PFIZER-UNIVERSIDAD DE GRANADA-JUNTA DE ANDALUCIA
CENTRE FOR GENOMICS AND ONCOLOGICAL RESEARCH

Universidad de Granada

Doctoral Program in Pharmacy

Doctoral Thesis:

Epigenetic control of the mobility of a human retrotransposon

PhD Candidate:
María Benítez Guijarro
FPU Fellowship

Thesis supervisors:
Dr. José Luis García Pérez
Dr. Antonio Sánchez Pozo

International PhD Mention.
GENYO - Centro Pfizer-Universidad de Granada-Junta de Andalucía de Genómica e Investigación Oncológica, Granada. España.
Laboratory of Cellular and Structural Biology, The Rockefeller University, Nueva York. EEUU (9 Months stay).

Granada, 2021.

Editor: Universidad de Granada. Tesis Doctorales
Autor: María Benítez Guijarro
ISBN: 978-84-1117-038-3
URI: http://hdl.handle.net/10481/70708

A mis padres y a mi hermano.

"A scientist in his laboratory is not a mere technician: he is also a child confronting natural phenomena that impress him as though they were fairy tales."

(Marie Curie, 1867-1934)

If at first you don't succeed, try two more times so that your failure is statistically significant.

AGRADECIMIENTOS / ACKNOWLEDGMENTS

No sé cómo agradecer a todos los que me habéis ayudado y acompañado hasta llegar a este momento, al final de mi Tesis. Han sido casi 6 años en los que además de aprender a investigar, he crecido como persona más que en ninguna otra etapa anterior de mi vida y todos los que habéis estado conmigo en este tiempo habéis formado parte de ese cambio. Solo puedo daros las **GRACIAS**.

Comenzaré por el principio.

Siempre me había atraído el mundo de la investigación, los laboratorios, la biología molecular, la genética… Sabía que quería ser doctora, pero no encontré mi sitio hasta que te conocí a ti, Cholo. Nunca olvidaré aquel día en que estando aún en el Máster nos presentaste el elemento móvil LINE-1. Simplemente me encantó y desde ese momento quise aprender más, y aprender de ti. Es un privilegio poder contar con un director de tesis con una mente como la tuya, y que además de un gran científico, me ha demostrado que también es una excelente persona. Por eso Cholo, solamente puedo agradecerte que me dieras la oportunidad de hacer mi tesis contigo y de aprender de ti. Muchas gracias por todos esos correos interminables, por tus consejos, por motivarme y compartir tu entusiasmo conmigo, por ayudarme en los momentos en que lo he necesitado, pero sobre todo, por confiar en mí. Te admiro como científico y como persona. Espero poder seguir colaborando contigo mucho más tiempo. **GRACIAS** por enseñarme a hacer ciencia, **MUCHÍSIMAS GRACIAS**.

A mis compañeros del Laboratorio 8, podría escribir páginas y páginas recordando momentos con vosotros. Mi querida Lauri, qué habría sido de mi sin ti. Has estado conmigo desde el principio y no solo me has ayudado y enseñado lo que sabes siempre que lo he necesitado, también has sido mi compañera, mi amiga, mi confidente. Me has acompañado de principio a fin y nadie sabe mejor que tú cómo han sido estos años para mí. No sabes cuánto he disfrutado trabajando contigo. Eres una bellísima persona y mejor trabajadora imposible. ¡¡Siempre podrás contar conmigo, te quiero, **MIL GRACIAS**!! A Sara, eres sencillamente admirable, siempre estás cuidando de todos y superándote a ti misma. **GRACIAS** por todas las veces que me has ayudado, por tus consejos y por ser un ejemplo a seguir. Vales muchísimo y no tengo dudas de que vas a llegar a donde tú quieras. **MUCHAS GRACIAS** por todo. A todos los que seguís en el labo: a Meri, eres un amor y te mereces todo por lo que has luchado y más y sé que tendrás tu recompensa, **GRACIAS** por todas las veces que me has ayudado y aconsejado, que sepas que siempre podrás contar conmigo; a mi Ana de los peces, **GRACIAS** por todas esas conversaciones sobre ciencia y sobre la vida y por tus consejos, eres capaz de todo lo que te propongas; a Martita, **GRACIAS** por tu ayuda y tus consejos, tanto experimentales y personales como de "Curly Girl", eres una mujer fuerte y luchadora y siempre aprendo algo nuevo de ti; a Alex, **GRACIAS** por tu ayuda, por las conversaciones, las paellas… eres un crack y te va a ir genial !; a mis queridas Anas, sois un amor de compañeras, alegráis el labo con vuestra presencia, disfrutad mucho del camino que os queda durante la Tesis, os va a ir genial

