. Let-7 miRNAs are expressed in differentiated cells (Roush and Slack, 2008), where plenty of mechanisms function to repress L1 expression and prevent somatic L1 insertions (Goodier, 2016; Pizarro and Cristofari, 2016). We speculate that alterations in let-7 expression in human cancers, such as lung cancer, results in increased retrotransposition of actively transcribed L1s, as well as L1-mediated mobilization of non-autonomous elements such as Alu and SVA. This increased retrotransposition potentially contributes to tumor progression by genomic instability **(Figure 45)**.

Further study of the regulation of endogenous L1 retrotransposition is granted using the lung cancer cellular model described in this Thesis **(Section R.6)**. Let-7 is downregulated in lung tumor samples that present somatic L1 insertions **(Section R.1)**. Briefly, we have generated a stable SK-MES-1 cell line in which effective expression of mature let-7 can be induced with doxycycline. Combination of long-term culture of these cells in presence or absence of doxycycline with analysis of WGS data from different passages by MELT (as described in **section R.1**) will enable us to gain insight into the effect of restoring endogenous let-7 levels in controlling endogenous L1 mobilization. We speculate that one of the tumor-suppressor effects of let-7 miRNAs could be restricting the mobility of active retrotransposons, therefore reducing the genomic impact of new insertions (see Section I.5).

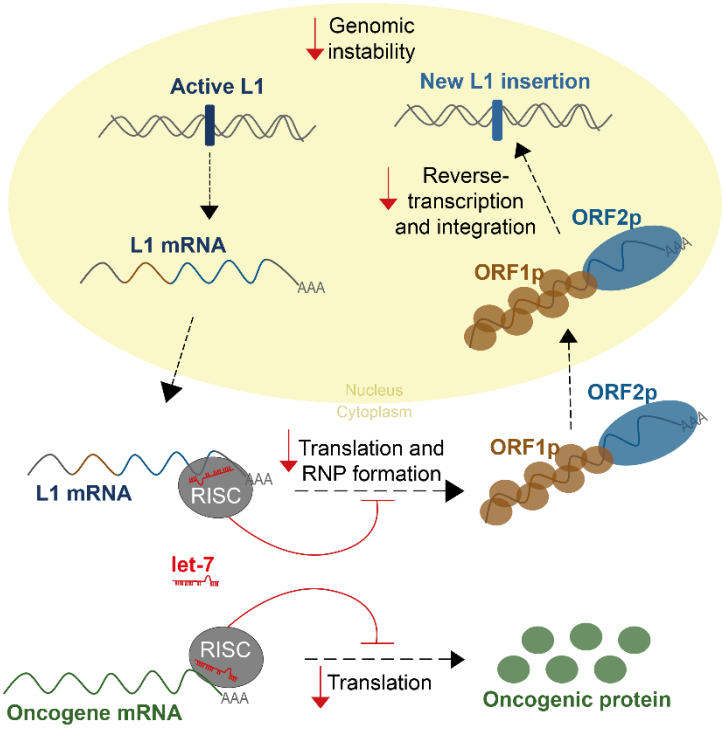


Figure 45. A new role for the let-7 family of tumor suppressor miRNAs: guardians of genome stability by restriction of active LINE-1 retrotransposons. Besides the well-known regulation of oncogenes (lower part of the scheme), we propose a novel tumor-suppressor role for let-7 microRNAs: binding the mRNA of active LINE-1 retrotransposons, impairing L1-ORF2p translation and, overall, reducing L1 mobilization to protect genome integrity. Consequently, downregulation of let-7 in cancer could be one of the mechanisms leading to increased somatic retrotransposition.

D.2. miR-20, a potential indirect activator of L1 mobilization

As described in section D.1, we used MELT (Gardner *et al.*, 2017) to analyze somatic L1 insertions in our NSCLC samples, and those samples with L1 insertions presented reduced levels of two tumor suppressor miRNAs: let-7 and miR-34 (Section R.1). However, when we repeated the statistical analysis considering the NSCLC samples analyzed by Helman and col. using by Transpo-Seq (Helman *et al.*, 2014), we observed another miRNA whose levels were significantly different between the samples with and without insertions: miR-20. Interestingly, this correlation was the opposite to what would be expected for a miRNA regulating its target (in this case L1): samples with increased L1 insertions presented significantly more expression of miR-20 **(section R.1)**. Using the assay that we developed to study the effect of miRNAs in regulating L1 mobilization (sRNA/L1 retrotransposition assay (Tristan-Ramos *et al.*, 2020)), we corroborated this direct correlation experimentally: depletion of miR-20 reduced engineered L1 retrotransposition in cultured human cells **(section R.7)**.

The microRNA miR-20 is generally considered an oncomiR (Esquela-Kerscher and Slack, 2006) and is part of the oncogenic miR17/92 cluster (Mogilyansky and Rigoutsos, 2013). Therefore, a potential explanation for our results could be that increased expression of miR-20 occurs simultaneously but independently of the increase of somatic L1 insertions in lung cancer. However, miR-20 could also be indirectly regulating L1 retrotransposition through regulation of proteins that restrict L1 mobilization. A recent genome wide CRISPR/Cas9 screening characterized thousands of genes as potential L1 repressors, as their genetic depletion via CRISPR/Cas9 increased L1 mobilization in K562 or HeLa cells (Liu *et al.*, 2018). Amongst these potential L1 repressors, we searched for miR-20 targets and found 7 that had been experimentally validated (Trompeter *et al.*, 2011), and 15 that were predicted to be miR-20 targets by three independent target prediction softwares (microrna.org, miRdb.org and

TargetScan.org). Interestingly, 17 of those 22 genes were classified as potential L1 repressors in the aforementioned genome-wide study **(section R.7)**. Further studies are needed to understand the potential role of miR-20 in L1 retrotransposition, and whether the upregulation of miR-20 in cancer can indirectly lead to increased somatic L1 insertions, and therefore to more genomic instability. Given the growing number of cellular factors involved in the post-transcriptional control of L1 mobilization (Goodier, 2016; Pizarro and Cristofari, 2016), we find reasonable to speculate that miR-20 may regulate one or more of these factors, indirectly resulting in increased L1 mobilization in cancer **(Figure 46)**.

Interestingly, gene ontology analysis showed that proteins involved in cell cycle regulation were enriched in the previous analysis of miR-20 targets **(Section R.7)**, consistent with recent reports suggesting that some cell cycle regulators such as p21 or p27 repress L1 retrotransposition (Kawano *et al.*, 2018). The relation between cell cycle, cell division and L1 retrotransposition is a field of active research. Some studies suggest that it does not require cell division (Kubo *et al.*, 2006; Macia *et al.*, 2017) while other suggest that cell division is required, or at least enhances retrotransposition (Xie *et al.*, 2013; Mita *et al.*, 2018). We speculate that miR-20 might target one of these cell cycle regulators that repress L1 mobilization, and therefore increased levels of miR-20 would indirectly lead to accumulation of L1 insertions in tumors.

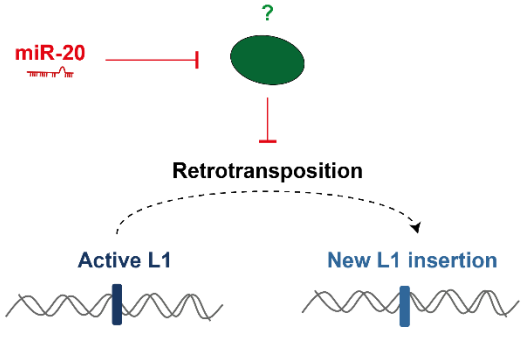


Figure 46. Model of indirect regulation of L1 retrotransposition mediated by miR-20. Many cellular factors (shown in green) control active L1 retrotransposition and generation of new L1 insertions. Based on our results, we speculate that miR-20 (shown in red) may regulate one of these cellular proteins involved in controlling L1 retrotransposition. Therefore, an increase in miR-20, frequently observed in many types of cancer, could reduce the levels of this regulator, indirectly resulting in increased levels of somatic L1 mobilization.

D.3. Identification of endogenous miRNA-L1 chimeras

Bioinformatic prediction of microRNA binding sites, despite notable advances in the past decade, remains challenging, and many predicted binding sites do not behave as such experimentally (Bartel, 2009). Additionally, despite canonical seeds are well characterized (section I.12), non-canonical sites can also mediate post-transcriptional regulation of gene expression (Kim *et al.*, 2016), adding another layer of complexity to target prediction. To overcome these issues, several techniques have recently been developed to enable unbiased discovery of new miRNA targets by analysing miRNA-mRNA interactions *in vivo* (section I.13). The first such approach is termed CLASH (Helwak *et al.*, 2013). Basically, CLASH relies on cross-linking of miRNAs and mRNAs to AGO proteins, immunoprecipitation of AGO, and intermolecular RNA-RNA ligation between the miRNA and the mRNAs in close vicinity. After small RNA library preparation and sequencing, chimeric reads containing miRNA and mRNA provide valuable information about which specific miRNA is interacting with each mRNA *in vivo* (Helwak *et al.*, 2013).

One of the main objectives of this Thesis is to identify miRNAs that regulate active retrotransposons. For this reason, during my stay at the Institute for Immunity and Infection Research (University of Edinburgh, UK) we have performed CLASH in a human embryonic teratocarcinoma cell line, PA-1, that presents high endogenous expression of LINE-1 retrotransposons but retains a stable karyotype **(section R.8)**. We immunoprecipitated endogenous AGO2, and prepared the small RNA sequencing library using a technology that avoids the generation of primer dimers (Shore *et al.*, 2016). Bioinformatic analysis was performed by our collaborator Dr. Greg Kudla, who originally developed this technique (Kudla *et al.*, 2011).

Interestingly, we were able to detect chimeras containing miRNAs bound to mRNAs or to transposon RNAs specifically in the AGO2 samples, and not in

the IgG samples. This suggest that we were able to sequence RNAs specifically associated with AGO2, and therefore the identified chimeras could represent *in vivo* targeting. However, the extremely limited yield of chimeras (~0.01% of all mapped reads) suggest that CLASH still requires significant optimization to generate a comprehensive map of the miRNA-mRNA interactions occurring in human cells. Nevertheless, we were able to identify several interactions for which experimental evidence was available in miRTarBase a database for validated microRNA-target interactions (Hsu et al., 2011; Huang et al., 2019). Next, we looked for chimeras containing sequences of active retrotransposons, and found chimeras containing L1-Ta RNA and AluS or AluY RNA, all members of active subfamilies. Next experiments will include our recently developed sRNA/L1 retrotransposition assay (Tristan-Ramos *et al.*, 2020), in order to experimentally validate whether these miRNAs regulate L1 retrotransposition in cultured cells, and optimization of this technique in order to test Alu retrotransposition *in trans*.

There are several limiting steps that could explain the low abundance of chimeras in our experiments. First of all, the RNA-RNA ligation that generates the chimeric RNAs is a highly inefficient process, requiring optimized conditions that are difficult to achieve in a cell lysate. Next, ligation of the 3' or 5' adapter is expected to have a 50% efficiency (Shore *et al.*, 2016), meaning that the probability of a chimeric RNA being ligated to the two adapters is 25%. Lastly, library generation requires a reverse transcription step. Cross-linked RNAs still have small peptides bound to them that limit reverse transcriptase efficiency. Considering that chimeras contain not one but two cross-linked peptides (one to the miRNA and other to mRNA/transposonRNA), many chimeras will probably not be reverse transcribed efficiently and will not be present in the sequencing library. In sum, we consider CLASH a technique with immense potential but with considerable room for improvement. Remarkably, several modifications are being

developed to overcome these limitations (Lee and Ule, 2018), which could lead to more efficient protocols in the near future, opening new lines of research.

One of these research fields, that remains largely unexplored, is the generation of miRNAs derived from transposable elements. In the past decade, it was proposed that TEs might have generated precursor microRNA genes upon insertion in the vicinity of another TE (Piriyapongsa, Mariño-Ramírez and Jordan, 2007). Recently, it has been reported that mature microRNAs can originate from LINE-2 elements which are also frequent in 3'UTRs, suggesting that ancient LINEs may have generated novel regulatory networks (Petri *et al.*, 2019). An improved, optimized CLASH could enable us to identify not only microRNAs regulating TEs, but also microRNAs derived from TEs that regulate the expression of cellular mRNAs. This is currently being further investigated, as the analysis of short reads corresponding to repeated sequences remains challenging. Nevertheless, the hypothesis of transposable elements being involved in the evolutionary origin of microRNAs is an exciting area of research that could result in striking discoveries in the near future.

Lastly, an optimized CLASH could be performed in tumor samples in which retrotransposition is known to occur at high frequencies (Burns, 2017; Scott and Devine, 2017), to identify miRNAs bound to L1 mRNAs *in vivo*. MicroRNAs are emerging as potential therapeutic tools, not only for cancer but for other diseases (Rupaimoole and Slack, 2017; Gilles and Slack, 2018). We speculate that one of the antitumoral consequences of restoring miRNA expression in cancer would be the restriction of active retrotransposons, maintaining genome instability and potentially reducing the aggressiveness of these tumors.

CONCLUSIONS

The main conclusions of this Thesis are:

1. The tumor suppressor miRNA let-7 binds directly to an offset 7mer site in ORF2 region of the L1 mRNA, impairs L1-ORF2p translation, and overall reduces L1 retrotransposition.
2. Downregulation of let-7 seems to contribute to the accumulation of somatic L1 insertions in human lung cancer.
3. The sRNA/L1 retrotransposition assay allows testing whether a miRNA regulates L1 mobilization in cultured human cells in a consistent and reproducible manner.
4. SK-MES-1/let-7a3 cells allow inducible expression of pri-let-7a3, and can be used to study the role of let-7 in the regulation of endogenous L1 retrotransposition.
5. Cross-Linking and Sequencing of Hybrids (CLASH) is a technique with immense potential to identify miRNAs regulating TEs in human cells, but still requires significant optimization to overcome limiting steps that result in a reduced yield of chimeric reads.

CONCLUSIONES

Las principales conclusiones de esta Tesis son:

1. El miARN supresor de tumores let-7 se une de forma directa a un sitio de tipo 7mer desplazado en la región ORF2 del ARNm de L1, disminuye la traducción de L1-ORF2p y reduce la retrotransposición de L1.
2. La disminución de los niveles de let-7 parece contribuir a la acumulación de inserciones somáticas de L1 en cáncer de pulmón humano.
3. El ensayo de retrotransposición 'sRNA/L1' permite comprobar si un miARN regula la movilización de L1 en células humanas en cultivo de forma consistente y reproducible.
4. Las células SK-MES-1/let-7a3 permiten la expresión inducible de pri-let-7a3, y pueden ser usadas para estudiar el papel de let-7 en la regulación de la retrotransposición endógena de L1.
5. La técnica 'Cross-Linking and Sequencing of Hybrids' (CLASH) tiene un gran potencial para identificar miARNs que regulen elementos móviles en células humanas, pero requiere de una considerable optimización para aumentar el número de quimeras identificadas, muy reducido en el actual protocolo.

MATERIALS & METHODS

M.1. Cell culture

HEK293T (DuBridge *et al.*, 1987), PA-1 (Zeuthen *et al.*, 1980) and U2OS (Ponten and Saksela, 1967) cells were obtained from ATCC (American Type Culture Collection). HeLa JVM cells were provided by John V. Moran (University of Michigan, US). Lung cancer cell lines (A549, SK-MES-1, H520 and H2170) were provided by Pedro Medina (University of Granada and GENYO, Spain). Stable Flp-In-293 cells expressing T7-tagged L1-ORF1p had been generated in the lab (MacLennan *et al.*, 2017).

HEK293T, U2OS, HeLa JVM, A549, SK-MES1 and Flp-In-293 cells expressing T7-tagged L1-ORF1p cells were cultured in high-glucose Dulbecco's Modified Eagle's Media (DMEM, Gibco) supplemented with GlutaMAX, 10% foetal bovine serum (FBS, Gibco) and 100 U/mL penicillin-streptomycin (Invitrogen).

PA-1 cells were cultured in Minimal Essential Medium (MEM, Gibco) supplemented with GlutaMAX, 10% heat-inactivated FBS (Gibco), 1% (100 U/mL) penicillin-streptomycin (Invitrogen) and 1% (0.1 mM) Non-Essential Amino Acids (Gibco).

H520 and H2170 cells were cultured in Roswell Park Memorial Institute-1640 Medium (RPMI, Gibco) supplemented with GlutaMAX, 10% foetal bovine serum (FBS, Gibco) and 100 U/mL penicillin-streptomycin (Invitrogen).

All cells were maintained in humidified incubators at 37ºC with 5-7% CO2, and passaged using standard methods. TrypLE (Gibco) was used for PA-1 cells, and 0.05% Trypsine (Gibco) for the rest of the cell lines. Absence of Mycoplasma spp. in cultured cells was confirmed once a month by a PCR-based assay (Minerva).

Freezing and unfreezing of cell lines was performed following standard procedures. For freezing, cells were trypsinised, pelleted at 200g for 4min, resuspended in FBS supplemented with 10% DMSO (Sigma) and placed at -80ºC. Within one month, they were transferred to liquid nitrogen for long-term storage.

For defrosting, vials were put in a water bath at 37ºC. Upon complete thaw, vial content was transferred to a 15mL tube containing 10mL of pre-warmed medium. The tube was then centrifuged for 4 min at 200g, supernatant containing DMSO removed, and cell pellet resuspended in an adequate volume of pre-warmed medium and transfer to a culture flask.

M.2. Antibiotic stocks

- Neomycin/G418: geneticin sulphate (Gibco) was diluted in milli-Q water to a final concentration of 40mg/mL, protected from light, and kept at 4ºC for a maximum of one month.
- Blasticidin: blasticidin-S (Thermo) was diluted in milli-Q water to a final concentration of 10mg/mL, and stored at -20ºC.
- Ampicillin: ampicillin (Sigma) was diluted to 100mg/mL in milli-Q water, and stored at -20ºC
- Puromycin: puromycin dihydrochloride 10mg/mL (Thermo) was aliquoted and stored at -20ºC
- Hygromycin: hygromycin B 50mg/mL (Thermo) was stored at 4ºC protected from light.

M.3. Bacterial transformation

Ultra-competent E. coli DH5α cells were incubated with DNA on ice for 15min, heat-shocked for 45sec at 42ºC and put back on ice for 2min. Pre-warmed LB was then added and bacteria were allowed to recover for 10-30min at room temperature before plating them in LB-agar-ampicillin plates. When DNA was to be digested by a methylation-sensitive restriction enzyme, INV110 bacteria were used following the same procedure.

M.4. Plasmids

All plasmids used for transfections were purified using Qiagen Plasmid Midi Kit (Qiagen). DNA was analyzed by electrophoresis in 1% agarose gels and only supercoiled DNA preparations were used. For any other application, plasmids were purified using WizardPlus SV Minipreps DNA Purification System (Promega). All enzymes used for cloning were purchased from New England Biolabs. Descriptions will be divided between plasmids used for retrotransposition assays and those used for other experiments.

M.4.A. Plasmids for retrotransposition assays

All these constructs contain a reporter cassette that is activated upon retrotransposition. All reporter cassettes are located in the 3'UTR of an active LINE, and consist of a reporter gene, a heterologous promoter, and a polyadenylation signal, all in antisense orientation with respect to the LINE. The reporter gene is interrupted by an intron of the γ-globin gene in sense orientation respect to the LINE (antisense respect to the reporter). This arrangement ensures that expression of the reporter can only be achieved after a round of successful retrotransposition (i.e. transcription from the LINE promoter, splicing of the intron, reverse transcription and integration in a different genomic location). An example is shown in **Figure 12, Section R.2**. The different reporter cassettes used in this study are:

- mneoI: previously described (Freeman, Goodchild and Mager, 1994), contains the neomycin phosphotransferase gene.
- mblastI: previously described (Morrish *et al.*, 2002), contains the blasticidin-S deaminase gene.
- megfpI: previously described (Ostertag *et al.*, 2000), contains the enhanced GFP gene.

- mflucI: previously described (Xie *et al.*, 2011), contains the firefly luciferase gene.