a las dos y ya sabéis, podéis contar conmigo para todo (PD. me debéis ir juntas a un concierto 😉), **GRACIAS**; a Laura Vera, no sabes cuánto te aprecio, cuánto me gustan nuestras reuniones, ni lo que he disfrutado trabajando juntas para conseguir que funcionen las CRISPR, me encanta cómo trabajas y tu entusiasmo inquebrantable, **GRACIAS** por tus consejos y por estar ahí siempre; a Cintia, **GRACIAS** por estar siempre dispuesta a ayudar, por tus consejos, ¡¡qué suerte tiene Laura de tenerte con ella!!; al recién regresado y versátil Thomas, muchas **GRACIAS** por tu ayuda, por todas esas conversaciones sobre ciencia, sobre el mundo científico y la vida en sí, por todos los consejos y ratos juntos en el laboratorio, mucho ánimo con esos peces; a Guillermo, quién fuera como tú, químico, informático, bioinformático, divulgador, profesor… eso es saber aprovechar el tiempo! ¡**GRACIAS** por todo! A todos con los que he coincidido durante este tiempo y ya se han marchado; Pablo, **GRACIAS** por tu ayuda, consejos, y por compartir toda tu sabiduría, me has enseñado mucho; a Kiko, ojalá ser tan buena científica cómo tú algún día, es una suerte poder escucharte, **GRACIAS** por todas las veces que me has ayudado; a Esther, eres un amor y no sabes cuánto me acuerdo de ti, vaya tía, puedes con lo que te echen!, **GRACIAS** por ayudarme siempre y por ser cómo eres; a mi Raquelilla, fuiste mi salvavidas cuando llegué al labo, **MIL GRACIAS** por todo lo que me ayudaste y por todas esas conversaciones; a David, **GRACIAS** por tu ayuda, eres un ejemplo a seguir como científico y como persona y es un gustazo hablar contigo; **GRACIAS** a Jenni, tan solo de pensar en ti alegras nuestros corazones, a Guille, eres un cielo de compañero y de persona y a Martín, **GRACIAS**. Me considero muy afortunada por haber coincidido con vosotros durante estos años, **MUCHÍSIMAS GRACIAS** por ayudarme a ser mejor científica, mejor compañera y mejor persona. Nunca olvidaré el tiempo que pasé con vosotros y todo lo que me enseñasteis.

MUCHAS GRACIAS a todos mis compañeros de la sala, a María, a Paulina, a Juane, a Araceli, a Marina, a Noelia, a Lidia, a Tere, a Marisa… y a todos aquellos otros con los que he tenido la suerte de compartir estos años. A todos los compañeros que habéis hecho que mi estancia en GENYO sea un placer, creando el ambiente de trabajo y compañerismo tan agradable que tenemos, y no empiezo porque terminaría escribiendo casi todo el Phonebook…!!. A Miguel Ángel, a Fernando, a Jorge, a Esmeralda, a Morante, a Raquel, a Olivia, Ana, a Carmen, a Lucía… por hacer tan bien vuestro trabajo y estar dispuestos a ayudarnos siempre.

A mi querido codirector Antonio Sánchez Pozo, por orientarme, por tus consejos, por tu apoyo y tu ayuda. Porque cada vez que hablo contigo me enseñas algo nuevo, es una suerte haber podido aprender de ti durante toda mi carrera, en el Grado, en el Máster, y ahora en el Doctorado, poder escucharte y contar contigo. **MUCHÍSIMAS GRACIAS.**

A todos los profesores del Departamento de Bioquímica y Biología Molecular II, **MUCHAS GRACIAS** por enseñarme a enseñar, por permitirme aprender de vosotros, y por contar conmigo. Rafa, Loles, Paloma, Alberto, Mª del Mar, Mercedes, Marina, David, Sergio, Esther, Jose Manuel, Mª Carmen…

I would also like to thank John LaCava and Michael P. Rout for welcoming me in their team and laboratory. Thank you for giving me the opportunity of learning from you, for including me and making me feel part of the group from the very first moment. Because since I arrived I felt at home. **THANK YOU** for everything you have taught me John, for trusting me, for motivating me. You have helped me to become the scientist I am today. I admire you and I love working with you. Thanks to Mehrnoosh, Hua, Leila, Greg and all the other members of the Rout Lab, Javi, Peter, Ryo, Natalia, Sam... You are all beautiful people and I am very lucky to have gotten to know you.