The following plasmids have been used for retrotransposition assays (unless otherwise indicated, L1 expression is driven by a CMV* promoter):

- JM101/L1.3: previously described (Sassaman *et al.*, 1997), contains a full-length human L1.3 element (accession number #L19088) tagged with the *mneoI* reporter cassette and cloned in pCEP4 (Life Technologies).
- JM101/bs2mut: a derivative of JM101/L1.3 in which mutations in the let-7 binding site located in ORF2 (bs2) have been generated using site-directed mutagenesis. Further details of the cloning procedure are provided in section M.11.
- JM105/L1.3: previously described (Wei *et al.*, 2001), a derivative of JM101/L1.3 that contains a missense mutation in the RT domain of L1-ORF2p (D702A).
- pCEP-TG$_F$21: previously described (Goodier *et al.*, 2001), contains a full-length mouse LINE-1 G$_F$21 element (L1Md- G$_F$21, accession number #AC021631.6, positions 62229-68991) tagged with the *mneoI* reporter cassette and cloned in pCEP4 (Life Technologies).
- Zfl2-2mneoI: previously described (Sugano, Kajikawa and Okada, 2006), contains a full-length zebrafish Zfl2-2 tagged with the *mneoI* reporter cassette inside the 3'UTR of the element and cloned in pCEP4 (Life Technologies).
- Zfl2-1mneoI: previously described (Sugano, Kajikawa and Okada, 2006), contains a full-length zebrafish Zfl2-1 tagged with the *mneoI* reporter cassette inside the 3'UTR of the element and cloned in pCEP4 (Life Technologies).
- pU6iNeo: previously described (Richardson *et al.*, 2014), contains the neomycin phosphotransferase (NEO) expression cassette from pEGFP-N1 (Clontech)

* Cytomegalovirus

cloned into a modified pBSKS-II(+) (Stratagene) that contains a U6 promoter in the multi-cloning site.

- JJ101/L1.3: previously described (Kopera *et al.*, 2011), contains a full-length human L1.3 element (accession number #L19088) tagged with the *mblastI* reporter cassette and cloned in pCEP4 (Life Technologies).

- TAM102/L1.3: previously described (Morrish *et al.*, 2002), a derivative of JJ101/L1.3 in which the L1.3 element lacks the 5'UTR.

- JJ101/L1.3Δ3'UTR: previously described (Heras *et al.*, 2013), a derivative of JJ101/L1.3 in which the L1.3 element lacks the 3'UTR.

- pCDNA6: contains an expression cassette for blasticidin-S deaminase (Invitrogen)

- 99-UB-LRE3: previously described (Coufal *et al.*, 2009), contains a full-length human LRE3 element (accession number #AC067958) (Brouha *et al.*, 2002) tagged with the *megfpI* reporter cassette and cloned in a modified version of pCEP4 (Life Technologies) in which the CMV promoter was substituted for the human ubiquitin C promoter (nucleotides 125398319-125399530 of human chromosome 12 [hg19]) and that contains a puromycin resistance instead of a hygromycin resistance marker (Garcia-Perez *et al.*, 2010).

- pYX014: previously described (Xie *et al.*, 2011), contains a full-length human L1RP element human L1RP element (accession number #AF148856.1) tagged with the *mflucI* indicator cassette cloned in a modified pCEP4 (Life Technologies) that lacks a CMV promoter and contains a puromycin resistance marker and an intact Renilla luciferase expression cassette.

- pYX015: previously described (Xie *et al.*, 2011), a derivative of pYX014 in which the L1RP element contains two engineered missense mutations in L1-ORF1p (R261A, R262A) that make it unable to mobilize.

- pYX017: previously described (Xie *et al.*, 2011), a derivative of pYX014 in which the 5'UTR of the L1 has been substituted for an exogenous, highly active CAG* promoter.

M.4.B. Plasmids used elsewhere

The following plasmids were used for the rest of the experiments:

- pSA500: a derivative of pAD500 (Doucet *et al.*, 2010) that contains the 5'UTR, 2xFLAG-tagged ORF2 and 3'UTR of an active human L1.3 element, cloned in pCEP4 (Life Technologies).
- pSA500 3'UTR-scrb: a derivative of pSA500 in which a scrambled sequence has been introduced in the 3'UTR of L1.3. Further details of the cloning procedure are provided in section M.12.
- pSA500 3'UTR-bs2: a derivative of pSA500 in which the let-7 binding site in ORF2 (bs2) has been introduced in the 3'UTR of L1.3. Further details of the cloning procedure are provided in section M.12.
- pSA500 3'UTR-8mer: a derivative of pSA500 in which a mutated the let-7 binding site that contains a canonical 8mer site for let-7 has been introduced in the 3'UTR of L1.3. Further details of the cloning procedure are provided in section M.12.
- pSA500 ORF2-8mer: a derivative of pSA500 in which the let-7 binding site in ORF2 (bs2) is mutated to a canonical 8mer site for let-7. Further details of the cloning procedure are provided in section M.11.

* A strong synthetic promoter containing: the cytomegalovirus (CMV) early enhancer element (C); the promoter, the first exon and the first intron of chicken beta-actin gene, (A); and the splice acceptor of the rabbit beta-globin gene (G).

- pVan583: a gift from Gael Cristofari, a derivative of JM101/L1.3 in which ORF1 and ORF2 are fused to the C-terminus of EGFP and mCherry respectively.
- psiCHECK2: contains a multicloning site in the 3'UTR of a Renilla luciferase gene and a Firefly luciferase gene in the backbone (Promega).
- psiCHECK2-nobs: derivative of psiCHECK2 in which a scrambled sequence that was not a let-7 binding site (Genscript) was cloned five times in tandem in the 3'UTR of the Renilla Luciferase gene. *
- psiCHECK2-perfectbs: derivative of psiCHECK2 in which a perfect let-7 binding site (Genscript) was cloned five times in tandem in the 3'UTR of the Renilla Luciferase gene. *
- psiCHECK2-bs1: derivative of psiCHECK2 in which a region of L1-ORF2 (bs1, 2650-2671) synthesised by Genscript was cloned five times in tandem in the 3'UTR of the Renilla Luciferase gene. *
- psiCHECK2-bs2: derivative of psiCHECK2 in which another region of L1-ORF2 (bs2, 4587-4608) was cloned five times in tandem in the 3'UTR of the Renilla Luciferase gene. *
- psiCHECK2-bs2mut: derivative of psiCHECK2 in which a mutated bs2 region (bs2mut) was cloned five times in tandem in the 3'UTR of the Renilla Luciferase gene. *
* All psiCHECK2 constructs were generated by cloning sequences synthesised and cloned in pUC57 (Genescript) into psiCHECK2 using XhoI and NotI restriction enzyme sites.

pFLAG-AGO1: a gift from Edward Chan (Addgene plasmid # 21533), contains a 3xFLAG-tagged AGO1 in its N-terminus, expressed from a CMV promoter (Lian *et al.*, 2009).

pFLAG-AGO2: a gift from Edward Chan (Addgene plasmid # 21538), contains a 3xFLAG-tagged AGO2 in its N-terminus, expressed from a CMV promoter (Lian *et al.*, 2009).

CEET-NL2-IS2 pri-let-7a-3: in CEET-NL2-IS2 (Benabdellah *et al.*, 2016), a gift from Francisco Martín, pri-let-7a3 was cloned using AscI and SbfI sites under the expression of a CMV-TetO promoter. In the same backbone, TetR expression is driven by the EF1α promoter, allowing repression of transgene expression until addition of doxycycline. Further description of how it was generated is provided in section M.20.

pCMV_R8.91: encodes the human immunodeficiency virus (HIV) packaging gene.

pMD2.G: encodes the vesicular stomatitis virus (VSV-G) envelope gene.

M.5. Bioinformatic analysis

The obtention and analysis of sequencing data was performed in collaboration with Guillermo Peris and Alejandro Rubio.

Sequencing data

Whole-genome sequencing (WGS) and miRNA expression files from TCGA were obtained using the Genomic Data Commons (GDC) Legacy Archive and the GDC Data Transfer Tool. Cases of paired tumor-normal WGS where tumor miRNAs expression data was available were retrieved for lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC). High coverage (28-95x) WGS files aligned to hg19 from primary tumor and solid tissue normal samples, and miRNA gene quantification files from primary tumor were downloaded for LUAD (17 patients) and LUSC (28 patients).

WGS analysis

Putative somatic LINE-1 insertion calls for both normal tissue (NT) and primary tumor (PT) were obtained using MELT version 2.1.5. Candidate somatic/tumor-specific insertions were further filtered to discard possible sequencing artifacts: only calls with a minimum of three split-reads passing all internal MELT filters were included. Polymorphic insertion calls were excluded from final results. Next, several quality values were checked. First, somatic insertions found in NT alone, and NT and PT simultaneously were expected to be zero, and only a maximum of one insertion was allowed for these values. All samples passed this additional filtering. Furthermore, polymorphic L1 insertions after MELT filtering were controlled, requiring that a similar number was found for PT and NT samples, independently of sample coverage. Four samples were excluded from the analysis after only a few polymorphic insertions passed all filters (TCGA-60-2695, TCGA-60-2722) and LUAD (TCGA-55-1594, TCGA-55-1596). Filtered LINE-1 calls were considered somatic insertions if detected in PT but not in NT set within a range of 100 bp.

Correlation with miRNA expression

Samples were divided in two groups depending on whether putative somatic insertions were found or not in PT. Only miRNAs with over 100 mapped reads per million (RPM) were considered. For some of the analysis, expression of specific miRNAs known to be involved in the development and progression of lung cancer was analyzed. First, for each miRNA, outliers were discarded (defined as values deviating more than two standard deviations from the mean in each group). Differentially expressed miRNAs were identified applying an unpaired two-tailed t test adjusted by FDR<0.01. Results were confirmed using a rank-sum test. RPM values were normalized to the highest expression value of each miRNA to enable visualization of all miRNAs in the same graph. Tumor

suppressor miRNAs and oncomiRs related to lung cancer used for this analysis were described in a recent revision (Inamura and Ishikawa, 2016). As a control, L1 insertion numbers were randomly re-assigned to each sample and analysis was repeated. Moreover, analysis was done with the number of somatic L1 insertions determined by Helman and col. using Transpo-seq in lung and breast cancer samples obtained from TGCA as well (Helman *et al.*, 2014). Data processing and analysis was performed as described above.

M.6. MicroRNA inhibitors and mimics

Let-7a/b mimic (C-300473-05 and C-300476-05), miR-34 mimic (C-300551-07), mimic control (scr, CN-002000-01), let-7 inhibitor (IH-300476-07), miR-20 inhibitor IH-300491-05) and inhibitor control (c-, IN-002005-01) were purchased from Dharmacon. They were resuspended in 1x siRNA Buffer (Thermo) to a working concentration of 20 μM, aliquoted and kept at -80ºC. MiRNA mimics/scr were used at a final concentration of 60nM, whereas miRNA inhibitors/c- were used at a final concentration of 40nM.

M.7. Retrotransposition assays

Slightly modified versions of previously established L1 retrotransposition assays were performed, each of which is described below. The sRNA/L1 retrotransposition assay, that allows testing the effect of miRNA overexpression or depletion in engineered retrotransposition, has recently been published in detail (Tristan-Ramos *et al.*, 2020). Each assay is described below.

Neomycin resistance assay.

Adapted from (Wei *et al.*, 2000). $2x10^5$ HeLa JVM cells were plated in 6-well dishes. Within 24h, cells were co-transfected with 0.5-1μg of JM101/L1.3 (or any of its derivatives) and 60nM of mimic/scr or 40nM of inhibitor/c- with

Dharmafect DUO (Dharmacon) following manufacturer's instructions. Selection with 400 μg/mL of G418 (Life) was started 48h post-transfection, and medium was changed every two days with fresh G418 to remove dead cells and debris. Between 12 and 14 days after transfection, cells were washed with 1X PBS (Gibco), fixed (2% formaldehyde, 0.2% glutaraldehyde in 1X PBS), and stained with 5% crystal violet. Colonies were manually counted. The number of antibiotic-resistant colonies was used to quantify retrotransposition levels in cultured cells. Clonability assay were performed transfecting $1x10^5$ HeLa JVM cells with 0.5 μg pU6iNeo and selecting with G418 as described above.

Blasticidin resistance assay.

Adapted from (Morrish *et al.*, 2002). $2x10^5$ HeLa JVM cells were plated in 6-well dishes. Within 24h, cells were co-transfected with 0.5-1μg of JJ101/L1.3 (or any of its derivatives) and 60nM of mimic/scr or 40nM of inhibitor/c- with Dharmafect DUO (Dharmacon) following manufacturer's instructions. Selection with 10 μg/mL of blasticidin-S (Millipore) was started 5 days post-transfection, and medium was changed every two days with fresh Blast to remove dead cells and debris. Between 12 and 14 days after transfection, cells were washed with 1X PBS (Gibco), fixed (2% formaldehyde, 0.2% glutaraldehyde in 1X PBS), and stained with 5% crystal violet. Colonies were manually counted. The number of antibiotic-resistant colonies was used to quantify retrotransposition levels in cultured cells. Clonability assay were performed transfecting $1x10^5$ HeLa JVM cells with 0.5 μg pCDNA6 and selecting with Blast as described above.

Luciferase assay

Adapted from (Xie *et al.*, 2011). $1x10^5$ HEK293T, $1x10^5$ SK-MES-1 or $7x10^4$ A549 cells HeLa JVM were plated in a 24-well plate. Within 24 hours, cells were co-transfected with 200 ng of pYX014, pYX015 or pYX017 and 60nM of let-7

mimic or 40nM of let-7 inhibitor using Lipofectamine 2000 (Life) following manufacturer's protocols. The next day, and for 24h, transfected cells were selected by adding puromycin to the media (5μg/mL for HeLa JVM and HEK293T, 1μg/mL for SK-MES-1 and A549). Firefly and Renilla luciferase activity were measured 96h post-transfection using Dual-Luciferase Reporter Assay System (Promega), in a GloMax Luminometer (Promega), following manufacturer's protocols.

EGFP assay

Adapted from (Ostertag *et al.*, 2000). $4x10^5$ HEK293T were plated in a 6-well plate. Within 24 hours, cells were co-transfected with 1μg of 99-UB-LRE3 and 40nM of let-7 inhibitor using Lipofectamine 2000 (Life). Transfected cells were selected with 5μg/mL puromycin for 6 days, starting 24h post-transfection, and GFP was measured 8 days post transfection in a FACS Canto cytometer (BD).

M.8. RNA Immunoprecipitation (RIP)

PA-1 cells ($2x10^6$) were transfected in 10cm tissue culture plates with 4μg FLAG-AGO2 and 25nM scr or let-7 mimic using Lipofectamine 2000 (Life), following manufacturer's instructions. Transfection with pBSKS (empty vector) was used as a negative control for the IP. Forty-eight hours post-transfection, cells were washed with ice-cold 1xPBS, scraped and transferred to a 1.5 ml tube. After centrifugation at 200g for 3 minutes, cells were resuspended in 200μl of cold resuspension buffer (20mM Tris pH=7.5, 150mM NaCl, 1mM EDTA, 1mM EGTA) containing 1U/μL RNAsin Plus (Promega) and lysed adding 800μl of cold lysis buffer (1% Triton X-100, 20mM Tris pH=7.5, 150mM NaCl, 1mM EDTA, 1mM EGTA, 1mM phenylmethyl-sulfonyl fluoride (PMSF, Sigma), 1X cOmplete EDTA-free Protease Inhibitor cocktail (Roche)) and incubating for 20 min on ice. After centrifugation (10,000g for 10min at 4ºC), 10μL of RQ1 DNase

(Promega) was added to the supernatant, and 10% (100μL was kept as 'input'). Immunoprecipitation of FLAG-AGO2 was performed with 25μL Dynabeads Protein G (Life) and 1μg anti-FLAG M2 mouse (Sigma, F3165) for 3h at 4ºC with rotation. After five washes with lysis buffer, 10% of sample-beads were used for protein extraction and western-blot, while 90% was incubated with RQ1 DNase at 37ºC for 30 min for later RNA extraction with Trizol LS (Ambion). Absolute standard curve was used for RT-qPCR quantification.

M.9. siCHECK luciferase assays

HEK293T ($1x10^5$) or HeLa JVM ($8x10^4$) cells were seeded in 24-well plates. Within 24h, cells were co-transfected with 10ng of siCHECK plasmids and 50-80nM scr/let-7 mimic using Lipofectamine 2000 (Life). Firefly and renilla luciferase measurements were performed 24h post-transfection transfection using Dual-Luciferase Reporter Assay System (Promega), in a GloMax Luminometer (Promega), following manufacturer's protocols. For HEK293T cells, 10μL of total extract was used, and whereas 20μL were used for HeLa JVM cells.

M.10. Site-directed mutagenesis

Binding site mutant 'bs2mut' was generated using an established protocol for site-directed mutagenesis (Heckman and Pease, 2007).

- To generate JM101/bs2mutL1.3: 2 sequential PCRs were performed, using an active L1.3 as a template. First, two PCRs were performed using the following primers under standard conditions: Let7-ORF2PCRafw/Let7-ORF2PCRa_PG2rv and Let7-ORF2PCRarv/Let7-ORF2PCRa_PG2fw. The products of both reactions were purified, mixed in equal amounts, and used as a template for a second PCR using primers Let7-ORF2PCRafw/ Let7-ORF2PCRarv. Conditions for this PCR were: 95ºC 5min, 10 cycles with (95ºC 15s, 50ºC 30s, 72ºC 60s), 25 cycles with (95ºC 15s, 55ºC 30s, 72ºC 60s), 72ºC

10min. The resulting product contained the mutated sequence in ORF2. This product was purified and cloned into a plasmid containing an active L1 (pJCC5/L1.3) using EcoNI and BsaBI sites, generating pJCC5/bs2mutL1.3. This mutant L1 was then cloned into pJM101 using NotI and BstZ17I sites, generating pJM101/bs2mutL1.3.

- To generate pSA500 ORF2-8mer: the same protocol described above was used, with minimal modifications. For the first two PCRs, primers used were Let7-Bcl1-ORF2bs-PCRaFw/Let7-ORF2PCRa_8mer and Let7-ORF2PCRb_8mer/ pCEP4_Rv. The products of both reactions were purified, mixed in equal amounts, and used as a template for a second PCR using primers Let7-Bcl1-ORF2bs-PCRaFw/pCEP4_Rv. The product was purified and cloned into pSA500 using BclI and BstZ17I sites, generating pSA500-ORF2-8mer.

Both constructs were checked by digestion and Sanger sequencing.

M.11. Generation of 3'UTR variants of pSA500

Sequences were ordered as oligos flanked by BstZ17I sites. First, 1μL of each Fw and Rv were annealed and phosphorylated with T4 Polynucleotide Kinase (PNK, NEB) using the following program: 30min at 37ºC, 5min at 95ºC, and then down to 25ºC at -5ºC/min. They were cloned in 3'UTR of pSA500 using a BstZ17I site, generating pSA500-3'UTR-scrb/bs2/8mer. All constructs were checked by digestion and Sanger sequencing.