Finalmente, **GRACIAS** a todas las maravillosas personas que tengo la suerte de tener en mi vida y que me habéis apoyado y habéis creído en mi desde el principio. A mis amigos de toda la vida, a mi familia, a mis primos y a mis tíos.

A mis padres y a mi hermano. Porque sin vosotros no sería quién soy ni habría llegado hasta aquí. Porque cada vez que me he caído me habéis ayudado a levantarme. Porque os debo tanto que no tengo palabras para agradecéroslo sin quedarme corta. Porque habéis vivido conmigo todos y cada uno de los momentos por los que he pasado hasta llegar a escribir la última palabra de esta Tesis, apoyándome, creyendo en mí y soportándome en los buenos y en los malos momentos.
MIL GRACIAS POR ESTAR SIEMPRE CONMIGO.

MIL GRACIAS A TODOS,

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

- Internship in a foreign research center.

May 2018 – July 2018 and May 2019 – October 2019 (total of 9 months). Research stay at the Laboratory of Cellular and Structural Biology, The Rockefeller University (New York, EEUU) supervised by Dr. John LaCava, Research Associate Professor at Dr. Michaels P. Rout Laboratory. These stays were founded by two Mobility grants for short term stays and temporary transfers for beneficiaries of the "Formación de Profesorado Universitario" (FPU) program (2017 and 2018) and an EMBO short-term Fellowship (European Molecular Biology Organization).

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

RNase H2, mutated in Aicardi-Goutières syndrome, promotes LINE-1 retrotransposition. **Benitez-Guijarro M**, Lopez-Ruiz C, Tarnauskaitė Ž, Murina O, Mian Mohammad M, Williams TC, Fluteau A, Sanchez L, Vilar-Astasio R, Garcia-Canadas M, Cano D, Kempen MJ HC, Sanchez-Pozo A, Heras SR, Jackson AP, Reijns M AM, Garcia-Perez JL. EMBO J. 2018;37(15):e98506. doi: 10.15252/embj.201798506

- Language used for Thesis writing and defense

In order to fulfil the requirements of the University of Granada to obtain an International degree PhD this thesis has been written and will be defended in English. The abstract and conclusion section have been written and will be defended in Spanish.

Thesis founded by a "Formación de Profesorado Universitario" (FPU) fellowship (FPU15/03294).

INDEX

I. RESUMEN 1

II. ABSTRACT 7

III. INTRODUCTION 13

1. Mobility of DNA fragments in the human genome: Transposable Elements 15

1.1. The discovery of mobile elements 15

1.2. Types of mobile elements 17

1.2.1. DNA transposons 21

1.2.2. Long Terminal Repeats (LTR) containing retrotransposons or Endogenous Retro Viruses (ERVs) 24

1.2.3. Non-LTR retrotransposons 27

- Long INterspersed Elements (LINEs) 27

- Short INterspersed Elements (SINEs) 29

1.3. Impact of mobile elements in genomes 29

2. Active LINE-1 retrotransposons 32

2.1. Structure of retrotransposition-competent L1s (RC-L1s) 32

2.2. LINE-1 retrotransposition mechanism 34

2.2.1. A system to study retrotransposition: The retrotransposition assay 36

3. LINE-1 retrotransposition in pluripotent cells 38

3.1. Impact of L1 retrotransposition in genomes 40

4. Mechanisms regulating LINE-1 retrotransposition 43

4.1. Epigenetic control of L1 expression: regulation at the chromatin level .. 43

4.1.1. DNA methylation of the L1 promoter 43

4.1.2. Histone deacetylation as a possible L1 restriction mechanism. 45

4.2. Control mechanisms against the L1-mRNA: Regulation at the transcriptional and post-transcriptional levels. 47

5. Aicardi Goutières Syndrome (AGS) 51

6. The LINE-1 interactome 53

6.1. Proteomic study of the LINE-1 interactome 53

6.2. Genomic studies of the LINE-1 interactome 66

IV. HYPOTHESIS 69

Theoretical background supporting the hypothesis 71

Hypothesis 73

V. OBJECTIVES 75

Thesis Objectives 77

VI. MATERIALS AND METHODS 79

1. Materials 81

2. Cell line culture. 83

2.1. Cell thawing 84

2.2. Subculture routine 84

2.3. Cell freezing 85

3. Cell transfection 86

4. Gel electrophoresis, staining and western blotting. 87

5. Plasmids 89

5.1. Plasmid production 89

5.2. Characteristics and list of the plasmids employed 90

5.3. Plasmids employed 91

6. Retrotransposition assays. 95

7. Transposition assays. 97

8. Generation of KO cell lines and RNase H2 mutants. 98

9. Complementation of RNase H2 KO cell lines. 98

10. Cell extracts preparation for RNase H2 experiments. 99

11. RNase H2 activity assay. 99

12. Ribonucleotide detection in genomic DNA. 100

13. Mutation analysis by PCR and cloning. 101

14. Statistical analysis of the retrotransposition / transposition assays and activity assays. 101

15. Large-scale culture of pluripotent PA-1 (PCs) and differentiated PA-1 (DCs). 102

16. Cell collection and freezing to prepare cell beads. 103

17. Cryomilling of frozen cells. 104

18. Antibody coupled magnetic beads preparation 105

19. Affinity capture of protein complexes from cell powder 105

20. Generation of inducible Tet-On PA-1 cell lines 107

21. Sample preparation and liquid Chromatography coupled to Mass spectrometry 108

22. Mass spectrometry data analysis 108

23. LINE-1 Expression analyses by RT-qPCR 110

24. DNA-methylation analyses of the LINE-1 promoter 111

25. Microarray and GO expression analyses 112

VII. RESULTS 115

1. RNase H2 and LINE-1 retrotransposition 117

1.1. L1 retrotransposition is compromised in RNase H2 knock out (KO) mutant cell lines. 117

1.2. RNase H2 facilitates the mobilization of non-LTR retroelements but is not required for LTR retrotransposons and DNA transposons. 127

1.3. Overexpression of RNase H2 increase L1 retrotransposition. 132

1.4. Complementation of RNase H2 KO cell lines with the RNASEH2A subunit, but not with a "separation of function" mutant version, recovers L1 retrotransposition. 136

1.5. RNase H2 KO cells do not exhibit an increased mutation rate on retrotranscribed L1 DNAs. 143

1.6. Overexpression of a separation-of-function (SoF) mutant RNASEH2A rescues L1 retrotransposition. 146

1.7. Overexpression of RNase H1 in RNase H2 KO cells partially rescues L1 retrotransposition. 147

1.8. RNase H2 may interact with LINE-1 through PCNA. 152

1.9. RNASEH2A mutations characterized in AGS patients also reduce L1 retrotransposition. 154

2. Study of the LINE-1 interactome in pluripotent (PCs) and differentiated (DCs) PA-1 cells. 158

2.1. Capture of ectopic L1-RNPs in PCs and DCs. 159

2.1.1. Generation of inducible TetOn Advanced (Clontech) PA-1 cell lines to overexpress LINE-1 proteins. 164

2.2. Capture of endogenous L1-RNPs using PC and DC PA-1 cells. 169

2.2.1. Optimization of endogenous L1-ORF1p affinity capture. 170

2.3. Capturing L1-RNPs and their interactors on PC and DC PA-1 cells. 176

2.3.1. Capture of L1-ORF1p and interactors from PC PA-1 using affinity capture and mouse monoclonal anti-human_L1-ORF1p antibody coupled magnetic dynabeads (mαORF1p). 177

2.3.2. Capture of L1-ORF1p and interactors from PC PA-1 using affinity capture and llama-derived anti-human_L1-ORF1p antibody coupled magnetic dynabeads (LαORF1p). 182

2.3.3. Capture of L1-ORF1p and interactors from DC PA-1 using affinity capture and mouse monoclonal anti-human_L1-ORF1p antibody coupled magnetic dynabeads (mαORF1p). 188