M.12. Western Blot

In this Thesis, two slightly different protocols have been used for Western Blot: both share protein extraction and blotting, but one uses 'homemade' gels and the other uses commercial 'pre-cast' gels, with different transfer methods. The main steps are described below:

- Cell lysis and protein extract preparation. Cells were washed with 1X PBS, trypsinised and pelleted at 200g for 4 minutes. Pellets were lysed in ice-cold RIPA buffer (Sigma) supplemented with 1X cOmplete EDTA-free Protease Inhibitor cocktail (Roche), 1mM PMSF (Sigma) and 35mM β-mercaptoethanol (Sigma), and incubated for 15 min on ice. Extract was then centrifuged at 18000g for 10 min at 4ºC, and debris-free supernatant was transferred to a new pre-chilled tube. Protein concentration was determined using the Micro BCA Protein Assay Kit (Thermo) following manufacturer's instructions.
- Run and transfer using 'homemade' gels. Proteins were resolved in an SDS-PAGE gel (8, 10 or 12%, depending on the experiment) at 100V for 60min in running buffer (25mM Tris, 192mM Glycine, 0.1% SDS, all from Sigma), and transferred to a nitrocellulose membrane (BioRad) at 250mA for 90min in transfer buffer (48mM Tris, 39mM Glycine, in 20% Methanol diluted in dH_2O). Ponceau staining was performed to check transferring of proteins to the membrane.
- Run and transfer using pre-cast gels. Proteins were resolved in an 4-15% Mini-PROTEAN TGX Precast Gels (BioRad) at 100V for 60min in running buffer (described above) and transferred to a PVDF membrane using Trans-Blot Turbo Mini PVDF Transfer Packs (BioRad) at 1.3A for 10min in the Trans-Blot Turbo Transfer System (BioRad).
- Blotting. After transfer, membranes were blocked using 5% milk in TBS-T (0.1% Tween-20 in 1XTBS) during 1h at room temperature. For blotting we used the following antibodies for an overnight incubation at 4ºC with gentle shaking: a polyclonal rabbit anti L1-ORF1p (1:1000, provided by Dr. Oliver Weichenrieder, Max-Planck, Germany), a monoclonal mouse anti-L1-ORF1p (1:2000, Millipore MABC1152), anti-HMGA2 (1:1000, Abcam), anti-DICER (1:1000, Cell Signalling #3363), anti-AGO2 (Millipore, MABE253), anti-tubulin (1:1000, Santa Cruz). For chemiluminescent detection we used anti-rabbit HRP (1:2000, Cell Signaling) or anti-mouse HRP (1:2000, Cell Signaling), and Clarity ECL

Western Blotting Substrate (BioRad) or SuperSignal West Femto Maximum Sensitivity Substrate (Thermo). Images were acquired with an ImageQuant LAS4000 and quantified using ImageJ software. For Odyssey analysis, anti-rabbit and anti-mouse fluorescent antibodies from LI-COR were used at 1:10000 dilution, and detection and quantification were performed in Odyssey (LI-COR). All secondary antibodies were incubated for 1-2h at room temperature. After incubation with primary and secondary antibody, three 5min washes with TBS-T in gentle shaking were performed to remove excess of antibody.

M.13. Genomic and plasmid DNA extraction

To extract genomic and plasmid DNA, cells were lysed by resuspending cell pellets in a buffer containing 10mM Tris pH=8.2, 10mM EDTA, 200mM NaCl, 0.5%SDS and 200μg/ml proteinase K, and incubating for 2h at 56ºC with shaking. Next, one volume of 25:24:1 Phenol:Chloroform:Isoamyl alcohol (Thermo) was added, mixture was vortexed for 1min then centrifuged at 20,000g for 5min. Aqueous phase was transferred to a new tube, and DNA was precipitated overnight at -20ºC after addition of 2.5 volumes of cold 100% EtOH and 0.1 volumes of 3M AcNa (pH=5.2). Next day, DNA was pelleted at 20,000g for 30min at 4ºC, washed with 70% EtOH, and resuspended in Nuclease-Free water using vortex at low speed.

M.14. RNA extraction

RNA was extracted from cells using Trizol (Invitrogen), following manufacturer's instructions. RNA was subsequently treated with 10 units of RQ1 DNase at 37ºC for 45min and purified by Phenol:Chloroform:Isoamyl extraction and overnight EtOH/AcNa precipitation as described above in the 'Genomic and plasmid DNA extraction' section.

M.15. qPCR and RT-qPCR

In all cases, the GoTaq qPCR mix (Promega) was used in a StepOne Plus instrument (Applied Biosystems). Program: 95ºC for 10min, 40 cycles of 15s at 95ºC then 1min at 60ºC, and finally melting curve.

qPCR

50ng of DNA and 0.2μM primer were used per reaction. An untransfected control and was used to discard contamination. Normalization was performed using genomicGAPDH primers unless otherwise indicated. Relative quantification method using standard curves for each primer was used.

RT-qPCR

First, 1μg RNA was treated with DNase I (RNase-free, Thermo), then cDNA was synthesized using High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems) following manufacturer's instructions in both cases. After a 1:4 dilution in nuclease-free H_2O, 25ng cDNA was used per reaction. Two controls were used to verify the absence of contaminating gDNA: no-RT and no-template. Endogenous L1 mRNA was quantified using N51 primers (Macia *et al.*, 2017). GAPDH was used to normalize. Transfected L1 mRNA was quantified using NEOjunct2 primers, designed to exclusively amplify the spliced neomycin cassette (Banuelos-Sanchez *et al.*, 2019) or SV40 primers when the cassette was absent. EBNA-1, expressed from the backbone of the plasmids, was used to normalize. Relative quantification method using standard curves for each primer was used in all cases.

Mature miRNA quantification

A 3'RACE RT-qPCR was used. First, 1μg of total RNA was polyadenylated and then cDNA was synthesised using an oligo(dT) with qScript microRNA

cDNA synthesis Kit (QuantaBio), following manufacturer's instructions. Quantitative PCR was performed using a universal primer against poly(A) (provided with the kit) and a miRNA-specific primer followed by 'AAA' that ensures the specific detection of polyadenylated mature miRNA and not its precursors (QuantaBio). qPCR was performed with primers Let-7aAAA, Let-7bAAA, miR-34AAA and miR-20AAA. Small nucleolar RNA SNORD44 was used to normalize.

M.16. Flow cytometry

For pVan583 experiments, $2x10^5$ HeLa cells were seeded per well in 6-well plates. Next day, cells were transfected with 1μg of pVan583 and 40nM let7 inhibitor or it control (c-) using Lipofectamine 2000 (Life). Seventy-two hours post-transfection, cells were washed with 1X PBS (Life), detached with TrypLE Express (Gibco) for 5 min at 37ºC, pelleted 4 min at 200g, resuspended in 1X PBS with 5% FBS and 5mM EDTA, and passed through a 70μM filter. After incubation with 10ug/mL 7AAD (Sigma) for 10min, fluorescence was quantified in a FACSAria (BD) cytometer. For each replicate, 105 cells were passed through the cytometer. Only live and transfected cells (7AAD- and GFP+, between 3600 and 9300 cells) were used for %mCherry analysis, which was performed using FlowJo software (LLC). Controls were used to set the threshold for each fluorescent channel of detection: untransfected cells, and cells expressing either GFP only or mCherry only.

For GFP retrotransposition assays, transfection and selection is described in the 'Retrotransposition assays' section. Cells were processed as described above, and GFP+ cells were analyzed in a FACS Canto (BD).

M.17. Confocal microscopy

$2x10^5$ U2OS per well were seeded in 6-well plates. Next day, cells were transfected with 1ug of DNA and 60nM scr/mimic using lipofectamine 2000, following standard protocols. Twenty-four hours post-transfection, cells were detached and re-seeded in 24-well plates where a glass slide had previously been placed. Forty-eight hours post-transfection, cells were washed in PBS 1X and fixed in paraformaldehyde (4% in PBS 1X) for 30 minutes at room temperature. Slides were then mounted with Slow-Fade Gold Antifade with DAPI (Life). Slides were imaged using a Zeiss LSM-710 confocal microscope (Leica).

M.18. Hybridization between let-7 and L1

The potential structure formed by let-7 and L1 or the binding site mutant was analyzed in RNAHybrid (Rehmsmeier *et al.*, 2004). Briefly, the region of L1Hs identified as 'bs2' with RNA22 was paired to the mature sequence of let-7, using default parameters.

M.19. Generation of SK-MES-1 cells with inducible expression of let-7a

To generate SK-MES-1 cells for stable and inducible expression of let-7, we modified a previously published second generation lentiviral vector, CEETnl2Is2 (Benabdellah *et al.*, 2016), a gift from Dr. Francisco Martín's lab.

- Cloning of pri-let-7a3 in CEETnl2Is2. Pri-let-7a3 was amplified from 150ng of SK-MES-1 genomic DNA with High Fidelity PCR System (Roche) following manufacturer's instructions, using primers that added AscI and SbfI restriction sites at the 5' and the 3' end of the amplicon, respectively (AscI_Prilet7a3_Fw and SbfI_Prilet7a3_Rv). Conditions for this PCR were: 95ºC 2min, 10 cycles with (94ºC 15s, 50ºC 30s, 72ºC 45s), 25 cycles with (94ºC 15s, 60ºC 30s, 72ºC 45s),

72ºC 10min. The resulting product was purified and cloned into CEETnl2Is2 using AscI and SbfI sites, generating CEET-NL2-IS2 pri-let-7a-3.

- Lentivirus production. In 10cm^2 dishes, 4-5x10^6 HEK293T were transfected with 10µg of CEET-NL2-IS2 pri-let-7a-3, 7µg of pCMV_R8.91, and 3µg of pMD2.G using 80µg of polyethylenimine (PEI). DNA and PEI were mixed in DMEM without serum, and incubated for 20min at room temperature before being added to the cells. Twenty-four hours post-transfection, fresh media was added to the cells. Forty-eight hours post-transfection, supernatant was removed from the cells, and viruses were concentrated by ultracentrifugation at 23000rpm for 2h at 4ºC, followed by resuspension in 1/100 of the initial volume of DMEM.

- Transduction and clonal cell line generation. 100,000 SK-MES-1 cells were pelleted at 200g for 4min, resuspended in 200µL of the concentrated lentiviruses and 100µL of DMEM for a total volume of 300µL, and seeded in a 24-well plate. Five hours after seeding, viruses were removed and fresh media was added to the cells. To generate clonal cell lines, cells were diluted and seeded in 96-well plates at 0.5 cells/well. Fresh medium was added a week after seeding. Around 14 days after seeding, positive wells (i.e. those with a single colony of cells in it) were passaged to a 48-well. Clonal cell lines were subsequently expanded until they reached 100% confluency in a 24-well, then they were detached with trypsin and 80% used for freezing while 20% was pelleted for gDNA extraction.

- PCR analysis of clones. PCRs were performed with 60ng of gDNA using KAPA2G Fast Hotstart ReadyMix PCR Kit (Kapa Biosystems) following manufacturer's instructions, and run in a 1% agarose gel. The primers used map to the LTR region of the lentivirus in a way that ensures that amplification can only occur upon viral integration ('ΔU3 (LTR) Fw', 'PBS (LTR) Rv').

- Copy number analysis of positive clones by qPCR. To analyze copy number, we used 60ng of gDNA in the first well and then four 1/10 serial dilutions. Primers used to quantify the number of integrated copies were 'ΔU3 (LTR) Fw' and

'PBS (LTR) Rv'). Albumin was used to normalize. Quantification was performed by comparison with a control cell line that contains one integrated copy per cell, provided to us by the group of Dr. Francisco Martín (Genyo). We estimated that 60ng corresponds to ~10,000 cells, therefore to 10,000 copies.

M.20. Cross-Linking, Amplification and Sequencing of Hybrids (CLASH)

CLASH was performed in the Institute for Infection and Immunity Research (University of Edinburgh, UK):

- Cell cross-linking and lysis. PA-1 cells ($2x10^7$) growing in 150mm plates were rinsed once with warm 1XPBS, covered with 10mL of cold 1XPBS, and irradiated with UV-C (254nm) for 400 mJ/cm^2 using a Stratagene Stratalinker 2400 (Agilent). Cells were then scraped and pelleted at 200g for 4min at 4ºC. Pellets were resuspended in 900µL of lysis buffer (50mM Tris/HCL pH=7.8, 300mM NaCl, 1% Triton X-100, 1X cOmplete protease inhibitor (Roche), 5mM EDTA, 10% glycerol), and lysed for 30min on ice.
- DNase and RNase treatment. Next, 30µL of RQ1 DNase (Promega) and 10µL of 1:250 RNase I (Ambion) were added to the lysate and incubated for 5min at 37ºC with shaking (1000rpm). Lysates were centrifuged at 16000g for 15min at 4ºC, and supernatant was transferred to a new tube.
- Immunoprecipitation and washes. For the AGO2 IP, 100µL of Dynabeads Protein G (Thermo) were incubated with 10µg of rat anti-AGO2 (Merck) or an IgG control (Sigma) in a total volume of 200µL with 0.02% Tween-20 in 1XPBS for 4h at 4ºC in rotation. During that time the same time, lysates were pre-cleared with 10µL of beads in rotation at 4ºC. Beads were then washed three times with lysis buffer, then mixed with the lysate and incubated overnight at 4ºC with rotation. Next day, CLASH washes were performed as follows:

- Once with LS (Low Salt) buffer (50 mM Tris-HCl pH 7.5, 300 mM NaCl, 5 mM $MgCl_2$, 0.5% Triton X-100, 2.5% glycerol)
- Twice with HS (High Salt) buffer (50 mM Tris-HCl pH 7.5, 800 mM NaCl, 5 mM $MgCl_2$, 0.5% Triton X-100, 2.5% glycerol) for 5 minutes at 4ºC with rotation
- Once with LS buffer
- Once with PNK buffer (50 mM Tris-HCl pH 7.5, 50 mM NaCl, 10 mM $MgCl_2$, 0.5% Triton X-100)

- 5'end phosphorylation and intermolecular RNA-RNA ligation. Next, beads were treated with 0.5U/μL T4 Polynucleotide Kinase (3' phosphatase minus, NEB) and 1mM rATP in a final volume of 80μL for 2.5h at 20ºC to phosphorylate cleaved mRNA 5'-ends. After another round of CLASH washes, intermolecular RNA-RNA ligation was performed using 0.5U/μL T4 RNA Ligase I (NEB) and 1mM ATP in a final volume of 160μL overnight at 16ºC.
- RNA dephosphorylation and 3'adaptor ligation. The following morning, after CLASH washes, beads were treated with 0.1U/μL TSAP (Promega) in a final volume of 80μL for 45min at 20ºC to remove 3' phosphates of the RNAs, then washed again. Pre-adenylated, IR800-tagged adaptor was added with 10U/μL Truncated RNA Ligase 2 (truncated KQ, NEB), with 1μM adaptor and 10% PEG 8000 (NEB), in a final volume of 160μL overnight at 16ºC protected from the light.
- Elution of AGO-RNA complexes from beads. CLASH washes were performed, then beads were resuspended in 20μL of buffer containing 1XPNK, 1.5X NuPAGE Reducing Agent (Invitrogen) and 1X NuPAGE LDS Sample Buffer (Invitrogen), and incubated at 70ºC for 10min with shaking. Samples were then resolved in a 4-12% Bis-Tris NuPAGE gel at 120V for 150min in the dark. Labelled RNA was then visualized directly in the gel using an Odyssey scanner (LI-COR). AGO-RNA complexes were excised from the gel, 'crushed', and

treated with 100µg Proteinase K in a final volume of 500µL for 2h at 55ºC shaking at 1000rpm.

- RNA purification. RNA was purified using Phenol/Chloroform extraction and overnight EtOH/AcNa precipitation, following standard protocols. Pellets were resuspended in 2.5µL of ddH2O.

- 5'adaptor ligation and library preparation. Ligation of the 5' adaptor, as well as the 3' sequencing primer containing the index to identify the samples, was performed using the CleanTag Small RNA Library Prep Kit (TriLink) following manufacturer's instructions. Notably, this library preparation kit adds specific chemical modifications to the adapters to prevent dimer formation (Shore et al., 2016). Libraries were run on a 6% TBE gel (Invitrogen) and purified, excluding primer dimers, from gel using standard DNA precipitation protocols. Libraries were quantified using Qubit 3.0 fluorometer (Invitrogen) and the Qubit dsDNA HS Assay Kit, analyzed with the 2100 Bioanalyzer (Agilent), and then pooled and sent for sequencing in Edinburgh Genomics (Edinburgh).

M.21. Primers

Primer name	Sequence (5' to 3')
let-7a-5pAAA	TGAGGTAGTAGGTTGTATAGTTAAA
let-7b-5pAAA	TGAGGTAGTAGGTTGTGTGGTTAAA
miR-34a-5pAAA	TGGCAGTGTCTTAGCTGGTTGTAAA
miR-20aAAA	TAAAGTGCTTATAGTGCAGGTAGAAA
SNORD44	GCAAATGCTGACTGAACATGAA
N51 Fw	GAATGATTTTGACGAGCTGAGAGAA
N51 Rv	GTCCTCCCGTAGCTCAGAGTAATT
HMGA2 Fw	TTGCTGCCTTTGGGTCTTCC
HMGA2 Rv	CAGCGCCTCAGAAGAGAGGACG
DICER1 Fw	AGTGGTAGGCTTTCACACAG
DICER1 Rv	AGAAAGGACCCATTGGTGAG
GAPDH Fw	TGCACCACCAACTGCTTAGC
GAPDH Rv	GGCATGGACTGTGGTCATGAG
SV40 Fw	TGGACAAACCACAACTAGAATGC
SV40 Rv	TTGCAGCTTATAATGGTTAC
CMV Fw	ACTGCCAAGTAGGAAAGTCCCA
CMV Rv	ATGCCAAGTACGCCCCCTAT
EBNA-1 Fw	CGTCATCTCCGTCATCACC
EBNA-1 Rv	AGATTTGCCTCCCTGGTTTC
NEOjunct2 Fw	TGCCTCGTCCTGAAGCTC
NEOjunct2 Rv	CAATCGGCTGCTCTGATG
genomicGAPDH Fw	CGTTTCCCAAAGTCCTCCTGT
genomicGAPDG Rv	AGGTGATCGGTGCTGGTTC
DGCR8 SURVEYOR 1 Fw	ACTCGCTTAGTCGCCAGTCA

DGCR8 SURVEYOR 1 Rv	TTACTCCTGCAGCTCTCGGT
Let7ORF2PCRa_ PG2rv	AAAATCCCCCAACGTGATTCCTCCAGCTTTGTTC
Let7-ORF2PCRb_ PG2fw	GAGGAATCACGTTGGGGGATTTTAAACTATACTAC
AscI_Prilet7a3_Fw	AAAGGCGCGCCTGCCCGCCAGAATCCCT
SbfI_Prilet7a3_Rv	TTTCCTGCAGGTCACACAGCAAGTGGCACCTAG
DU3 (LTR) Fw	GATCTGCTTTTTGCTTGTACT
PBS (LTR) Rv	GAGTCCTGCGTCGAGAGAGC
hAlb Fw	GCTGTCATCTCTTGTGGGCTGT
hAlb Rv	ACTCATGGGAGCTGCTGGTTC
bs2_scrambled Fw	TACAGTTGCGTTGTAGAACGATATAGAGGAACTACGC AGTAAGGTA
bs2_scrambled Rv	TACCTTACTGCGTAGTTCCTCTATATCGTTCTACAACGC AACTGTA
bs2 Fw	TACGAACAAAGCTGGAGGCATCACACTACCTGACTTCA AACTAGTA
bs2 Rv	TACTAGTTTGAAGTCAGGTAGTGTGATGCCTCCAGCTT TGTTCGTA
bs2_8mer Fw	TACGAACAAAGCTGGAGGCATCACTCTACCTCACTTCA AACTAGTA
bs2_8mer Rv	TACTAGTTTGAAGTGAGGTAGAGTGATGCCTCCAGCTT TGTTCGTA
Let7-ORF2PCRa_ 8mer	GAAGTGAGGTAGAGTGATGCCTCCAGCTTTGTTC
Let7-ORF2PCRb_ 8mer	GAGGCATCACTCTACCTCACTTCAAACTATACTAC
Let7-Bcl1-ORF2bs-PCRaFw	TGGATTCACAGCCGAATTCTACC
pCEP4 Rv	GTGGTTTGTCCAAACTCATC

BIBLIOGRAPHY

Adams, B. D. *et al.* (2016) 'MiR-34a silences c-SRC to attenuate tumor growth in triple-negative breast cancer', *Cancer Research*, 76(4), pp. 927–939. doi: 10.1158/0008-5472.CAN-15-2321.

Adney, E. M. *et al.* (2019) 'Comprehensive scanning mutagenesis of human retrotransposon LINE-1 identifies motifs essential for function', *Genetics*, 213(4), pp. 1401–1414. doi: 10.1534/genetics.119.302601.

Agarwal, V. *et al.* (2015) 'Predicting effective microRNA target sites in mammalian mRNAs', *eLife*, 4, pp. 1–38. doi: 10.7554/eLife.05005.