2.4. Comparison of L1-ORF1p interactomes from isogenic PCs and DCs. 198

2.5. A list of PC-specific L1-ORF1p interactors potentially associated with L1-regulation in PCs. 204

VIII. DISCUSSION 219

1. Section I: RNase H2 role in LINE-1 retrotransposition 221

2. Section II: Study of the LINE-1 interactome 227

IX. CONCLUSIONS 241

X. CONCLUSIONES 247

BIBLIOGRAPHY 251

I. RESUMEN

Más de la mitad del genoma humano está formado por secuencias derivadas de elementos transponibles (TEs, por *Transposable Elements*), y algunos TEs continúan impactando nuestro genoma. Los retroelementos de tipo LINE-1 o L1 (*Long Interspersed Element class 1 elements*) son los únicos TEs activos y autónomos que se movilizan actualmente en el genoma humano utilizando un mecanismo de copia-pega. A lo largo de la evolución, los elementos LINE-1 se han amplificado hasta alcanzar cifras asombrosas en el genoma humano, de forma que >17% de nuestro genoma está formado por secuencias LINE-1s (>500000 copias de L1 por genoma). A pesar de su abundancia, sólo un pequeño subgrupo de 80-100 elementos sigue generando variabilidad interindividual en humanos, y estos elementos se conocen como *Retrotransposition Competent L1s* (RC-L1s). Los RC-L1s humanos tienen una longitud de 6 kb, y de 5´a 3´contienen: una región no traducida (5´UTR, de ~900pb) con actividad promotora interna, dos marcos abiertos de lectura (o ORFs, por *Open Reading Frames,* denominados L1-ORF1p y L1-ORF2p), y una corta 3´UTR (250pb) que contiene una señal de poliadenilación. Los RC-L1s se expresan y movilizan durante la embriogénesis humana temprana y, en una proporción mucho menor, en células germinales. De hecho, la actividad de L1 en el genoma de la línea germinal asegura su éxito evolutivo a lo largo del tiempo. Curiosamente, los RC-L1s también se expresan y movilizan en neuronas y células de cáncer; aunque la actividad de L1 en la línea somática no tiene un beneficio evolutivo claro. Aun así, algunas hipótesis sugieren que la actividad de L1 en células somáticas podría influenciar la biología humana. Debido a su carácter móvil, las inserciones de LINE-1 pueden tener un notable impacto en el genoma a través de múltiples mecanismos; de esta forma, y aunque esporádicamente, la actividad de L1 durante la embriogénesis temprana puede desencadenar nuevos trastornos genéticos en neonatos. En consecuencia, numerosos mecanismos que regulan y restringen la movilización de L1 han surgido durante la evolución humana. Así, estudios previos han caracterizado varios mecanismos de regulación de L1, tanto transcripcionales como postranscripcionales. De hecho, en esta Tesis he estudiado cómo se regulan los retroelementos L1 humanos, ya que sabemos relativamente poco sobre los mecanismos y vías que regulan la expresión y retrotransposición de L1.

Se ha demostrado que varios factores relacionados con el metabolismo de los ácidos nucleicos, como TREX1, SAMHD1 y ADAR-1 regulan la retrotransposición de LINE-1. Sorprendentemente, mutaciones inactivantes en estos genes desencadenan una rara enfermedad humana, conocida como Síndrome de Aicardi Goutières (AGS). El AGS es un trastorno genético caracterizado por una fuerte respuesta inmunitaria a ácidos nucleicos endógenos. Por otro lado, las mutaciones más comunes en pacientes de AGS se producen en cualquiera de las tres subunidades de la enzima humana RNasa H2

(subunidad A, B o C). Antes de esta Tesis, el potencial papel regulador de la RNasa H2 sobre los retrotransposones L1 no se había estudiado, aunque, por analogía con otros genes asociados con AGS se había sugerido que podrían actuar inhibiendo la retrotransposición de L1. Por tanto, en esta Tesis he diseccionado el papel de la enzima humana RNasa H2 en la regulación de LINE-1. Aprovechando las nuevas estrategias de edición del genoma (CRISPR/Cas9), generé varias líneas celulares que carecen de RNasa H2; a continuación, utilicé estas células para determinar si la movilización de un panel de elementos móviles activos se ve afectada en las células RNasa H2 KnockOut (KO). Utilizando ensayos de movilización de TEs *in vitro*, así como métodos bioquímicos, he podido demostrar que la actividad RNasa H2 es paradójicamente necesaria para la retrotransposición de L1, aunque es prescindible para otros retrotransposones de mamíferos que codifican su propio dominio/actividad RNasa H (retrotransposones Mus D de tipo LTR) o para los transposones de ADN (que se movilizan utilizando un mecanismo diferente de corte-pega). Además, demostré que mutaciones inactivantes en la subunidad RNASEH2A caracterizadas en pacientes AGS impedirían la retrotransposición de L1, lo que sugiere que no todos los pacientes AGS se caracterizarían por una alta actividad de L1 en sus genomas.