Agrawal, A., Eastman, Q. M. and Schatz, D. G. (1998) 'Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system', *Nature*, 394(6695), pp. 744–751. doi: 10.1038/29457.

Ahl, V. *et al.* (2015) 'Retrotransposition and Crystal Structure of an Alu RNP in the Ribosome-Stalling Conformation', *Molecular Cell*, 60(5), pp. 715–727. doi: 10.1016/j.molcel.2015.10.003.

Alisch, R. S. *et al.* (2006) 'Unconventional translation of mammalian LINE-1 retrotransposons', *Genes and Development*, 20(2), pp. 210–224. doi: 10.1101/gad.1380406.

Anastasiadou, E., Jacob, L. S. and Slack, F. J. (2018) 'Non-coding RNA networks in cancer', *Nature Reviews Cancer*. Nature Publishing Group, 18(1), pp. 5–18. doi: 10.1038/nrc.2017.99.

Aneichyk, T. *et al.* (2018) 'Dissecting the Causal Mechanism of X-Linked Dystonia-Parkinsonism by Integrating Genome and Transcriptome Assembly', *Cell*, 172(5), pp. 897-909.e21. doi: 10.1016/j.cell.2018.02.011.

Ardeljan, D. *et al.* (2019) 'LINE-1 ORF2p expression is nearly imperceptible in human cancers', *Mobile DNA*, 11(1), pp. 1–19. doi: 10.1186/s13100-019-0191-2.

Ardeljan, D. *et al.* (2020) 'Cell fitness screens reveal a conflict between LINE-1 retrotransposition and DNA replication', *Nature Structural and Molecular Biology*, 27(2), pp. 168–178. doi: 10.1038/s41594-020-0372-1.

Athanikar, J. N., Badge, R. M. and Moran, J. V. (2004) 'A YY1-binding site is required for accurate human LINE-1 transcription initiation', *Nucleic Acids Research*, 32(13), pp. 3846–3855. doi: 10.1093/nar/gkh698.

Attig, J. *et al.* (2018) 'Heteromeric RNP Assembly at LINEs Controls Lineage-Specific RNA Processing', *Cell*, 174(5), pp. 1067-1081.e17. doi: 10.1016/j.cell.2018.07.001.

Baek, D. *et al.* (2008) 'The impact of microRNAs on protein output', *Nature*, 455(7209), pp. 64–71. doi: 10.1038/nature07242.

Baillat, D. and Shiekhattar, R. (2009) 'Functional Dissection of the Human TNRC6 (GW182-Related) Family of Proteins', *Molecular and Cellular Biology*, 29(15), pp. 4144–4155. doi: 10.1128/mcb.00380-09.

Baillie, J. K. *et al.* (2011) 'Somatic retrotransposition alters the genetic landscape of the human brain', *Nature*, 479(7374), pp. 534–537. doi: 10.1038/nature10531.

Balzeau, J. *et al.* (2017) 'The LIN28/let-7 Pathway in Cancer', *Frontiers in Genetics*, 8, pp. 1–16. doi: 10.3389/fgene.2017.00031.

Bannert, N. and Kurth, R. (2006) 'The Evolutionary Dynamics of Human Endogenous Retroviral Families', *Annual Review of Genomics and Human Genetics*, 7(1), pp. 149–173. doi: 10.1146/annurev.genom.7.080505.115700.

Banuelos-Sanchez, G. *et al.* (2019) 'Synthesis and Characterization of Specific Reverse Transcriptase Inhibitors for Mammalian LINE-1 Retrotransposons.', *Cell Chemical Biology*, 26(8), pp. 1095-1109.e14. doi: 10.1016/j.chembiol.2019.04.010.

Barau, J. *et al.* (2016) 'The DNA methyltransferase DNMT3C protects male germ cells from transposon activity', *Science*, 354(6314), pp. 909–912. doi: 10.1126/science.aah5143.

Bartel, D. P. (2009) 'MicroRNAs: target recognition and regulatory functions.', *Cell*, 136(2), pp. 215–33. doi: 10.1016/j.cell.2009.01.002.

Bartel, D. P. (2018) 'Metazoan MicroRNAs', *Cell*, 173(1), pp. 20–51. doi: 10.1016/j.cell.2018.03.006.

Baskerville, S. and Bartel, D. P. (2005) 'Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes', *RNA*, 11(3), pp. 241–247. doi: 10.1261/rna.7240905.

Batzer, M. A. and Deininger, P. L. (2002) 'Alu repeats and human genomic diversity', *Nature Reviews Genetics*, 3(5), pp. 370–379. doi: 10.1038/nrg798.

Beck, C. R. *et al.* (2010) 'L1 retrotransposition acivity in human genomes', *Cell*, pp. 1159–1170.

Beck, C. R. *et al.* (2011) 'LINE-1 Elements in Structural Variation and Disease', *Annu Rev Genomics Hum Genet*, 12(60), pp. 187–215. doi: 10.1146/annurev-genom-082509-141802.

Becker, K. G. *et al.* (1993) 'Binding of the ubiquitous nuclear transcription factor YY1 to a cis regulatory sequence in the human LINE-1 transposable element', *Human Molecular Genetics*, 2(10), pp. 1697–1702. doi: 10.1093/hmg/2.10.1697.

Belancio, V. P., Hedges, D. J. and Deininger, P. (2006) 'LINE-1 RNA splicing and influences on mammalian gene expression', *Nucleic Acids Research*, 34(5), pp. 1512–1521. doi: 10.1093/nar/gkl027.

Benabdellah, K. *et al.* (2016) 'Lent-On-Plus Lentiviral vectors for conditional expression in human stem cells', *Scientific Reports*, 6(December 2015), pp. 1–17. doi: 10.1038/srep37289.

Benitez-Guijarro, M. *et al.* (2018) 'RNase H2, mutated in Aicardi-Goutières syndrome, promotes LINE-1 retrotransposition', *The EMBO Journal*, 37(15), p. e98506. doi: 10.15252/embj.201798506.

Bennett, E. A. *et al.* (2008) 'Active Alu retrotransposons in the human genome.', *Genome Research*, 18(12), pp. 1875–83. doi: 10.1101/gr.081737.108.

Bernstein, E. *et al.* (2001) 'Role for a bidentate ribonuclease in the initiation step of RNA interference', *Nature*, 409(6818), pp. 363–366. doi: 10.1038/35053110.

Bodak, M. *et al.* (2017) 'Dicer, a new regulator of pluripotency exit and LINE-1 elements in mouse embryonic stem cells', *FEBS Open Bio*, 7(2), pp. 204–220. doi: 10.1002/2211-5463.12174.

Bogerd, H. P. *et al.* (2006) 'Cellular inhibitors of long interspersed element 1 and Alu retrotransposition.', *Proceedings of the National Academy of Sciences of the United States of America*, 103(23), pp. 8780–5. doi: 10.1073/pnas.0603313103.

Bohnsack, M. T., Czaplinski, K. and Gorlich, D. (2004) 'Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs.', *RNA*, 10(2), pp. 185–91. doi: 10.1261/rna.5167604.

Boissinot, S., Chevret, P. and Furano, A. V. (2000) 'L1 (LINE-1) retrotransposon evolution and amplification in recent human history', *Molecular Biology and Evolution*, 17(6), pp. 915–928. doi: 10.1093/oxfordjournals.molbev.a026372.

Borchert, G. M., Lanier, W. and Davidson, B. L. (2006) 'RNA polymerase III transcribes human microRNAs', *Nature Structural and Molecular Biology*, 13(12), pp. 1097–1101. doi: 10.1038/nsmb1167.

Boudreau, R. L. *et al.* (2014) 'Transcriptome-wide discovery of microRNA binding sites in Human Brain', *Neuron*, 81(2), pp. 294–305. doi: 10.1016/j.neuron.2013.10.062.

Bourc'his, D. and Bestor, T. H. (2004) 'Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L.', *Nature*, 431(7004), pp. 96–99. doi: 10.1038/nature02886.

Bourque, G. *et al.* (2018) 'Ten things you should know about transposable elements', *Genome Biology*, 19(1), p. 199. doi: 10.1186/s13059-018-1577-z.

Branciforte, D. and Martin, S. L. (1994) 'Developmental and cell type specificity of LINE-1 expression in mouse testis: implications for transposition.', *Molecular and cellular biology*, 14(4), pp. 2584–92. doi: 10.1128/mcb.14.4.2584.

Braun, J. E. *et al.* (2011) 'GW182 proteins directly recruit cytoplasmic deadenylase complexes to miRNA targets', *Molecular Cell*, 44(1), pp. 120–133. doi: 10.1016/j.molcel.2011.09.007.

Britten, R. J. and Kohne, D. E. (1968) 'Repeated Sequences in DNA', *Science*, 161(3841), pp.

529–540. doi: 10.1126/science.161.3841.529.

Broughton, J. P. *et al.* (2016) 'Pairing beyond the Seed Supports MicroRNA Targeting Specificity', *Molecular Cell*, 64(2), pp. 320–333. doi: 10.1016/j.molcel.2016.09.004.

Broughton, J. P. and Pasquinelli, A. E. (2016) 'A tale of two sequences: microRNA-target chimeric reads', *Genetics Selection Evolution*, 48(1), p. 31. doi: 10.1186/s12711-016-0209-x.

Brouha, B. *et al.* (2002) 'Evidence consistent with human L1 retrotransposition in maternal meiosis I', *American Journal of Human Genetics*, 71(2), pp. 327–336. doi: 10.1086/341722.

Brouha, B. *et al.* (2003) 'Hot L1s account for the bulk of retrotransposition in the human population', *Proceedings of the National Academy of Sciences*, 100(9), pp. 5280–5285. doi: 10.1073/pnas.0831042100.

Brümmer, A. and Hausser, J. (2014) 'MicroRNA binding sites in the coding region of mRNAs: Extending the repertoire of post-transcriptional gene regulation', *BioEssays*, 36(6), pp. 617–626. doi: 10.1002/bies.201300104.

Burns, K. H. (2017) 'Transposable elements in cancer', *Nature Reviews Cancer*, 17(7), pp. 415–424. doi: 10.1038/nrc.2017.35.

Burns, K. H. (2020) 'Our Conflict with Transposable Elements and Its Implications for Human Disease', *Annual Review of Pathology: Mechanisms of Disease*, 15(1), pp. 51–70. doi: 10.1146/annurev-pathmechdis-012419-032633.

Buzdin, A. *et al.* (2003) 'The human genome contains many types of chimeric retrogenes generated through in vivo RNA recombination', *Nucleic Acids Research*, 31(15), pp. 4385–4390. doi: 10.1093/nar/gkg496.

Cai, X., Hagedorn, C. H. and Cullen, B. R. (2004) 'Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs', *RNA*, 10(12), pp. 1957–1966. doi: 10.1261/rna.7135204.

Calin, G. A. *et al.* (2002) 'Frequent deletions and down-regulation of micro-RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia', *Proceedings of the National Academy of Sciences of the United States of America*, 99(24), pp. 15524–15529. doi: 10.1073/pnas.242606799.

Carreira, P. E. *et al.* (2016) 'Evidence for L1-associated DNA rearrangements and negligible L1 retrotransposition in glioblastoma multiforme', *Mobile DNA*, 7(1), p. 21. doi: 10.1186/s13100-016-0076-6.

Castro-Diaz, N. *et al.* (2014) 'Evolutionally dynamic L1 regulation in embryonic stem cells', *Genes and Development*, 28(13), pp. 1397–1409. doi: 10.1101/gad.241661.114.

De Cecco, M. *et al.* (2019) 'L1 drives IFN in senescent cells and promotes age-associated inflammation', *Nature*, 566(7742), pp. 73–78. doi: 10.1038/s41586-018-0784-9.

Chang, T. C. *et al.* (2007) 'Transactivation of miR-34a by p53 Broadly Influences Gene Expression and Promotes Apoptosis', *Molecular Cell*, 26(5), pp. 745–752. doi: 10.1016/j.molcel.2007.05.010.

Chang, T. C. *et al.* (2008) 'Widespread microRNA repression by Myc contributes to tumorigenesis', *Nature Genetics*, 40(1), pp. 43–50. doi: 10.1038/ng.2007.30.

Cheetham, S. W., Faulkner, G. J. and Dinger, M. E. (2020) 'Overcoming challenges and dogmas to understand the functions of pseudogenes', *Nature Reviews Genetics*, 21(3), pp. 191–201. doi: 10.1038/s41576-019-0196-1.

Chekulaeva, M. *et al.* (2011) 'MiRNA repression involves GW182-mediated recruitment of CCR4-NOT through conserved W-containing motifs', *Nature Structural and Molecular Biology*, 18(11), pp. 1218–1226. doi: 10.1038/nsmb.2166.

Chen, Y. *et al.* (2014) 'A DDX6-CNOT1 Complex and W-Binding Pockets in CNOT9 Reveal Direct Links between miRNA Target Recognition and Silencing', *Molecular Cell*, 54(5), pp. 737–750. doi: 10.1016/j.molcel.2014.03.034.

Chendrimada, T. P. *et al.* (2005) 'TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing', *Nature*, 436(7051), pp. 740–744. doi: 10.1038/nature03868.

Chi, S. W. *et al.* (2009) 'Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps',

Nature, 460(7254), pp. 479–486. doi: 10.1038/nature08170.

Chin, L. J. *et al.* (2008) 'A SNP in a let-7 microRNA complementary site in the KRAS 3' untranslated region increases non-small cell lung cancer risk', *Cancer Research*, 68(20), pp. 8535–8540. doi: 10.1158/0008-5472.CAN-08-2129.

Chu, W. M., Liu, W. M. and Schmid, C. W. (1995) 'RNA polymerase III promoter and terminator elements affect Alu RNA expression', *Nucleic Acids Research*, 23(10), pp. 1750–1757. doi: 10.1093/nar/23.10.1750.

Chuong, E. B., Elde, N. C. and Feschotte, C. (2016) 'Regulatory evolution of innate immunity through co-option of endogenous retroviruses', *Science*, 351(6277), pp. 1083–1087. doi: 10.1126/science.aad5497.

Chuong, E. B., Elde, N. C. and Feschotte, C. (2017) 'Regulatory activities of transposable elements: From conflicts to benefits', *Nature Reviews Genetics*, 18(2), pp. 71–86. doi: 10.1038/nrg.2016.139.

Conrad, T. *et al.* (2014) 'Microprocessor activity controls differential miRNA biogenesis in vivo', *Cell Reports*, 9(2), pp. 542–554. doi: 10.1016/j.celrep.2014.09.007.

Cordaux, R. and Batzer, M. A. (2009) 'The impact of retrotransposons on human genome evolution.', *Nature Reviews Genetics*, 10(10), pp. 691–703. doi: 10.1038/nrg2640.

Cosby, R. L., Chang, N. C. and Feschotte, C. (2019) 'Host–transposon interactions: Conflict, cooperation, and cooption', *Genes and Development*, 33(17–18), pp. 1098–1116. doi: 10.1101/gad.327312.119.

Cost, G. J. *et al.* (2002) 'Human L1 target-primed reverse transcription in vitro', *EMBO Journal*, 21(21), pp. 5899–5910. doi: 10.1093/emboj/cdf592.

Coufal, N. G. *et al.* (2009) 'L1 retrotransposition in human neural progenitor cells', *Nature*, 460(7259), pp. 1127–1131. doi: 10.1038/nature08248.

Coufal, N. G. *et al.* (2011) 'Ataxia telangiectasia mutated (ATM) modulates long interspersed element-1 (L1) retrotransposition in human neural stem cells', *Proceedings of the National Academy of Sciences of the United States of America*, 108(51), pp. 20382–20387. doi: 10.1073/pnas.1100273108.

Criscione, S. W. *et al.* (2016) 'Genome-wide characterization of human L1 antisense promoter-driven transcripts', *BMC Genomics*, 17(1), pp. 1–15. doi: 10.1186/s12864-016-2800-5.

Croce, C. M. (2009) 'Causes and consequences of microRNA dysregulation in cancer.', *Nature Reviews Genetics*, 10(10), pp. 704–14. doi: 10.1038/nrg2634.

Denli, A. M. *et al.* (2015) 'Primate-Specific ORF0 Contributes to Retrotransposon-Mediated Diversity', *Cell*, 163(3), pp. 583–893. doi: 10.1016/j.cell.2015.09.025.

Dewannieux, M., Esnault, C. and Heidmann, T. (2003) 'LINE-mediated retrotransposition of marked Alu sequences.', *Nature Genetics*, 35(1), pp. 41–48. doi: 10.1038/ng1223.

Dexheimer, P. J. and Cochella, L. (2020) 'MicroRNAs: From Mechanism to Organism', *Frontiers in Cell and Developmental Biology*, 8(June), pp. 1–18. doi: 10.3389/fcell.2020.00409.

Dmitriev, S. E. *et al.* (2007) 'Efficient Translation Initiation Directed by the 900-Nucleotide-Long and GC-Rich 5' Untranslated Region of the Human Retrotransposon LINE-1 mRNA Is Strictly Cap Dependent Rather than Internal Ribosome Entry Site Mediated', *Molecular and Cellular Biology*, 27(13), pp. 4685–4697. doi: 10.1128/MCB.02138-06.

Dombroski, B. *et al.* (1991) 'Isolation of an active human transposable element.', *Science*, 254(5039), pp. 1805–1808. doi: 10.1126/science.1662412.

Doolittle, W. F. and Sapienza, C. (1980) 'Selfish genes, the phenotype paradigm and genome evolution', *Nature*, 284(5757), pp. 601–603. doi: 10.1038/284601a0.

Doucet-O'Hare, T. T. *et al.* (2015) 'LINE-1 expression and retrotransposition in Barrett's esophagus and esophageal carcinoma.', *Proceedings of the National Academy of Sciences of the United States of America*, 112(35), pp. E4894-900. doi: 10.1073/pnas.1502474112.

Doucet, A. J. *et al.* (2010) 'Characterization of LINE-1 ribonucleoprotein particles', *PLoS Genetics*, 6(10), pp. 1–19. doi: 10.1371/journal.pgen.1001150.

Doucet, A. J., Wilusz, J. E., *et al.* (2015) 'A 3' Poly(A) Tract Is Required for LINE-1

Retrotransposition', *Molecular Cell*, 60(5), pp. 728–741. doi: 10.1016/j.molcel.2015.10.012.

Doucet, A. J., Droc, G., *et al.* (2015) 'U6 snRNA Pseudogenes: Markers of Retrotransposition Dynamics in Mammals', *Molecular Biology and Evolution*, 32(7), pp. 1815–1832. doi: 10.1093/molbev/msv062.

Doudna, J. A. and Charpentier, E. (2014) 'The new frontier of genome engineering with CRISPR-Cas9', *Science*, 346(6213), pp. 1258096–1258096. doi: 10.1126/science.1258096.

DuBridge, R. B. *et al.* (1987) 'Analysis of mutation in human cells by using an Epstein-Barr virus shuttle system.', *Molecular and Cellular Biology*, 7(1), pp. 379–387. doi: 10.1128/mcb.7.1.379.

Dueck, A. *et al.* (2012) 'MicroRNAs associated with the different human Argonaute proteins', *Nucleic Acids Research*, 40(19), pp. 9850–9862. doi: 10.1093/nar/gks705.

Duursma, A. M. *et al.* (2008) 'miR-148 targets human DNMT3b protein coding region', *RNA*, 14(5), pp. 872–877. doi: 10.1261/rna.972008.

Ebert, M. S., Neilson, J. R. and Sharp, P. a (2007) 'MicroRNA sponges: competitive inhibitors of small RNAs in mammalian cells.', *Nature Methods*, 4(9), pp. 721–726. doi: 10.1038/nmeth1079.

Ecco, G., Imbeault, M. and Trono, D. (2017) 'KRAB zinc finger proteins', *Development*, 144(15), pp. 2719–2729. doi: 10.1242/dev.132605.

Eichhorn, S. W. *et al.* (2014) 'MRNA Destabilization Is the dominant effect of mammalian microRNAs by the time substantial repression ensues', *Molecular Cell*, 56(1), pp. 104–115. doi: 10.1016/j.molcel.2014.08.028.

Elbarbary, R. A., Lucas, B. A. and Maquat, L. E. (2016) 'Retrotransposons as regulators of gene expression.', *Science*, 351(6274), p. aac7247. doi: 10.1126/science.aac7247.

Ergün, S. *et al.* (2004) 'Cell type-specific expression of LINE-1 open reading frames 1 and 2 in fetal and adult human tissues', *Journal of Biological Chemistry*, 279(26), pp. 27753–27763. doi: 10.1074/jbc.M312985200.

Esnault, C., Maestre, J. and Heidmann, T. (2000) 'Human LINE retrotransposons generate processed pseudogenes.', *Nature Genetics*, 24(april), pp. 363–367. doi: 10.1038/74184.

Esquela-Kerscher, A. and Slack, F. J. (2006) 'Oncomirs - microRNAs with a role in cancer.', *Nature Reviews Cancer*, 6(4), pp. 259–69. doi: 10.1038/nrc1840.

Evrony, G. D. *et al.* (2012) 'Single-neuron sequencing analysis of l1 retrotransposition and somatic mutation in the human brain', *Cell*, 151(3), pp. 483–496. doi: 10.1016/j.cell.2012.09.035.

Evrony, G. D. *et al.* (2016) 'Resolving rates of mutation in the brain using single-neuron genomics', *eLife*, 5, pp. 1–32. doi: 10.7554/eLife.12966.

Ewing, A. D. *et al.* (2015) 'Widespread somatic L1 retrotransposition occurs early during gastrointestinal cancer evolution', *Genome Research*, 25(10), pp. 1536–1545. doi: 10.1101/gr.196238.115.

Ewing, A. D. and Kazazian, H. H. (2010) 'High-throughput sequencing reveals extensive variation in human-specific L1 content in individual human genomes.', *Genome research*, 20(9), pp. 1262–70. doi: 10.1101/gr.106419.110.

Fabian, M. R. *et al.* (2011) 'MiRNA-mediated deadenylation is orchestrated by GW182 through two conserved motifs that interact with CCR4-NOT', *Nature Structural and Molecular Biology*, 18(11), pp. 1211–1217. doi: 10.1038/nsmb.2149.

Fang, W. and Bartel, D. P. (2015) 'The Menu of Features that Define Primary MicroRNAs and Enable De Novo Design of MicroRNA Genes', *Molecular Cell*, 60(1), pp. 131–145. doi: 10.1016/j.molcel.2015.08.015.

Farabaugh, P. J. and Fink, G. R. (1980) 'Insertion of the eukaryotic transposable element Ty1 creates a 5-base pair duplication', *Nature*, 286(5771), pp. 352–356. doi: 10.1038/286352a0.

Fassina, A., Cappellesso, R. and Fassan, M. (2011) 'Classification of non-small cell lung carcinoma in transthoracic needle specimens using microRNA expression profiling', *Chest*, 140(5), pp. 1305–1311. doi: 10.1378/chest.11-0708.

Faulkner, G. J. and Garcia-Perez, J. L. (2017) 'L1 Mosaicism in Mammals: Extent, Effects, and Evolution', *Trends in Genetics*, 33(11), pp. 802–816. doi: 10.1016/j.tig.2017.07.004.

Fedoroff, N., Wessler, S. and Shure, M. (1983) 'Isolation of the transposable maize controlling elements Ac and Ds', *Cell*, 35(1), pp. 235–242. doi: 10.1016/0092-8674(83)90226-X.

Feng, Q. *et al.* (1996) 'Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition', *Cell*, 87(5), pp. 905–916. doi: 10.1016/S0092-8674(00)81997-2.

Feusier, J. *et al.* (2019) 'Pedigree-based estimation of human mobile element retrotransposition rates', *Genome Research*, 29(10), pp. 1567–1577. doi: 10.1101/gr.247965.118.

Finnegan, D. J. (1989) 'Eukaryotic transposable elements and genome evolution.', *Trends in Genetics*, 5(4), pp. 103–7. doi: 10.1016/0168-9525(89)90039-5.

Fire, A. *et al.* (1998) 'Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans', *Nature*, 391(6669), pp. 806–811. doi: 10.1038/35888.

Flasch, D. A. *et al.* (2019) 'Genome-wide de novo L1 Retrotransposition Connects Endonuclease Activity with Replication', *Cell*, 177(4), pp. 837-851.e28. doi: 10.1016/j.cell.2019.02.050.

Flores, O. *et al.* (2014) 'Differential RISC association of endogenous human microRNAs predicts their inhibitory potential', *Nucleic Acids Research*, 42(7), pp. 4629–4639. doi: 10.1093/nar/gkt1393.

Forman, J. J., Legesse-Miller, A. and Coller, H. A. (2008) 'A search for conserved sequences in coding regions reveals that the let-7 microRNA targets Dicer within its coding sequence', *Proceedings of the National Academy of Sciences*, 105(39), pp. 14879–14884. doi: 10.1073/pnas.0803230105.

Freeman, J. D., Goodchild, N. L. and Mager, D. L. (1994) 'A modified indicator gene for selection of retrotransposition events in mammalian cells.', *BioTechniques*, 17(1), pp. 46,48-49,52.

Friedli, M. and Trono, D. (2015) 'The Developmental Control of Transposable Elements and the Evolution of Higher Species', *Annual Review of Cell and Developmental Biology*, 31(1), pp. 429–451. doi: 10.1146/annurev-cellbio-100814-125514.

Friedman, R. C. *et al.* (2009) 'Most mammalian mRNAs are conserved targets of microRNAs', *Genome Research*, 19(1), pp. 92–105. doi: 10.1101/gr.082701.108.

Fromm, B. *et al.* (2015) 'A Uniform System for the Annotation of Vertebrate microRNA Genes and the Evolution of the Human microRNAome', *Annual Review of Genetics*, 49(1), pp. 213–242. doi: 10.1146/annurev-genet-120213-092023.

Fu, X. *et al.* (2020) 'Dynamic miRNA-mRNA interactions coordinate gene expression in adult Anopheles gambiae', *PLOS Genetics*, 16(4), p. e1008765. doi: 10.1371/journal.pgen.1008765.

Fukao, A. *et al.* (2014) 'MicroRNAs trigger dissociation of eIF4AI and eIF4AII from target mRNAs in humans', *Molecular Cell*, 56(1), pp. 79–89. doi: 10.1016/j.molcel.2014.09.005.

Garcia-Perez, J. L., Doucet, A. J., *et al.* (2007) 'Distinct mechanisms for trans-mediated mobilization of cellular RNAs by the LINE-1 reverse transcriptase', *Genome Research*, 17(5), pp. 602–611. doi: 10.1101/gr.5870107.

Garcia-Perez, J. L., Marchetto, M. C. N., *et al.* (2007) 'LINE-1 retrotransposition in human embryonic stem cells', *Human Molecular Genetics*, 16(13), pp. 1569–1577. doi: 10.1093/hmg/ddm105.

Garcia-Perez, J. L. *et al.* (2010) 'Epigenetic silencing of engineered L1 retrotransposition events in human embryonic carcinoma cells', *Nature*, 466(7307), pp. 769–773. doi: 10.1038/nature09209.

Garcia-Perez, J. L., Widmann, T. J. and Adams, I. R. (2016) 'The impact of transposable elements on mammalian development', *Development*, 143(22), pp. 4101–4114. doi: 10.1242/dev.132639.

Gardner, E. J. *et al.* (2017) 'The Mobile Element Locator Tool (MELT): population-scale mobile element discovery and biology', *Genome Research*, 27(11), pp. 1916–1929. doi: 10.1101/gr.218032.116.

Gebert, L. F. R. and MacRae, I. J. (2019) 'Regulation of microRNA function in animals', *Nature Reviews Molecular Cell Biology*, 20(1), pp. 21–37. doi: 10.1038/s41580-018-0045-7.

Georgiou, I. *et al.* (2009) 'Retrotransposon RNA expression and evidence for retrotransposition events in human oocytes', *Human Molecular Genetics*, 18(7), pp. 1221–1228. doi: 10.1093/hmg/ddp022.

Gilbert, C. and Feschotte, C. (2018) 'Horizontal acquisition of transposable elements and viral sequences: patterns and consequences', *Current Opinion in Genetics and Development*, 49(2018), pp. 15–24. doi: 10.1016/j.gde.2018.02.007.

Gilbert, N., Lutz-Prigge, S. and Moran, J. V. (2002) 'Genomic deletions created upon LINE-1 retrotransposition', *Cell*, 110(3), pp. 315–325. doi: 10.1016/S0092-8674(02)00828-0.

Gilles, M. E. and Slack, F. J. (2018) 'Let-7 microRNA as a potential therapeutic target with implications for immunotherapy', *Expert Opinion on Therapeutic Targets*, 22(11), pp. 929–939. doi: 10.1080/14728222.2018.1535594.

Goerner-Potvin, P. and Bourque, G. (2018) 'Computational tools to unmask transposable elements', *Nature Reviews Genetics*, 19(11), pp. 688–704. doi: 10.1038/s41576-018-0050-x.

Goodier, J. L. *et al.* (2001) 'A novel active L1 retrotransposon subfamily in the mouse', *Genome Research*, 11(10), pp. 1677–1685. doi: 10.1101/gr.198301.

Goodier, J. L. *et al.* (2007) 'LINE-1 ORF1 protein localizes in stress granules with other RNA-binding proteins, including components of RNA interference RNA-induced silencing complex.', *Molecular and cellular biology*, 27(18), pp. 6469–6483. doi: 10.1128/MCB.00332-07.

Goodier, J. L. *et al.* (2015) 'The Broad-Spectrum Antiviral Protein ZAP Restricts Human Retrotransposition', *PLoS Genetics*, 11(5), pp. 1–32. doi: 10.1371/journal.pgen.1005252.

Goodier, J. L. (2016) 'Restricting retrotransposons: a review', *Mobile DNA*, 7(1), p. 16. doi: 10.1186/s13100-016-0070-z.

Goodier, J. L., Cheung, L. E. and Kazazian, H. H. (2012) 'MOV10 RNA Helicase Is a Potent Inhibitor of Retrotransposition in Cells', *PLoS Genetics*, 8(10). doi: 10.1371/journal.pgen.1002941.

Goodier, J. L., Cheung, L. E. and Kazazian, H. H. (2013) 'Mapping the LINE1 ORF1 protein interactome reveals associated inhibitors of human retrotransposition', *Nucleic Acids Research*, 41(15), pp. 7401–7419. doi: 10.1093/nar/gkt512.

Goodier, J. L. and Kazazian, H. H. (2008) 'Retrotransposons Revisited: The Restraint and Rehabilitation of Parasites', *Cell*, 135(1), pp. 23–35. doi: 10.1016/j.cell.2008.09.022.

Gregory, R. I. *et al.* (2004) 'The Microprocessor complex mediates the genesis of microRNAs.', *Nature*, 432(7014), pp. 235–240. doi: 10.1038/nature03120.

Griffiths-Jones, S. *et al.* (2006) 'miRBase: microRNA sequences, targets and gene nomenclature.', *Nucleic acids research*, 34(Database issue), pp. D140-4. doi: 10.1093/nar/gkj112.

Griffiths-Jones, S. *et al.* (2008) 'miRBase: Tools for microRNA genomics', *Nucleic Acids Research*, 36, pp. 154–158. doi: 10.1093/nar/gkm952.

Grimaldi, G., Skowronski, J. and Singer, M. F. (1984) 'Defining the beginning and end of KpnI family segments.', *The EMBO journal*, 3(8), pp. 1753–9. doi: 10.1002/j.1460-2075.1984.tb02042.x.

Grishok, A. *et al.* (2001) 'Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control C. elegans developmental timing', *Cell*, 106(1), pp. 23–34. doi: 10.1016/S0092-8674(01)00431-7.

Gu, S. *et al.* (2009) 'Biological basis for restriction of microRNA targets to the 3' untranslated region in mammalian mRNAs', *Nature Structural and Molecular Biology*, 16(2), pp. 144–150. doi: 10.1038/nsmb.1552.

Guichard, E. *et al.* (2018) 'Impact of non-LTR retrotransposons in the differentiation and evolution of anatomically modern humans', *Mobile DNA*, 9(1), pp. 1–19. doi: 10.1186/s13100-018-0133-4.

Guil, S. and Cáceres, J. F. (2007) 'The multifunctional RNA-binding protein hnRNP A1 is required for processing of miR-18a', *Nature Structural and Molecular Biology*, 14(7), pp. 591–596. doi: 10.1038/nsmb1250.

Guo, H. *et al.* (2010) 'Mammalian microRNAs predominantly act to decrease target mRNA

levels', *Nature*, 466(7308), pp. 835–840. doi: 10.1038/nature09267.

Ha, M. and Kim, V. N. (2014) 'Regulation of microRNA biogenesis.', *Nature reviews. Molecular cell biology*, 15(8), pp. 509–524. doi: 10.1038/nrm3838.

Hamdorf, M. *et al.* (2015) 'MiR-128 represses L1 retrotransposition by binding directly to L1 RNA', *Nature Structural and Molecular Biology*, 22(10), pp. 824–831. doi: 10.1038/nsmb.3090.

Han, J. *et al.* (2009) 'Posttranscriptional Crossregulation between Drosha and DGCR8', *Cell*, 136(1), pp. 75–84. doi: 10.1016/j.cell.2008.10.053.

Han, J. S., Szak, S. T. and Boeke, J. D. (2004) 'Transcriptional disruption by the L1 retrotransposon and implications for mammalian transcriptomes', *Nature*, 429(6989), pp. 268–274. doi: 10.1038/nature02536.

Hancks, D. C. *et al.* (2011) 'Retrotransposition of marked SVA elements by human L1s in cultured cells', *Human Molecular Genetics*, 20(17), pp. 3386–3400. doi: 10.1093/hmg/ddr245.

Hancks, D. C. and Kazazian, H. H. (2010) 'SVA retrotransposons: Evolution and genetic instability', *Seminars in Cancer Biology*, 20(4), pp. 234–245. doi: 10.1016/j.semcancer.2010.04.001.

Hancks, D. C. and Kazazian, H. H. (2016) 'Roles for retrotransposon insertions in human disease', *Mobile DNA*, 7(1), p. 9. doi: 10.1186/s13100-016-0065-9.

Hannan, A. J. (2018) 'Tandem repeats mediating genetic plasticity in health and disease', *Nature Reviews Genetics*, 19(5), pp. 286–298. doi: 10.1038/nrg.2017.115.

Hansen, T. B. *et al.* (2013) 'Natural RNA circles function as efficient microRNA sponges', *Nature*, 495(7441), pp. 384–388. doi: 10.1038/nature11993.

Hata, K. and Sakaki, Y. (1997) 'Identification of critical CpG sites for repression of L1 transcription by DNA methylation', *Gene*, 189(2), pp. 227–234. doi: 10.1016/S0378-1119(96)00856-6.

Hatley, M. E. *et al.* (2010) 'Modulation of K-Ras-dependent lung tumorigenesis by MicroRNA-21', *Cancer Cell*, 18(3), pp. 282–293. doi: 10.1016/j.ccr.2010.08.013.

Hausser, J. *et al.* (2013) 'Analysis of CDS-located miRNA target sites suggests that they can effectively inhibit translation', *Genome Research*, 23(4), pp. 604–615. doi: 10.1101/gr.139758.112.

Hausser, J. and Zavolan, M. (2014) 'Identification and consequences of miRNA-target interactions - beyond repression of gene expression.', *Nature Reviews Genetics*, 15(9), pp. 599–612. doi: 10.1038/nrg3765.

He, L. *et al.* (2005) 'A microRNA polycistron as a potential human oncogene', *Nature*, 435(7043), pp. 828–833. doi: 10.1038/nature03552.

He, L. *et al.* (2007) 'A microRNA component of the p53 tumour suppressor network', *Nature*, 447(7148), pp. 1130–1134. doi: 10.1038/nature05939.

Heckman, K. L. and Pease, L. R. (2007) 'Gene splicing and mutagenesis by PCR-driven overlap extension', *Nature Protocols*, 2(4), pp. 924–932. doi: 10.1038/nprot.2007.132.

Helman, E. *et al.* (2014) 'Somatic retrotransposition in human cancer revealed by whole-genome and exome sequencing', *Genome Research*, 24(7), pp. 1053–1063. doi: 10.1101/gr.163659.113.

Helwak, A. *et al.* (2013) 'Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding', *Cell*, 153(3), pp. 654–665. doi: 10.1016/j.cell.2013.03.043.

Helwak, A. and Tollervey, D. (2014) 'Mapping the miRNA interactome by cross-linking ligation and sequencing of hybrids (CLASH)', *Nature Protocols*, 9(3), pp. 711–728. doi: 10.1038/nprot.2014.043.

Hentze, M. W. *et al.* (2018) 'A brave new world of RNA-binding proteins', *Nature Reviews Molecular Cell Biology*, 19(5), pp. 327–341. doi: 10.1038/nrm.2017.130.

Heo, I. *et al.* (2008) 'Lin28 Mediates the Terminal Uridylation of let-7 Precursor MicroRNA', *Molecular Cell*, 32(2), pp. 276–284. doi: 10.1016/j.molcel.2008.09.014.

Heras, S. R. *et al.* (2013) 'The Microprocessor controls the activity of mammalian retrotransposons', *Nature Structural and Molecular Biology*, 20(10), pp. 1173–1183. doi: 10.1038/nsmb.2658.

Hertel, J. *et al.* (2012) 'Evolution of the let-7 microRNA Family', *RNA Biology*, 9(3), pp. 231–241. doi: 10.4161/rna.18974.

Hickman, A. B. and Dyda, F. (2015) 'Mechanisms of DNA Transposition', in *Mobile DNA III*, pp. 529–553. doi: 10.1128/9781555819217.ch25.

Hiom, K., Melek, M. and Gellert, M. (1998) 'DNA transposition by the RAG1 and RAG2 proteins: a possible source of oncogenic translocations.', *Cell*, 94(4), pp. 463–70. doi: 10.1016/s0092-8674(00)81587-1.

Hocq, R. *et al.* (2018) 'Monitored ECLIP: High accuracy mapping of RNA-protein interactions', *Nucleic Acids Research*, 46(21), pp. 11553–11565. doi: 10.1093/nar/gky858.

Hohjoh, H. and Singer, M. F. (1996) 'Cytoplasmic ribonucleoprotein complexes containing human LINE-1 protein and RNA.', *The EMBO Journal*, 15(3), pp. 630–639. doi: 10.1002/j.1460-2075.1996.tb00395.x.

Hohjoh, H., Singer, M. F. and Nw, P. S. (1997) 'Sequence-specific single-strand RNA binding protein encoded by the human LINE-1 retrotransposon', *EMBO Journal*, 16(19), pp. 6034–6043.

Houck, C. M., Rinehart, F. P. and Schmid, C. W. (1979) 'A ubiquitous family of repeated DNA sequences in the human genome.', *Journal of molecular biology*, 132(3), pp. 289–306. doi: 10.1016/0022-2836(79)90261-4.

Howe, K. *et al.* (2013) 'The zebrafish reference genome sequence and its relationship to the human genome', *Nature*, 496(7446), pp. 498–503. doi: 10.1038/nature12111.

Howell, R. and Usdin, K. (1997) 'The ability to form intrastrand tetraplexes is an evolutionarily conserved feature of the 3' end of L1 retrotransposons', *Molecular Biology and Evolution*, 14(2), pp. 144–155. doi: 10.1093/oxfordjournals.molbev.a025747.

Hsu, S.-D. *et al.* (2011) 'miRTarBase: a database curates experimentally validated microRNA–target interactions', *Nucleic Acids Research*, 39, pp. D163–D169. doi: 10.1093/nar/gkq1107.

Huang, C. R. L. *et al.* (2010) 'Mobile interspersed repeats are major structural variants in the human genome', *Cell*, 141(7), pp. 1171–1182. doi: 10.1016/j.cell.2010.05.026.

Huang, D. *et al.* (2018) 'MiR-20a, a novel promising biomarker to predict prognosis in human cancer: A meta-analysis', *BMC Cancer*. BMC Cancer, 18(1), pp. 1–14. doi: 10.1186/s12885-018-4907-3.