En la segunda parte de mi Tesis también estudié procesos de regulación de L1, pero utilizando análisis proteómicos innovadores. En estudios previos de nuestro laboratorio, demostramos que las nuevas inserciones de L1 en Células Pluripotentes (CP) humanas se silencian epigenéticamente durante, o inmediatamente después de su inserción, de una forma independiente de su secuencia y con la participación de enzimas que inducen desacetilación de histonas. Además, se demostró que el silenciamiento de L1 se atenúa en Células Diferenciadas (CDs) isogénicas, indicando que el silenciamiento de L1 podría ser un nuevo mecanismo regulador de L1. Nuestra hipótesis de trabajo sugiere que este mecanismo específico de CPs operaría normalmente en células embrionarias pluripotentes, un nicho celular donde naturalmente se acumulan nuevas inserciones de L1 en humanos; de esta forma, el silenciamiento de L1 reduciría la carga global de retrotransposición acumulada en el genoma humano heredable a lo largo de la evolución. A pesar de conocer ciertos aspectos del proceso de silenciamiento de L1, no conocemos ni su mecanismo ni cómo las CPs podrían reconocer las nuevas inserciones de L1. Por ello, decidí utilizar métodos proteómicos para diseccionar el mecanismo de silenciamiento de L1 en CPs. Concretamente, caractericé el interactoma de elementos LINE-1s endógenos tanto en CPs como en CDs isogénicas, utilizando una línea celular de carcinoma embrionario como modelo pluripotente (células PA-1). Para lograr este

objetivo, procedí a capturar por afinidad los intermediarios de la retrotransposición de L1 endógenos (es decir, L1-*RiboNucleoprotein Particles* o L1-RNPs) que son expresados naturalmente en CPs y CDs. Posteriormente, analicé la composición del interactoma de L1-ORF1p en CPs y CDs mediante espectrometría de masas. Sorprendentemente, descubrimos que el interactoma de L1 es muy dinámico durante la diferenciación celular y encontramos un gran número de nuevos interactores de L1, tanto en CPs como en CDs. Al comparar los interactomas de células pluripotentes y diferenciadas, generé una lista de genes candidatos que podrían participar en el silenciamiento de L1 específico de CPs. En futuros experimentos, se analizará el papel de estos factores en el silenciamiento de L1, utilizando CPs y CDs y enfoques de pérdida y ganancia de función, respectivamente.

En resumen, los resultados incluidos en esta Tesis han aumentado nuestro conocimiento mecanístico de cómo se regula la retrotransposición de L1 en células pluripotentes, un nicho celular donde L1 se expresa y moviliza en humanos. Un conocimiento más profundo de los mecanismos de regulación de L1 revelaría, en última instancia, cómo la actividad de L1 puede influenciar la evolución del genoma humano, y permitiría definir su papel causal en trastornos humanos como el AGS.

II. ABSTRACT

More than half of the human genome is made of transposable element (TE) derived sequences, and active TEs continue to impact our genome. In humans, Long INterspersed Elements class 1 (LINE-1 or L1) retrotransposons are active and autonomous TEs that move in genomes using a copy-and-paste mechanism. Over evolution, LINE-1s have amplified to astonishing numbers in the human genome, comprising >17% of our genome (i.e., >500000 L1 copies per genome). Despite their abundance, only a small subset of 80-100 LINE-1s continue to generate interindividual variability in humans, and are known as Retrotransposition-Competent L1s (RC-L1s). Human RC-L1s are 6-kb in length and, from 5´to 3´, contain: a 5´UnTranslated Region (UTR, of ~900bp) with internal promoter activity, two non-overlapping Open Reading Frames (ORFs, termed L1-ORF1p and L1-ORF2p), and end in a short 3´UTR containing a polyadenylation signal. RC-L1s are expressed and mobilize during early human embryogenesis and, at much lower rate, in germ cells. Indeed, L1 activity in the germline genome ensure their evolutionary success over time. Interestingly, LINE-1s are expressed and can mobilize in brain and cancer cells. Although L1 activity in the soma might not have any impact on L1 evolution, provocative hypotheses suggest that L1 activity in somatic cells might play a role on human biology. New LINE-1 insertions can impact the genome by a myriad of mechanisms, and therefore their activity during early embryogenesis can, sporadically, result in new genetic disorders in newborns. Thus, cells have evolved numerous mechanisms to regulate and restrict L1 mobilization. Several L1 regulation mechanisms, acting at the transcriptional and/or post-transcriptional levels have been characterized in the past. In fact, in this Thesis I studied how L1 retroelements are regulated, as we know relatively little about mechanisms and pathways that regulate L1 expression and retrotransposition.