Huang, H.-Y. *et al.* (2019) 'miRTarBase 2020: updates to the experimentally validated microRNA–target interaction database', *Nucleic Acids Research*, 48(D1), pp. D148–D154. doi: 10.1093/nar/gkz896.

Huang, S. *et al.* (2016) 'Discovery of an Active RAG Transposon Illuminates the Origins of V(D)J Recombination', *Cell*, 166(1), pp. 102–114. doi: 10.1016/j.cell.2016.05.032.

van den Hurk, J. A. J. M. *et al.* (2007) 'L1 retrotransposition can occur early in human embryonic development', *Human Molecular Genetics*, 16(13), pp. 1587–1592. doi: 10.1093/hmg/ddm108.

Hutvágner, G. *et al.* (2001) 'A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA.', *Science*, 293(5531), pp. 834–8. doi: 10.1126/science.1062961.

Imbeault, M., Helleboid, P. Y. and Trono, D. (2017) 'KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks', *Nature*, 543(7646), pp. 550–554. doi: 10.1038/nature21683.

Inamura, K. and Ishikawa, Y. (2016) 'MicroRNA In Lung Cancer: Novel Biomarkers and Potential Tools for Treatment', *Journal of Clinical Medicine*, 5(3), p. 36. doi: 10.3390/jcm5030036.

Iskow, R. C. *et al.* (2010) 'Natural mutagenesis of human genomes by endogenous retrotransposons', *Cell*, 141(7), pp. 1253–1261. doi: 10.1016/j.cell.2010.05.020.

Iwasaki, S. *et al.* (2010) 'Hsc70/Hsp90 chaperone machinery mediates ATP-dependent RISC loading of small RNA duplexes', *Molecular Cell*, 39(2), pp. 292–299. doi: 10.1016/j.molcel.2010.05.015.

Izaurralde, E. (2015) 'Breakers and blockers - miRNAs at work', *Science*, 349(6246), pp. 380–

382. doi: 10.1126/science.1260969.

Jacob, F. and Monod, J. (1961) 'Genetic regulatory mechanisms in the synthesis of proteins', *Journal of Molecular Biology*, 3(3), pp. 318–356. doi: 10.1016/S0022-2836(61)80072-7.

Jacobs, F. M. J. *et al.* (2014) 'An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons.', *Nature*, 516(7530), pp. 242–5. doi: 10.1038/nature13760.

John, B. *et al.* (2004) 'Human microRNA targets', *PLoS Biology*, 2(11). doi: 10.1371/journal.pbio.0020363.

Johnson, C. D. *et al.* (2007) 'The let-7 microRNA represses cell proliferation pathways in human cells', *Cancer Research*, 67(16), pp. 7713–7722. doi: 10.1158/0008-5472.CAN-07-1083.

Johnson, S. M. *et al.* (2005) 'RAS is regulated by the let-7 microRNA family', *Cell*, 120(5), pp. 635–647. doi: 10.1016/j.cell.2005.01.014.

Jonas, S. and Izaurralde, E. (2015) 'Towards a molecular understanding of microRNA-mediated gene silencing', *Nature Reviews Genetics*, 16(7), pp. 421–433. doi: 10.1038/nrg3965.

Jönsson, M. E. *et al.* (2019) 'Activation of neuronal genes via LINE-1 elements upon global DNA demethylation in human neural progenitors', *Nature Communications*, 10(1). doi: 10.1038/s41467-019-11150-8.

Jurka, J., Zietkiewicz, E. and Labuda, D. (1995) 'Ubiquitous mammalian-wide interspersed repeats (MIRs) are molecular fossils from the mesozoic era', *Nucleic Acids Research*, 23(1), pp. 170–175. doi: 10.1093/nar/23.1.170.

Kallen, A. N. *et al.* (2013) 'The Imprinted H19 LncRNA Antagonizes Let-7 MicroRNAs', *Molecular Cell*, 52(1), pp. 101–112. doi: 10.1016/j.molcel.2013.08.027.

Kannan, M. *et al.* (2017) 'Dynamic silencing of somatic L1 retrotransposon insertions reflects the developmental and cellular contexts of their genomic integration', *Mobile DNA*, 8(1), p. 8. doi: 10.1186/s13100-017-0091-2.

Kano, H. *et al.* (2009) 'L1 retrotransposition occurs mainly in embryogenesis and creates somatic mosaicism', *Genes and Development*, 23(11), pp. 1303–1312. doi: 10.1101/gad.1803909.

Kapitonov, V. V. and Jurka, J. (2005) 'RAG1 core and V(D)J recombination signal sequences were derived from Transib transposons', *PLoS Biology*, 3(6), pp. 0998–1011. doi: 10.1371/journal.pbio.0030181.

Karube, Y. *et al.* (2005) 'Reduced expression of Dicer associated with poor prognosis in lung cancer patients.', *Cancer science*, 96(2), pp. 111–5. doi: 10.1111/j.1349-7006.2005.00015.x.

Kawano, K. *et al.* (2018) 'HIV-1 Vpr and p21 restrict LINE-1 mobility', *Nucleic Acids Research*, 46(16), pp. 8454–8470. doi: 10.1093/nar/gky688.

Kazazian, H. H. *et al.* (1988) 'Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man.', *Nature*, 332(6160), pp. 164–166. doi: 10.1038/332164a0.

Kazazian, H. H. and Moran, J. V. (2017) 'Mobile DNA in Health and Disease', *New England Journal of Medicine*, 377(4), pp. 361–370. doi: 10.1056/NEJMra1510092.

Kelley, D. and Rinn, J. (2012) 'Transposable elements reveal a stem cell-specific class of long noncoding RNAs', *Genome biology*, 13(11), p. R107. doi: 10.1186/gb-2012-13-11-r107.

Khadgi, B. B., Govindaraju, A. and Christensen, S. M. (2019) 'Completion of LINE integration involves an open "4-way" branched DNA intermediate', *Nucleic Acids Research*, 47(16), pp. 8708–8719. doi: 10.1093/nar/gkz673.

Khan, A. A. *et al.* (2009) 'Transfection of small RNAs globally perturbs gene regulation by endogenous microRNAs.', *Nature biotechnology*, 27(6), pp. 549–55. doi: 10.1038/nbt.1543.

Khan, H., Smit, A. and Boissinot, S. (2006) 'Molecular evolution and tempo of amplification of human LINE-1 retrotransposons since the origin of primates', *Genome Research*, 16(1), pp. 78–87. doi: 10.1101/gr.4001406.

Khazina, E. *et al.* (2011) 'Trimeric structure and flexibility of the L1ORF1 protein in human L1 retrotransposition', *Nature Structural and Molecular Biology*, 18(9), pp. 1006–1014. doi:

10.1038/nsmb.2097.

Khazina, E. and Weichenrieder, O. (2009) 'Non-LTR retrotransposons encode noncanonical RRM domains in their first open reading frame', *Proceedings of the National Academy of Sciences*, 106(3), pp. 731–736. doi: 10.1073/pnas.0809964106.

Kim, D. *et al.* (2016) 'General rules for functional microRNA targeting', *Nature Genetics*, 48(12), pp. 1517–1526. doi: 10.1038/ng.3694.

Kim, V. N., Han, J. and Siomi, M. C. (2009) 'Biogenesis of small RNAs in animals', *Nat Rev Mol Cell Biol*, 10(2), pp. 126–139. doi: 10.1038/nrm2632.

Kim, Y. K., Kim, B. and Kim, V. N. (2016) 'Re-evaluation of the roles of DROSHA, Exportin 5, and DICER in microRNA biogenesis', *Proceedings of the National Academy of Sciences of the United States of America*, 113(13), pp. E1881–E1889. doi: 10.1073/pnas.1602532113.

Kimberland, M. L. *et al.* (1999) 'Full-length human L1 insertions retain the capacity for high frequency retrotransposition in cultured cells.', *Human molecular genetics*, 8(8), pp. 1557–60. doi: 10.1093/hmg/8.8.1557.

Klawitter, S. *et al.* (2016) 'Reprogramming triggers endogenous L1 and Alu retrotransposition in human induced pluripotent stem cells', *Nature Communications*, 7, p. 10286. doi: 10.1038/ncomms10286.

Kleaveland, B. *et al.* (2018) 'A Network of Noncoding Regulatory RNAs Acts in the Mammalian Brain', *Cell*, 174(2), pp. 350-362.e17. doi: 10.1016/j.cell.2018.05.022.

Kloosterman, W. P. and Plasterk, R. H. A. (2006) 'The Diverse Functions of MicroRNAs in Animal Development and Disease', *Developmental Cell*, 11(4), pp. 441–450. doi: 10.1016/j.devcel.2006.09.009.

de Koning, A. P. J. *et al.* (2011) 'Repetitive elements may comprise over Two-Thirds of the human genome', *PLoS Genetics*, 7(12). doi: 10.1371/journal.pgen.1002384.

Konkel, M. K. *et al.* (2015) 'Sequence analysis and characterization of active human alu subfamilies based on the 1000 genomes pilot project', *Genome Biology and Evolution*, 7(9), pp. 2608–2622. doi: 10.1093/gbe/evv167.

Konkel, M. K. and Batzer, M. A. (2010) 'A mobile threat to genome stability: The impact of non-LTR retrotransposons upon the human genome', *Seminars in Cancer Biology*, 20(4), pp. 211–221. doi: 10.1016/j.semcancer.2010.03.001.

Kopera, H. C. *et al.* (2011) 'Similarities between long interspersed element-1 (LINE-1) reverse transcriptase and telomerase.', *Proceedings of the National Academy of Sciences of the United States of America*, 108(51), pp. 20345–50. doi: 10.1073/pnas.1100275108.

Kopera, H. C. *et al.* (2016) 'LINE-1 Cultured Cell Retrotransposition Assay.', *Methods in molecular biology*, 1400, pp. 139–56. doi: 10.1007/978-1-4939-3372-3_10.

Kozomara, A., Birgaoanu, M. and Griffiths-Jones, S. (2019) 'miRBase: from microRNA sequences to function', *Nucleic Acids Research*, 47(D1), pp. D155–D162. doi: 10.1093/nar/gky1141.

Krebs, J. E., Goldstein, E. S. and Kilpatrick, S. T. (2014) *Lewin's Genes XI*, *Transfusion Medicine*.

Kubo, S. *et al.* (2006) 'L1 retrotransposition in nondividing and primary human somatic cells', *Proceedings of the National Academy of Sciences of the United States of America*, 103(21), pp. 8036–8041. doi: 10.1073/pnas.0601954103.

Kudla, G. *et al.* (2011) 'Cross-linking, ligation, and sequencing of hybrids reveals RNA-RNA interactions in yeast', *Proceedings of the National Academy of Sciences*, 108(24), pp. 10010–10015. doi: 10.1073/pnas.1017386108.

Kulpa, D. A. and Moran, J. V. (2005) 'Ribonucleoprotein particle formation is necessary but not sufficient for LINE-1 retrotransposition', *Human Molecular Genetics*, 14(21), pp. 3237–3248. doi: 10.1093/hmg/ddi354.

Kulpa, D. A. and Moran, J. V. (2006) 'Cis-preferential LINE-1 reverse transcriptase activity in ribonucleoprotein particles', *Nature Structural and Molecular Biology*, 13(7), pp. 655–660. doi: 10.1038/nsmb1107.

Kumar, M. S. *et al.* (2009) 'Dicer1 functions as a haploinsufficient tumor suppressor', *Genes and Development*, 23(23), pp. 2700–2704. doi: 10.1101/gad.1848209.

Kuppusamy, K. T. *et al.* (2015) 'Let-7 family of microRNA is required for maturation and adult-like metabolism in stem cell-derived cardiomyocytes', *Proceedings of the National Academy of Sciences*, p. 201424042. doi: 10.1073/pnas.1424042112.

Lagos-Quintana, M. *et al.* (2001) 'Identification of novel genes coding for small expressed RNAs', *Science*, 294(5543), pp. 853–858. doi: 10.1126/science.1064921.

Lander, E. S. *et al.* (2001) 'Initial sequencing and analysis of the human genome.', *Nature*, 409(6822), pp. 860–921. doi: 10.1038/35057062.

Larson, P. A. *et al.* (2018) 'Spliced integrated retrotransposed element (SpIRE) formation in the human genome', *PLOS Biology*, 16(3), p. e2003067. doi: 10.1371/journal.pbio.2003067.

Lau, N. C. *et al.* (2001) 'An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans', *Science*, 294(5543), pp. 858–862. doi: 10.1126/science.1065062.

Lee, E. *et al.* (2012) 'Landscape of Somatic Retrotransposition in Human Cancers', *Science*, 337(6097), pp. 967–971. doi: 10.1126/science.1222077.

Lee, F. C. Y. and Ule, J. (2018) 'Advances in CLIP Technologies for Studies of Protein-RNA Interactions', *Molecular Cell*, 69(3), pp. 354–369. doi: 10.1016/j.molcel.2018.01.005.

Lee, R. C. and Ambros, V. (2001) 'An extensive class of small RNAs in Caenorhabditis elegans', *Science*, 294(5543), pp. 862–864. doi: 10.1126/science.1065329.

Lee, R. C., Feinbaum, R. L. and Ambros, V. (1993) 'The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14', *Cell*, 75(5), pp. 843–854. doi: 10.1016/0092-8674(93)90529-Y.

Lee, Y. *et al.* (2003) 'The nuclear RNase III Drosha initiates microRNA processing.', *Nature*, 425(6956), pp. 415–419. doi: 10.1038/nature01957.

Lee, Y. *et al.* (2004) 'MicroRNA genes are transcribed by RNA polymerase II', *EMBO Journal*, 23(20), pp. 4051–4060. doi: 10.1038/sj.emboj.7600385.

Lee, Y. S. *et al.* (2009) 'A novel class of small RNAs: tRNA-derived RNA fragments (tRFs)', *Genes and Development*, 23(22), pp. 2639–2649. doi: 10.1101/gad.1837609.

Lee, Y. S. and Dutta, A. (2007) 'The tumor suppressor microRNA let-7 represses the HMGA2 oncogene', *Genes & Development*, 21(9), pp. 1025–1030. doi: 10.1101/gad.1540407.

Lee, Y. S. and Dutta, A. (2009) 'MicroRNAs in Cancer', *Annual Review of Pathology: Mechanisms of Disease*, 4(1), pp. 199–227. doi: 10.1146/annurev.pathol.4.110807.092222.

Di Leva, G., Garofalo, M. and Croce, C. M. (2014) 'MicroRNAs in Cancer', *Annual Review of Pathology: Mechanisms of Disease*, 9(1), pp. 287–314. doi: 10.1146/annurev-pathol-012513-104715.

Lewis, B. P., Burge, C. B. and Bartel, D. P. (2005) 'Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets', *Cell*, 120(1), pp. 15–20. doi: 10.1016/j.cell.2004.12.035.

Li, P. *et al.* (2017) 'Aicardi-Goutières syndrome protein TREX1 suppresses L1 and maintains genome integrity through exonuclease-independent ORF1p depletion', *Nucleic Acids Research*, 45(8), pp. 4619–4631. doi: 10.1093/nar/gkx178.

Li, X. *et al.* (2013) 'The MOV10 helicase inhibits LINE-1 mobility', *Journal of Biological Chemistry*, 288(29), pp. 21148–21160. doi: 10.1074/jbc.M113.465856.

Lian, S. L. *et al.* (2009) 'The C- Terminal half of human Ago2 binds to multiple GW- rich regions of GW182 and requires GW182 to mediate silencing', *RNA*, 15(5), pp. 804–813. doi: 10.1261/rna.1229409.

Lin, R. J. *et al.* (2010) 'microRNA signature and expression of Dicer and Drosha can predict prognosis and delineate risk groups in neuroblastoma', *Cancer Research*, 70(20), pp. 7841–7850. doi: 10.1158/0008-5472.CAN-10-0970.

Lin, S. and Gregory, R. I. (2015) 'MicroRNA biogenesis pathways in cancer', *Nature Reviews Cancer*, 15(6), pp. 321–333. doi: 10.1038/nrc3932.

Lindtner, S., Felber, B. K. and Kjems, J. (2002) 'An element in the 3' untranslated region of

human LINE-1 retrotransposon mRNA binds NXF1(TAP) and can function as a nuclear export element', *RNA*, 8(3), pp. 345–356. doi: 10.1017/S1355838202027759.

Liu, J. *et al.* (2005) 'A role for the P-body component GW182 in microRNA function', *Nature Cell Biology*, 7(12), pp. 1161–1166. doi: 10.1038/ncb1333.

Liu, N. *et al.* (2018) 'Selective silencing of euchromatic L1s revealed by genome-wide screens for L1 regulators.', *Nature*, 553(7687), pp. 228–232. doi: 10.1038/nature25179.

Lu, J. *et al.* (2005) 'MicroRNA expression profiles classify human cancers', *Nature*, 435(7043), pp. 834–838. doi: 10.1038/nature03702.

Luan, D. D. *et al.* (1993) 'Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: A mechanism for non-LTR retrotransposition', *Cell*, 72(4), pp. 595–605. doi: 10.1016/0092-8674(93)90078-5.

Lubelsky, Y. and Ulitsky, I. (2018) 'Sequences enriched in Alu repeats drive nuclear localization of long RNAs in human cells', *Nature*, 555(7694), pp. 107–111. doi: 10.1038/nature25757.

Lujambio, A. and Lowe, S. W. (2012) 'The microcosmos of cancer', *Nature*, 482(7385), pp. 347–355. doi: 10.1038/nature10888.

Luna, J. M. *et al.* (2015) 'Hepatitis C virus RNA functionally sequesters miR-122', *Cell*, 160(6), pp. 1099–1110. doi: 10.1016/j.cell.2015.02.025.

Lund, E. *et al.* (2004) 'Nuclear export of microRNA precursors.', *Science*, 303(5654), pp. 95–98. doi: 10.1126/science.1090599.

Lytle, J. R., Yario, T. A. and Steitz, J. A. (2007) 'Target mRNAs are repressed as efficiently by microRNA-binding sites in the 5' UTR as in the 3' UTR', *Proceedings of the National Academy of Sciences*, 104(23), pp. 9667–9672. doi: 10.1073/pnas.0703820104.

Ma, L., Teruya-Feldstein, J. and Weinberg, R. A. (2007) 'Tumour invasion and metastasis initiated by microRNA-10b in breast cancer', *Nature*, 449(7163), pp. 682–688. doi: 10.1038/nature06174.

Macia, A. *et al.* (2011) 'Epigenetic control of retrotransposon expression in human embryonic stem cells', *Mol Cell Biol*, 31(2), pp. 300–316. doi: 10.1128/MCB.00561-10.

Macia, A. *et al.* (2017) 'Engineered LINE-1 retrotransposition in nondividing human neurons', *Genome Research*, 27(3), pp. 335–348. doi: 10.1101/gr.206805.116.

MacLennan, M. *et al.* (2017) 'Mobilization of LINE-1 retrotransposons is restricted by Tex19.1 in mouse embryonic stem cells', *eLife*, 6, pp. 1–32. doi: 10.7554/eLife.26152.

Mager, D. L. and Stoye, J. P. (2015) 'Mammalian Endogenous Retroviruses', *Microbiology Spectrum*, pp. 1079–1100. doi: 10.1128/microbiolspec.mdna3-0009-2014.

Malki, S. *et al.* (2014) 'A Role for retrotransposon LINE-1 in fetal oocyte attrition in mice', *Developmental Cell*, 29(5), pp. 521–533. doi: 10.1016/j.devcel.2014.04.027.

Malone, C. D. and Hannon, G. J. (2009) 'Small RNAs as Guardians of the Genome', *Cell*, 136(4), pp. 656–668. doi: 10.1016/j.cell.2009.01.045.

Marson, A. *et al.* (2008) 'Connecting microRNA Genes to the Core Transcriptional Regulatory Circuitry of Embryonic Stem Cells', *Cell*, 134(3), pp. 521–533. doi: 10.1016/j.cell.2008.07.020.