Several factors related to the metabolism of nucleic acids, such as TREX1, SAMHD1 and ADAR-1, have been shown to regulate LINE-1 retrotransposition. Remarkably, inactivating mutations in these genes have also been associated with the development of Aicardi Goutières Syndrome (AGS). AGS is a rare disorder characterized for a strong immune response to endogenous nucleic acids. However, inactivating mutations in any of the three subunits of RNASEH2 (A, B, or C) is the most common mutation in AGS patients. Prior to this Thesis, the potential regulatory role of human RNase H2 on L1 retrotransposons was unexplored, although, by analogy with other AGS genes, it was suggested that RNase H2 might inhibit L1 retrotransposition. Thus, in this Thesis I dissected the role of human RNase H2 on LINE-1 regulation. Taking advantage of novel genome editing strategies (CRISPR/Cas9), I generated several cell lines lacking RNase H2 activity; then, I used these cells to determine whether the mobilization of a panel

of active TEs is affected in KnockOut (KO) RNase H2 cells. Using engineered TE mobilization assays and biochemical methods, I demonstrated that RNase H2 is paradoxically required for human L1 retrotransposition, although it is dispensable for other mammalian retrotransposons that encode their own RNase H domain/activity (Mus D LTR-retrotransposons) or for DNA-Transposons (which mobilize using a fundamentally different cut-and-paste mechanism). I also demonstrated that RNASEH2A inactivating mutations characterized in AGS patients would prevent L1 retrotransposition, further suggesting that not all AGS patients would be characterized for high L1 activity in their genomes.

In the second part of my Thesis I also studied L1 regulation processes, using innovative and advanced proteomic analyses of L1-interactors. Previous research from our lab demonstrated that new L1 insertions in human Pluripotent Cells (PCs) are epigenetically silenced during/shortly after L1 retrotransposition, in a sequence independent manner and involving histone deacetylation of newly inserted L1s. Because further research demonstrated that L1-silencing is attenuated in isogenic Differentiated Cells (DCs), we propose that L1-silencing might be a novel L1 regulatory mechanism of PCs. Our working hypothesis suggests that this pluripotent specific mechanism might normally operate in pluripotent embryonic cells, a cellular niche where *de novo* L1 insertions accumulate in humans, reducing the overall load of retrotransposition accumulated in the heritable human genome over evolution. Despite this knowledge, we don't know how L1-silencing works at a mechanistic level, nor how *de novo* L1 insertions are recognized in PCs. Thus, I decided to use proteomic methods to dissect the mechanism of L1-silencing in PCs. Here, I characterized the interactome of endogenous LINE-1s using the embryonic carcinoma cell line PA-1 as a pluripotent model, comparing isogenic PCs and DCs. To do that, I affinity captured endogenous L1 retrotransposition intermediates (i.e., L1 RiboNucleoprotein Particles) naturally expressed in PCs and DCs. Subsequently, I analyzed the composition of the L1-ORF1p interactome using mass spectrometry. Remarkably, we found that the L1-interactome is highly dynamic during cellular differentiation, and we found that most are novel L1 interactors. Furthermore, by comparing the interactomes of pluripotent and differentiated cells, I generated a list of candidate genes that could participate in L1-silencing. Future research would analyze the involvement of these factors in L1-silencing, using PCs and DCs and lost and gain of function approaches, respectively.

In summary, the results included in this Thesis have augmented our mechanistic knowledge of how L1 retrotransposition is regulated in pluripotent human cells, a cellular niche where L1 is expressed and retrotranspose. A deeper

understanding of L1-regulatory mechanisms would ultimately reveal how L1s can drive human genome evolution and would allow to define their causative role in human disorders like AGS.

III. INTRODUCTION

1. Mobility of DNA fragments in the human genome: Transposable Elements

1.1. The discovery of mobile elements

Along the 20th century, genes were considered as stable units arranged in a linear and sequential fashion on chromosomes, like beads on a necklace (Morgan, 1922). However, in the late 1940s, Barbara McClintock challenged existing knowledge about genes by discovering a chromosomal locus capable of changing its position within chromosomes, demonstrating that some genome sequences could mobilize within genomes. McClintock, a geneticist at Cold Spring Harbor Laboratory (New York, US), first identified what today we know as mobile DNA elements (or Transposable Elements, TEs) while studying the pattern of coloration in corn. In fact, she discovered what we nowadays known as a *DNA Transposon*, a type of mobile DNA. Mobile DNA can be defined as stretches of DNA that have the capability to move within genomes, and are also known as "jumping genes".

McClintock discovered that, depending on where transposable elements were inserted within a chromosome, these mobile elements could reversibly alter the expression of nearby genes (McClintock, 1950). In the scientific community of the time, the initial reaction to McClintock´s discoveries was rejection. Decades of genetic data mapping had shown that genes were linearly arranged in fixed positions relative to each other, making it difficult for researchers to accept that genes could move within genomes. As a result, scientific colleagues at the time were unable to completely accept these findings or take them seriously, and used the term "junk DNA" to refer to mobile genetic elements; however, several decades on, the widespread nature and importance of mobile genetic elements began to be recognized by the scientific community.