Martin, S. L. (2006) 'The ORF1 Protein Encoded by LINE-1: Structure and Function During L1 Retrotransposition', *Journal of Biomedicine and Biotechnology*, 2006, pp. 1–6. doi: 10.1155/JBB/2006/45621.

Martin, S. L. and Bushman, F. D. (2001) 'Nucleic acid chaperone activity of the ORF1 protein from the mouse LINE-1 retrotransposon.', *Molecular and cellular biology*, 21(2), pp. 467–75. doi: 10.1128/MCB.21.2.467-475.2001.

Mathias, S. *et al.* (1991) 'Reverse transcriptase encoded by a human transposable element', *Science*, 254(5039), pp. 1808–1810. doi: 10.1126/science.1722352.

Mathys, H. *et al.* (2014) 'Structural and Biochemical Insights to the Role of the CCR4-NOT Complex and DDX6 ATPase in MicroRNA Repression', *Molecular Cell*, 54(5), pp. 751–765. doi: 10.1016/j.molcel.2014.03.036.

Mayr, C. and Bartel, D. P. (2009) 'Widespread Shortening of 3'UTRs by Alternative Cleavage and Polyadenylation Activates Oncogenes in Cancer Cells', *Cell*, 138(4), pp. 673–684. doi: 10.1016/j.cell.2009.06.016.

McClintock, B. (1950) 'The Origin and Behavior of Mutable Loci in Maize', *Genetics*, 36(6), pp. 344–355. doi: 10.1073/pnas.36.6.344.

Medina, P. P., Nolde, M. and Slack, F. J. (2010) 'OncomiR addiction in an in vivo model of microRNA-21-induced pre-B-cell lymphoma', *Nature*, 467(7311), pp. 86–90. doi: 10.1038/nature09284.

Meijer, H. A. *et al.* (2013) 'Translational repression and eIF4A2 activity are critical for microRNA-mediated gene regulation', *Science*, 340(6128), pp. 82–85. doi: 10.1126/science.1231197.

Meister, G. *et al.* (2004) 'Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs', *Molecular Cell*, 15(2), pp. 185–197. doi: 10.1016/j.molcel.2004.07.007.

Meister, G. *et al.* (2005) 'Identification of novel argonaute-associated proteins', *Current Biology*, 15(23), pp. 2149–2155. doi: 10.1016/j.cub.2005.10.048.

Meister, G. (2013) 'Argonaute proteins: Functional insights and emerging roles', *Nature Reviews Genetics*, 14(7), pp. 447–459. doi: 10.1038/nrg3462.

Melo, S. A. *et al.* (2010) 'A genetic defect in exportin-5 traps precursor MicroRNAs in the nucleus of cancer cells', *Cancer Cell*, 18(4), pp. 303–315. doi: 10.1016/j.ccr.2010.09.007.

Merritt, W. M. *et al.* (2008) 'Dicer, Drosha, and outcomes in patients with ovarian cancer.', *The New England journal of medicine*, 359(25), pp. 2641–50. doi: 10.1056/NEJMoa0803785.

Mi, H. *et al.* (2019) 'PANTHER version 14: More genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools', *Nucleic Acids Research*. Oxford University Press, 47(D1), pp. D419–D426. doi: 10.1093/nar/gky1038.

Mi, S. *et al.* (2000) 'Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis.', *Nature*, 403(6771), pp. 785–9. doi: 10.1038/35001608.

Michlewski, G. and Cáceres, J. F. (2019) 'Post-transcriptional control of miRNA biogenesis', *RNA*, 25(1), pp. 1–16. doi: 10.1261/rna.068692.118.

Miki, Y. *et al.* (1992) 'Disruption of the APC Gene by a Retrotransposal Insertion of L1 Sequence in a Colon Cancer', *Cancer Research*, 52(3), pp. 643–645. doi: citeulike-article-id:13533271.

Mir, A. A., Philippe, C. and Cristofari, G. (2015) 'euL1db: The European database of L1HS retrotransposon insertions in humans', *Nucleic Acids Research*, 43(D1), pp. D43–D47. doi: 10.1093/nar/gku1043.

Miranda, K. C. *et al.* (2006) 'A Pattern-Based Method for the Identification of MicroRNA Binding Sites and Their Corresponding Heteroduplexes', *Cell*, 126(6), pp. 1203–1217. doi: 10.1016/j.cell.2006.07.031.

Mita, P. *et al.* (2018) 'LINE-1 protein localization and functional dynamics during the cell cycle', *eLife*, 7, pp. 1–35. doi: 10.7554/eLife.30058.

Mita, P. *et al.* (2020) 'BRCA1 and S phase DNA repair pathways restrict LINE-1 retrotransposition in human cells', *Nature Structural and Molecular Biology*, 27(2), pp. 179–191. doi: 10.1038/s41594-020-0374-z.

Mogilyansky, E. and Rigoutsos, I. (2013) 'The miR-17/92 cluster: A comprehensive update on its genomics, genetics, functions and increasingly important and numerous roles in health and disease', *Cell Death and Differentiation*, 20(12), pp. 1603–1614. doi: 10.1038/cdd.2013.125.

Moldovan, J. B. *et al.* (2019) 'RNA ligation precedes the retrotransposition of U6/LINE-1 chimeric RNA', *Proceedings of the National Academy of Sciences of the United States of America*, 116(41), pp. 20612–20622. doi: 10.1073/pnas.1805404116.

Moldovan, J. B. and Moran, J. V. (2015) 'The Zinc-Finger Antiviral Protein ZAP Inhibits LINE and Alu Retrotransposition', *PLoS Genetics*, 11(5), pp. 1–34. doi: 10.1371/journal.pgen.1005121.

Moore, M. J. *et al.* (2015) 'MiRNA-target chimeras reveal miRNA 3'-end pairing as a major determinant of Argonaute target specificity', *Nature Communications*, 6(May), pp. 1–17. doi: 10.1038/ncomms9864.

Moran, J., Holmes, S. and Naas, T. (1996) 'High frequency retrotransposition in cultured mammalian cells', *Cell*, 87, pp. 917–927. doi: 10.1016/S0092-8674(00)81998-4.

Moran, J. V., DeBerardinis, R. J. and Kazazian, H. H. (1999) 'Exon shuffling by L1 retrotransposition', *Science*, 283(5407), pp. 1530–1534. doi: 10.1126/science.283.5407.1530.

Morrish, T. A. *et al.* (2002) 'DNA repair mediated by endonuclease-independent LINE-1 retrotransposition', *Nature Genetics*, 31(2), pp. 159–165. doi: 10.1038/ng898.

Morrish, T. a *et al.* (2007) 'Endonuclease-independent LINE-1 retrotransposition at mammalian telomeres.', *Nature*, 446(7132), pp. 208–12. doi: 10.1038/nature05560.

Mouse Genome Sequencing Consortium *et al.* (2002) 'Initial sequencing and comparative analysis of the mouse genome.', *Nature*, 420(6915), pp. 520–62. doi: 10.1038/nature01262.

Muckenfuss, H. *et al.* (2006) 'APOBEC3 proteins inhibit human LINE-1 retrotransposition', *Journal of Biological Chemistry*, 281(31), pp. 22161–22172. doi: 10.1074/jbc.M601716200.

Muñoz-Lopez, M. *et al.* (2016) 'Study of Transposable Elements and Their Genomic Impact.', *Methods in molecular biology*, 1400, pp. 1–19. doi: 10.1007/978-1-4939-3372-3_1.

Muotri, A. R. *et al.* (2005) 'Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition', *Nature*, 435(7044), pp. 903–910. doi: 10.1038/nature03663.

Nair, V. S., Maeda, L. S. and Ioannidis, J. P. A. (2012) 'Clinical outcome prediction by MicroRNAs in human cancer: A systematic review', *Journal of the National Cancer Institute*, 104(7), pp. 528–540. doi: 10.1093/jnci/djs027.

Nguyen, T. A. *et al.* (2015) 'Functional Anatomy of the Human Microprocessor.', *Cell*, 161(6), pp. 1374–87. doi: 10.1016/j.cell.2015.05.010.

Nguyen, T. H. M. *et al.* (2018) 'L1 Retrotransposon Heterogeneity in Ovarian Tumor Cell Evolution', *Cell Reports*, 23(13), pp. 3730–3740. doi: 10.1016/j.celrep.2018.05.090.

Nicoloso, M. S. *et al.* (2009) 'MicroRNAs - The micro steering wheel of tumour metastases', *Nature Reviews Cancer*, 9(4), pp. 293–302. doi: 10.1038/nrc2619.

Nigumann, P. *et al.* (2002) 'Many human genes are transcribed from the antisense promoter of L1 retrotransposon', *Genomics*, 79(5), pp. 628–634. doi: 10.1006/geno.2002.6758.

Nussbacher, J. K. and Yeo, G. W. (2018) 'Systematic Discovery of RNA Binding Proteins that Regulate MicroRNA Levels', *Molecular Cell*, 69(6), pp. 1005-1016.e7. doi: 10.1016/j.molcel.2018.02.012.

O'Donnell, K. A. *et al.* (2005) 'c-Myc-regulated microRNAs modulate E2F1 expression', *Nature*, 435(7043), pp. 839–843. doi: 10.1038/nature03677.

Orecchini, E. *et al.* (2016) 'ADAR1 restricts LINE-1 retrotransposition.', *Nucleic acids research*, p. gkw834. doi: 10.1093/nar/gkw834.

Orgel, L. E. and Crick, F. H. C. (1980) 'Selfish DNA: The ultimate parasite', *Nature*, 284(5757), pp. 604–607. doi: 10.1038/284604a0.

Ostertag, E. M. *et al.* (2000) 'Determination of L1 retrotransposition kinetics in cultured cells.', *Nucleic acids research*, 28(6), pp. 1418–23. doi: 10.1093/nar/28.6.1418.

Ostertag, E. M. *et al.* (2002) 'A mouse model of human L1 retrotransposition', *Nature Genetics*, 32(4), pp. 655–660. doi: 10.1038/ng1022.

Ozsolak, F. *et al.* (2008) 'Chromatin structure analyses identify miRNA promoters', *Genes and Development*, 22(22), pp. 3172–3183. doi: 10.1101/gad.1706508.

Pace, J. K. and Feschotte, C. (2007) 'The evolutionary history of human DNA transposons: Evidence for intense activity in the primate lineage', *Genome Research*, 17(4), pp. 422–432. doi: 10.1101/gr.5826307.

Pasquinelli, a E. *et al.* (2000) 'Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA.', *Nature*, 408(6808), pp. 86–89. doi: 10.1038/35040556.

Paterson, A. L. *et al.* (2015) 'Mobile element insertions are frequent in oesophageal

adenocarcinomas and can mislead paired-end sequencing analysis', *BMC Genomics*. BMC Genomics, pp. 1–14. doi: 10.1186/s12864-015-1685-z.

Payer, L. M. *et al.* (2019) 'Alu insertion variants alter mRNA splicing', *Nucleic Acids Research*, 47(1), pp. 421–431. doi: 10.1093/nar/gky1086.

Peddigari, S. *et al.* (2013) 'HnRNPL and nucleolin bind LINE-1 RNA and function as host factors to modulate retrotransposition', *Nucleic Acids Research*, 41(1), pp. 575–585. doi: 10.1093/nar/gks1075.

Percharde, M. *et al.* (2018) 'A LINE1-Nucleolin Partnership Regulates Early Development and ESC Identity.', *Cell*, 174(2), pp. 391-405.e19. doi: 10.1016/j.cell.2018.05.043.

Perepelitsa-Belancio, V. and Deininger, P. (2003) 'RNA truncation by premature polyadenylation attenuates human mobile element activity', *Nature Genetics*, 35(4), pp. 363–366. doi: 10.1038/ng1269.

Petri, R. *et al.* (2019) 'LINE-2 transposable elements are a source of functional human microRNAs and target sites', *PLOS Genetics*, 15(3), p. e1008036. doi: 10.1371/journal.pgen.1008036.

Philippe, C. *et al.* (2016) 'Activation of individual L1 retrotransposon instances is restricted to cell-type dependent permissive loci', *eLife*, 5, pp. 1–30. doi: 10.7554/eLife.13926.

Pickeral, O. K. *et al.* (2000) 'Frequent human genomic DNA transduction driven by line-1 retrotransposition', *Genome Research*, 10(4), pp. 411–415. doi: 10.1101/gr.10.4.411.

Pillai, R. S. (2005) 'Inhibition of Translational Initiation by Let-7 MicroRNA in Human Cells', *Science*, 309(5740), pp. 1573–1576. doi: 10.1126/science.1115079.

Piriyapongsa, J., Mariño-Ramírez, L. and Jordan, I. K. (2007) 'Origin and evolution of human microRNAs from transposable elements', *Genetics*, 176(2), pp. 1323–1337. doi: 10.1534/genetics.107.072553.

Piskounova, E. *et al.* (2011) 'Lin28A and Lin28B inhibit let-7 MicroRNA biogenesis by distinct mechanisms', *Cell*, 147(5), pp. 1066–1079. doi: 10.1016/j.cell.2011.10.039.

Pitkänen, E. *et al.* (2014) 'Frequent L1 retrotranspositions originating from TTC28 in colorectal cancer.', *Oncotarget*, 5(3), pp. 853–9. doi: 10.18632/oncotarget.1781.

Pizarro, J. G. and Cristofari, G. (2016) 'Post-Transcriptional Control of LINE-1 Retrotransposition by Cellular Host Factors in Somatic Cells', *Frontiers in Cell and Developmental Biology*, 4(March). doi: 10.3389/fcell.2016.00014.

Pobezinsky, L. a *et al.* (2015) 'Let-7 microRNAs target the lineage-specific transcription factor PLZF to regulate terminal NKT cell differentiation and effector function', *Nature Immunology*, 16(April), pp. 8–11. doi: 10.1038/ni.3146.

Ponten, J. and Saksela, E. (1967) 'Two established in vitro cell lines from human mesenchymal tumours', *International Journal of Cancer*, 2(5), pp. 434–447. doi: 10.1002/ijc.2910020505.

Potter, S. S. *et al.* (1979) 'Transposition of elements of the 412, copia and 297 dispersed repeated gene families in drosophila', *Cell*, 17(2), pp. 415–427. doi: 10.1016/0092-8674(79)90168-5.

Powers, J. T. *et al.* (2016) 'Multiple mechanisms disrupt the let-7 microRNA family in neuroblastoma', *Nature*, 535(7611), pp. 246–251. doi: 10.1038/nature18632.

Rahkonen, N. *et al.* (2016) 'Mature Let-7 miRNAs fine tune expression of LIN28B in pluripotent human embryonic stem cells', *Stem Cell Research*, 17(3), pp. 498–503. doi: 10.1016/j.scr.2016.09.025.

Raiz, J. *et al.* (2012) 'The non-autonomous retrotransposon SVA is trans-mobilized by the human LINE-1 protein machinery', *Nucleic Acids Research*, 40(4), pp. 1666–1683. doi: 10.1093/nar/gkr863.

Ravà, M. *et al.* (2017) 'Mutual epithelium-macrophage dependency in liver carcinogenesis mediated by ST18.', *Hepatology*, 65(5), pp. 1708–1719. doi: 10.1002/hep.28942.

Ray *et al.* (2008) 'Multiple waves of recent DNA transposon activity in the bat, Myotis lucifugus', *Genome Research*, 18(5), pp. 717–728. doi: 10.1101/gr.071886.107.6.

Ray, D. A. *et al.* (2007) 'Bats with hATs: Evidence for recent DNA transposon activity in genus Myotis', *Molecular Biology and Evolution*, 24(3), pp. 632–639. doi: 10.1093/molbev/msl192.

Reczko, M. *et al.* (2012) 'Functional microRNA targets in protein coding sequences', *Bioinformatics*, 28(6), pp. 771–776. doi: 10.1093/bioinformatics/bts043.

Rehmsmeier, M. *et al.* (2004) 'Fast and effective prediction of microRNA/target duplexes.', *RNA*, 10(10), pp. 1507–17. doi: 10.1261/rna.5248604.

Reichholf, B. *et al.* (2019) 'Time-Resolved Small RNA Sequencing Unravels the Molecular Principles of MicroRNA Homeostasis', *Molecular Cell*, 75(4), pp. 756-768.e7. doi: 10.1016/j.molcel.2019.06.018.

Reinhart, B. J. *et al.* (2000) 'The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans.', *Nature*, 403(6772), pp. 901–906. doi: 10.1038/35002607.

Reznik, B. *et al.* (2019) 'Heterogeneity of transposon expression and activation of the repressive network in human fetal germ cells', *Development*, 146(12). doi: 10.1242/dev.171157.

Richardson, S. R. *et al.* (2014) 'APOBEC3A deaminates transiently exposed single-strand DNA during LINE-1 retrotransposition', *eLife*, 2014(3), pp. 1–20. doi: 10.7554/eLife.02008.

Richardson, S. R. *et al.* (2015) 'The Influence of LINE-1 and SINE Retrotransposons on Mammalian Genomes.', *Microbiology spectrum*, 3(2). doi: 10.1128/microbiolspec.MDNA3-0061-2014.

Richardson, S. R. *et al.* (2017) 'Heritable L1 retrotransposition in the mouse primordial germline and early embryo', *Genome Research*, 27(8), pp. 1395–1405. doi: 10.1101/gr.219022.116.

Robertson, B. *et al.* (2010) 'Specificity and functionality of microRNA inhibitors', *Silence*, 1(1), pp. 1–9. doi: 10.1186/1758-907X-1-10.

Rodić, N. *et al.* (2014) 'Long interspersed element-1 protein expression is a hallmark of many human cancers', *American Journal of Pathology*, 184(5), pp. 1280–1286. doi: 10.1016/j.ajpath.2014.01.007.

Rodić, N. *et al.* (2015) 'Retrotransposon insertions in the clonal evolution of pancreatic ductal adenocarcinoma', *Nature Medicine*, 21(9), pp. 1060–1064. doi: 10.1038/nm.3919.

Rodriguez-Martin, B. *et al.* (2020) 'Pan-cancer analysis of whole genomes identifies driver rearrangements promoted by LINE-1 retrotransposition', *Nature Genetics*, 52(3), pp. 306–319. doi: 10.1038/s41588-019-0562-0.

Roush, S. and Slack, F. J. (2008) 'The let-7 family of microRNAs', *Trends in Cell Biology*, 18(10), pp. 505–516. doi: 10.1016/j.tcb.2008.07.007.

Rupaimoole, R. and Slack, F. J. (2017) 'MicroRNA therapeutics: Towards a new era for the management of cancer and other diseases', *Nature Reviews Drug Discovery*, 16(3), pp. 203–221. doi: 10.1038/nrd.2016.246.

Sahakyan, A. B. *et al.* (2017) 'G-quadruplex structures within the 3' UTR of LINE-1 elements stimulate retrotransposition', *Nature Structural and Molecular Biology*, 24(3), pp. 243–247. doi: 10.1038/nsmb.3367.

Saito, K. *et al.* (2010) 'Long interspersed nuclear element 1 hypomethylation is a marker of poor prognosis in stage IA non-small cell lung cancer', *Clinical Cancer Research*, 16(8), pp. 2418–2426. doi: 10.1158/1078-0432.CCR-09-2819.

Sanchez-Luque, F. J. *et al.* (2019) 'LINE-1 Evasion of Epigenetic Repression in Humans.', *Molecular cell*, 75(3), pp. 590-604.e12. doi: 10.1016/j.molcel.2019.05.024.

Sandberg, R. *et al.* (2008) 'Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites.', *Science*, 320(5883), pp. 1643–7. doi: 10.1126/science.1155390.

Sassaman, D. M. *et al.* (1997) 'Many human L1 elements are capable of retrotransposition', *Nature Genetics*, 16(1), pp. 37–43. doi: 10.1038/ng0597-37.

Sayah, D. M. *et al.* (2004) 'Cyclophilin A retrotransposition into TRIM5 explains owl monkey resistance to HIV-1.', *Nature*, 430(6999), pp. 569–573. doi: 10.1038/nature02777.

Schauer, S. N. *et al.* (2018) 'L1 retrotransposition is a common feature of mammalian

hepatocarcinogenesis', *Genome Research*, pp. 1–15. doi: 10.1101/gr.226993.117.

Schirle, N. T., Sheu-Gruttadauria, J. and MacRae, I. J. (2014) 'Structural basis for microRNA targeting', *Science*, 346(6209), pp. 608–613. doi: 10.1126/science.1258040.

Schnable, P. S. *et al.* (2009) 'The B73 maize genome: complexity, diversity, and dynamics.', *Science*, 326(5956), pp. 1112–5. doi: 10.1126/science.1178534.

Schrader, L. and Schmitz, J. (2019) 'The impact of transposable elements in adaptive evolution', *Molecular Ecology*, 28(6), pp. 1537–1549. doi: 10.1111/mec.14794.

Schumann, G. G. *et al.* (2019) 'The impact of transposable element activity on therapeutically relevant human stem cells', *Mobile DNA*, 10(1), pp. 1–23. doi: 10.1186/s13100-019-0151-x.

Scott, E. C. *et al.* (2016) 'A hot L1 retrotransposon evades somatic repression and initiates human colorectal cancer', *Genome Research*, 26(6), pp. 745–755. doi: 10.1101/gr.201814.115.

Scott, E. C. and Devine, S. E. (2017) 'The role of somatic L1 retrotransposition in human cancers', *Viruses*, 9(6), p. 131. doi: 10.3390/v9060131.

Selbach, M. *et al.* (2008) 'Widespread changes in protein synthesis induced by microRNAs', *Nature*, 455(7209), pp. 58–63. doi: 10.1038/nature07228.

Sen, S. K. *et al.* (2007) 'Endonuclease-independent insertion provides an alternative pathway for L1 retrotransposition in the human genome', *Nucleic Acids Research*, 35(11), pp. 3741–3751. doi: 10.1093/nar/gkm317.

Servant, G. *et al.* (2017) 'The nucleotide excision repair pathway limits L1 retrotransposition', *Genetics*, 205(1), pp. 139–153. doi: 10.1534/genetics.116.188680.

Shenoy, A., Danial, M. and Blelloch, R. H. (2015) 'Let-7 and miR-125 cooperate to prime progenitors for astrogliogenesis.', *The EMBO Journal*, 34(9), pp. 1180–94. doi: 10.15252/embj.201489504.

Shore, S. *et al.* (2016) 'Small RNA library preparation method for next-generation sequencing using chemical modifications to prevent adapter dimer formation', *PLoS ONE*, 11(11), pp. 1–26. doi: 10.1371/journal.pone.0167009.

Shukla, R. *et al.* (2013) 'Endogenous retrotransposition activates oncogenic pathways in hepatocellular carcinoma', *Cell*, 153(1), pp. 101–111. doi: 10.1016/j.cell.2013.02.032.

Si, W. *et al.* (2018) 'A miR-20a/MAPK1/c-Myc regulatory feedback loop regulates breast carcinogenesis and chemoresistance', *Cell Death and Differentiation*, 25(2), pp. 406–420. doi: 10.1038/cdd.2017.176.

Siomi, M. C. *et al.* (2011) 'PIWI-interacting small RNAs: the vanguard of genome defence.', *Nature Reviews Molecular Cell Biology*, 12(4), pp. 246–258. doi: 10.1038/nrm3089.

Skowronski, J., Fanning, T. G. and Singer, M. F. (1988) 'Unit-length line-1 transcripts in human teratocarcinoma cells.', *Molecular and Cellular Biology*, 8(4), pp. 1385–1397. doi: 10.1128/mcb.8.4.1385.

Slack, F. J. and Chinnaiyan, A. M. (2019) 'The Role of Non-coding RNAs in Oncology', *Cell*, 179(5), pp. 1033–1055. doi: 10.1016/j.cell.2019.10.017.

Smit, A. F. A. and Riggs, A. D. (1995) 'MIRs are classic, tRNA-derived SINEs that amplified before the mammalian radiation', *Nucleic Acids Research*, 23(1), pp. 98–102. doi: 10.1093/nar/23.1.98.

Smith, A. M. *et al.* (2008) 'A novel mode of enhancer evolution: The Tal1 stem cell enhancer recruited a MIR element to specifically boost its activity', *Genome Research*, 18(9), pp. 1422–1432. doi: 10.1101/gr.077008.108.

Solyom, S. *et al.* (2012) 'Extensive somatic L1 retrotransposition in colorectal tumors', *Genome Research*, 22(12), pp. 2328–2338. doi: 10.1101/gr.145235.112.

Speek, M. (2001) 'Antisense Promoter of Human L1 Retrotransposon Drives Transcription of Adjacent Cellular Genes', *Molecular and Cellular Biology*, 21(6), pp. 1973–1985. doi: 10.1128/MCB.21.6.1973-1985.2001.

Spengler, R. M. *et al.* (2016) 'Elucidation of transcriptome-wide microRNA binding sites in human cardiac tissues by Ago2 HITS-CLIP', *Nucleic Acids Research*, 44(15), pp. 7120–7131. doi:

10.1093/nar/gkw640.

Startek, M. *et al.* (2015) 'Genome-wide analyses of LINE-LINE-mediated nonallelic homologous recombination', *Nucleic Acids Research*, 43(4), pp. 2188–2198. doi: 10.1093/nar/gku1394.

Stetson, D. B. *et al.* (2008) 'Trex1 Prevents Cell-Intrinsic Initiation of Autoimmunity', *Cell*, 134(4), pp. 587–598. doi: 10.1016/j.cell.2008.06.032.

Subtelny, A. O. *et al.* (2014) 'Poly(A)-tail profiling reveals an embryonic switch in translational control', *Nature*, 508(1), pp. 66–71. doi: 10.1038/nature13007.

Sugano, T., Kajikawa, M. and Okada, N. (2006) 'Isolation and characterization of retrotransposition-competent LINEs from zebrafish', *Gene*, 365(1-2 SPEC. ISS.), pp. 74–82. doi: 10.1016/j.gene.2005.09.037.

Sultana, T. *et al.* (2019) 'The Landscape of L1 Retrotransposons in the Human Genome Is Shaped by Pre-insertion Sequence Biases and Post-insertion Selection', *Molecular Cell*, 74(3), pp. 555-570.e7. doi: 10.1016/j.molcel.2019.02.036.

Swergold, G. D. (1990) 'Identification, characterization, and cell specificity of a human LINE-1 promoter.', *Molecular and cellular biology*, 10(12), pp. 6718–29. doi: 10.1128/mcb.10.12.6718.

Symer, D. E. *et al.* (2002) 'Human L1 retrotransposition is associated with genetic instability in vivo', *Cell*, 110(3), pp. 327–338. doi: 10.1016/S0092-8674(02)00839-5.

Szak, S. T. *et al.* (2002) 'Molecular archeology of L1 insertions in the human genome.', *Genome biology*, 3(10), p. research0052.1–0052.18. doi: 10.1186/gb-2002-3-10-research0052.

Takamizawa, J. *et al.* (2004) 'Reduced expression of the let-7 microRNAs in human lung cancers in association with shortened postoperative survival', *Cancer Research*, 64(11), pp. 3753–3756. doi: 10.1158/0008-5472.CAN-04-0637.

Tang, Z. *et al.* (2017) 'Human transposon insertion profiling: Analysis, visualization and identification of somatic LINE-1 insertions in ovarian cancer', *Proceedings of the National Academy of Sciences*, 114(5), pp. E733–E740. doi: 10.1073/pnas.1619797114.

Tarasov, V. *et al.* (2007) 'Differential regulation of microRNAs by p53 revealed by massively parallel sequencing: miR-34a is a p53 target that induces apoptosis and G 1-arrest', *Cell Cycle*, 6(13), pp. 1586–1593. doi: 10.4161/cc.6.13.4436.

Taylor, M. S. *et al.* (2013) 'Affinity proteomics reveals human host factors implicated in discrete stages of LINE-1 retrotransposition', *Cell*, 155(5), pp. 1034–1048. doi: 10.1016/j.cell.2013.10.021.

Taylor, M. S. *et al.* (2018) 'Dissection of affinity captured LINE-1 macromolecular complexes', *eLife*, 7, pp. 1–40. doi: 10.7554/eLife.30094.

Tchénio, T., Casella, J. F. and Heidmann, T. (2000) 'Members of the SRY family regulate the human LINE retrotransposons.', *Nucleic Acids Research*, 28(2), pp. 411–5. doi: 10.1093/nar/28.2.411.

Terry, D. M. and Devine, S. E. (2020) 'Aberrantly High Levels of Somatic LINE-1 Expression and Retrotransposition in Human Neurological Disorders', *Frontiers in Genetics*, 10(January), pp. 1–14. doi: 10.3389/fgene.2019.01244.

Thomas, C. A. *et al.* (2017) 'Modeling of TREX1-Dependent Autoimmune Disease using Human Stem Cells Highlights L1 Accumulation as a Source of Neuroinflammation.', *Cell Stem Cell*, 21(3), pp. 319-331.e8. doi: 10.1016/j.stem.2017.07.009.

Thomson, J. M. *et al.* (2006) 'Extensive post-transcriptional regulation of microRNAs and its implications for cancer', *Genes and Development*, 20(16), pp. 2202–2207. doi: 10.1101/gad.1444406.

Thornton, J. E. *et al.* (2014) 'Selective microRNA uridylation by Zcchc6 (TUT7) and Zcchc11 (TUT4)', *Nucleic Acids Research*, 42(18), pp. 11777–11791. doi: 10.1093/nar/gku805.

Tokumaru, S. *et al.* (2008) 'let-7 regulates Dicer expression and constitutes a negative feedback loop', *Carcinogenesis*, 29(11), pp. 2073–2077. doi: 10.1093/carcin/bgn187.

Treiber, T. *et al.* (2017) 'A Compendium of RNA-Binding Proteins that Regulate MicroRNA Biogenesis', *Molecular Cell*, 66(2), pp. 270-284.e13. doi: 10.1016/j.molcel.2017.03.014.

Treiber, T., Treiber, N. and Meister, G. (2019) 'Regulation of microRNA biogenesis and its crosstalk with other cellular pathways', *Nature Reviews Molecular Cell Biology*, 20(1), pp. 5–20. doi: 10.1038/s41580-018-0059-1.

Triboulet, R., Pirouz, M. and Gregory, R. I. (2015) 'A Single Let-7 MicroRNA Bypasses LIN28-Mediated Repression', *Cell Reports*, 13(2), pp. 260–266. doi: 10.1016/j.celrep.2015.08.086.

Tristan-Ramos, P. *et al.* (2020) 'sRNA/L1 retrotransposition: using siRNAs and miRNAs to expand the applications of the cell culture-based LINE-1 retrotransposition assay', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 375(1795), p. 20190346. doi: 10.1098/rstb.2019.0346.

Trompeter, H. I. *et al.* (2011) 'MicroRNAs MiR-17, MiR-20a, and MiR-106b Act in concert to modulate E2F activity on cell cycle arrest during neuronal lineage differentiation of USSC', *PLoS ONE*, 6(1). doi: 10.1371/journal.pone.0016138.

Tubio, J. M. C. *et al.* (2014) 'Extensive transduction of nonrepetitive DNA mediated by L1 retrotransposition in cancer genomes', *Science*, 345(6196), pp. 1251343–1251343. doi: 10.1126/science.1251343.

Ullu, E. and Tschudi, C. (1984) 'Alu sequences are processed 7SL RNA genes', *Nature*, 312(5990), pp. 171–172. doi: 10.1038/312171a0.

Upton, K. R. *et al.* (2015) 'Ubiquitous L1 mosaicism in hippocampal neurons', *Cell*, 161(2), pp. 228–239. doi: 10.1016/j.cell.2015.03.026.

Ustianenko, D. *et al.* (2018) 'LIN28 Selectively Modulates a Subclass of Let-7 MicroRNAs', *Molecular Cell*, 71(2), pp. 271-283.e5. doi: 10.1016/j.molcel.2018.06.029.

Vasudevan, S., Tong, Y. and Steitz, J. A. (2007) 'Switching from repression to activation: MicroRNAs can up-regulate translation', *Science*, 318(5858), pp. 1931–1934. doi: 10.1126/science.1149460.

Ventura, A. and Jacks, T. (2009) 'MicroRNAs and Cancer: Short RNAs Go a Long Way', *Cell*, 136(4), pp. 586–591. doi: 10.1016/j.cell.2009.02.005.

Viswanathan, S. R. and Daley, G. Q. (2010) 'Lin28: A MicroRNA Regulator with a Macro Role', *Cell*, pp. 445–449. doi: 10.1016/j.cell.2010.02.007.

Viswanathan, S. R., Daley, G. Q. and Gregory, R. I. (2008) 'Selective blockade of microRNA processing by Lin28', *Science*, 320(5872), pp. 97–100. doi: 10.1126/science.1154040.

Volinia, S. *et al.* (2006) 'A microRNA expression signature of human solid tumors defines cancer gene targets', *Proceedings of the National Academy of Sciences of the United States of America*, 103(7), pp. 2257–2261. doi: 10.1073/pnas.0510565103.

Wang, H. *et al.* (2005) 'SVA elements: A hominid-specific retroposon family', *Journal of Molecular Biology*, 354(4), pp. 994–1007. doi: 10.1016/j.jmb.2005.09.085.

Wang, J. *et al.* (2014) 'Primate-specific endogenous retrovirus-driven transcription defines naive-like stem cells', *Nature*, 516(7531), pp. 405–409. doi: 10.1038/nature13804.

Ward, J. R. *et al.* (2017) 'Condensin II and GAIT complexes cooperate to restrict LINE-1 retrotransposition in epithelial cells.', *PLoS genetics*, 13(10), p. e1007051. doi: 10.1371/journal.pgen.1007051.

Waring, M. and Britten, R. J. (1966) 'Nucleotide sequence repetition: A rapidly reassociating fraction of mouse DNA', *Science*, 154(3750), pp. 791–794. doi: 10.1126/science.154.3750.791.

Warkocki, Z. *et al.* (2018) 'Uridylation by TUT4/7 Restricts Retrotransposition of Human LINE-1s', *Cell*, pp. 1–12. doi: 10.1016/j.cell.2018.07.022.

Wei, W. *et al.* (2000) 'A transient assay reveals that cultured human cells can accommodate multiple LINE-1 retrotransposition events', *Analytical Biochemistry*, 284(2), pp. 435–438. doi: 10.1006/abio.2000.4675.

Wei, W. *et al.* (2001) 'Human L1 Retrotransposition: cis Preference versus trans Complementation', *Molecular and Cellular Biology*, 21(4), pp. 1429–1439. doi: 10.1128/MCB.21.4.1429-1439.2001.

Wightman, B., Ha, I. and Ruvkun, G. (1993) 'Posttranscriptional regulation of the

heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans', *Cell*, 75(5), pp. 855–862. doi: 10.1016/0092-8674(93)90530-4.

Wildschutte, J. H. *et al.* (2016) 'Discovery of unfixed endogenous retrovirus insertions in diverse human populations', *Proceedings of the National Academy of Sciences of the United States of America*, 113(16), pp. E2326–E2334. doi: 10.1073/pnas.1602336113.

Wilhelm, M. and Wilhelm, F. X. (2001) 'Reverse transcription of retroviruses and LTR retrotransposons.', *Cellular and Molecular Life Sciences*, 58(9), pp. 1246–62. doi: 10.1007/PL00000937.

Willems, L. and Gillet, N. A. (2015) 'APOBEC3 interference during replication of viral genomes', *Viruses*, 7(6), pp. 2999–3018. doi: 10.3390/v7062757.

Williams, Z. *et al.* (2015) 'Discovery and Characterization of piRNAs in the Human Fetal Ovary', *Cell Reports*, 13(4), pp. 854–863. doi: 10.1016/j.celrep.2015.09.030.

Wissing, S. *et al.* (2012) 'Reprogramming somatic cells into iPS cells activates LINE-1 retroelement mobility.', *Human Molecular Genetics*, 21(1), pp. 208–18. doi: 10.1093/hmg/ddr455.

Wu, L., Fan, J. and Belasco, J. G. (2006) 'MicroRNAs direct rapid deadenylation of mRNA', *Proceedings of the National Academy of Sciences of the United States of America*, 103(11), pp. 4034–4039. doi: 10.1073/pnas.0510928103.

Xie, Y. *et al.* (2011) 'Characterization of L1 retrotransposition with high-throughput dual-luciferase assays', *Nucleic Acids Research*, 39(3), pp. 1–11. doi: 10.1093/nar/gkq1076.

Xie, Y. *et al.* (2013) 'Cell division promotes efficient retrotransposition in a stable L1 reporter cell line.', *Mobile DNA*, 4(1), p. 10. doi: 10.1186/1759-8753-4-10.

Yanaihara, N. *et al.* (2006) 'Unique microRNA molecular profiles in lung cancer diagnosis and prognosis', *Cancer Cell*, 9(3), pp. 189–198. doi: 10.1016/j.ccr.2006.01.025.

Yang, N. *et al.* (2003) 'An important role for RUNX3 in human L1 transcription and retrotransposition', *Nucleic Acids Research*, 31(16), pp. 4929–4940. doi: 10.1093/nar/gkg663.

Yang, W. R. *et al.* (2019) 'SQuIRE reveals locus-specific regulation of interspersed repeat expression', *Nucleic Acids Research*. Oxford University Press, 47(5), pp. 1–16. doi: 10.1093/nar/gky1301.

Yi, R. *et al.* (2003) 'Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs', *Genes and Development*, 17(24), pp. 3011–3016. doi: 10.1101/gad.1158803.

Yoder, J. A., Walsh, C. P. and Bestor, T. H. (1997) 'Cytposine Methylation and the Ecology of Intragenomic Parasites', *Trends in Genetics*, 13(8), pp. 335–340. doi: 10.1016/S0168-9525(97)01181-5.

Yu, F. *et al.* (2007) 'let-7 Regulates Self Renewal and Tumorigenicity of Breast Cancer Cells', *Cell*, 131(6), pp. 1109–1123. doi: 10.1016/j.cell.2007.10.054.

Zeuthen, J. *et al.* (1980) 'Characterization of a human ovarian teratocarcinoma-derived cell line.', *International journal of cancer*. United States, 25(1), pp. 19–32.

Zhang, A. *et al.* (2014) 'RNase L restricts the mobility of engineered retrotransposons in cultured human cells', *Nucleic Acids Research*, 42(6), pp. 3803–3820. doi: 10.1093/nar/gkt1308.

Zhang, K. *et al.* (2018) 'A novel class of microRNA-recognition elements that function only within open reading frames', *Nature Structural & Molecular Biology*, 25(11), pp. 1019–1027. doi: 10.1038/s41594-018-0136-3.

Zhang, L. *et al.* (2006) 'microRNAs exhibit high frequency genomic alterations in human cancer.', *Proceedings of the National Academy of Sciences of the United States of America*, 103(24), pp. 9136–41. doi: 10.1073/pnas.0508889103.

Zhang, Z. *et al.* (2003) 'Millions of years of evolution preserved: A comprehensive catalog of the processed pseudogenes in the human genome', *Genome Research*, 13(12), pp. 2541–2558. doi: 10.1101/gr.1429003.

Zhao, C. *et al.* (2010) 'MicroRNA let-7b regulates neural stem cell proliferation and differentiation by targeting nuclear receptor TLX signaling.', *Proceedings of the National Academy of Sciences of the United States of America*, 107(5), pp. 1876–1881. doi: 10.1073/pnas.0908750107.

Zhao, K. *et al.* (2013) 'Modulation of LINE-1 and Alu/SVA Retrotransposition by Aicardi-Goutières Syndrome-Related SAMHD1', *Cell Reports*, 4(6), pp. 1108–1115. doi: 10.1016/j.celrep.2013.08.019.

Zhao, S. *et al.* (2015) 'MiR-20a promotes cervical cancer proliferation and metastasis in vitro and in vivo', *PLoS ONE*, 10(3), pp. 1–11. doi: 10.1371/journal.pone.0120905.

Zhu, G. fang *et al.* (2019) 'Mir20a/106a-WTX axis regulates RhoGDIa/CDC42 signaling and colon cancer progression', *Nature Communications*, 10(1). doi: 10.1038/s41467-018-07998-x.