By the mid-1960s, the processes of DNA to mRNA transcription and mRNA to protein translation were well established, known as the "Central Dogma". In this paradigm, genes were no longer abstract concepts but discrete molecular entities that could be manipulated in a test tube. In this context, shortly before Barbara McClintock retired in the late 1960s, mobile genetic elements were discovered in bacteriophages, viruses that infect bacteria (Taylor, 1963), bacteria, and finally, in Drosophila (Shapiro, 1969; Engels & Preston, 1981). These discoveries led the scientific community to gradually recognize that transposable elements were not unique to maize but were widespread among species.

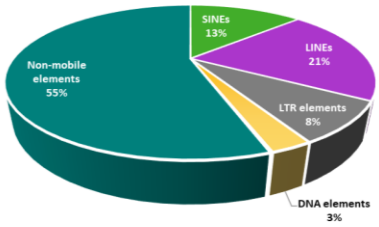
Remarkably, Barbara McClintock was recognized with a Nobel Prize in 1983, 35 years after the publication of her first seminal paper on mobile elements in corn (Ravindran, 2012). Since these initial events, research on transposable elements continued, leading to the discovery of multiple types of mobile elements, their noticeable impact on genomes, etc.

Mobile elements can be found in all living organisms, from bacteria to plants, animals and archaea (Singh *et al*, 2014). Their prevalence in genomes suggests that far from being junk DNA, as they were initially considered, they may exert some role beyond being drivers of genome evolution. Humans are not an exception, and our genome contains a large amount of transposable element-derived DNA sequences. Indeed, in the late 1960s, Britten and colleagues used DNA reassociation kinetic assays to discover that the human genome contains large amounts of repetitive DNA sequences (Britten & Kohne, 1968). Further studies revealed that, as other organisms, much of our DNA is made of transposable elements. However, through the completion of the first human genome reference sequence, it was confirmed that DNA was not the static, stable and immobile genetic material previously thought (Lander *et al*, 2001). The sequencing of the human genome revealed that, while only 5% or less of our genome corresponds to coding sequences, an astonishing 50% of our genome is made up of repetitive sequences. Further bioinformatic analyses have revealed that up to 70% of our genome might have been generated by the activity of mobile DNA elements (de Koning *et al*, 2011). In the human genome, repeated or repetitive DNA sequences include pseudogenes, simple sequence repeats, duplicated DNA segments of between 10 and 300 kb, blocks of tandem repeats (telomeres, centromeres, long arms of acrocentric chromosomes, ribosomal gene clusters) and mobile elements or transposable elements. Among all these types of repeat sequences, mobile elements constitute the most abundant, accounting for at least 45% (Lander *et al*, 2001). Most repeated sequences in genomes consist of "fixed" insertions present in all individuals of a given species. However, a small fraction of repeated DNA sequences are variable or polymorphic, and some of these correspond to active mobile DNA sequences which are able to mobilize in a given species (Mills *et al*, 2007).

1.2. Types of mobile elements

The sequencing of genomes has uncovered that there are multiple types of mobile DNA elements, and that their prevalence and activity is variable within species. Indeed, there is significant variability in the types of transposable elements found in genomes, even for closely related species, likely due to the inherent interaction with their host genome (Lander *et al*, 2001; Singh *et al*, 2014). Thus, at difference with evolutionary forces driving evolution of coding sequences, no generalization can be made for transposable elements, and each genome is a different entity. In this Thesis, I will describe the main types of transposable elements found in the human genome, and will refer to other genomes when relevant.

The human genome contains a significant percentage of repeated sequences and up to 70% of these sequences are made up of mobile elements (Lander *et al*, 2001; de Koning *et al*, 2011). Mobile elements or Transposable Elements (TEs) have driven genome evolution in a variety of ways. In most mammalian genomes, retrotransposons have accumulated to astonishing high numbers, eventually constituting a large fraction of the genome, shaping both genes and the entire genome. Although the host can control their number over evolution, massive expansions of retrotransposons have been tolerated. Therefore, mobile elements have become useful tools for learning more about genome evolution and gene function (Kazazian, 2004).



	Number of copies (x 1,000)	Total number of bases in the draft genome sequence (Mb)	Fraction of the draft genome sequence (%)
SINEs	1,558	359.6	13.14
Alu	1,090	290.1	10.60
MIR	393	60.1	2.20
MIR3	75	9.3	0.34
LINEs	868	558.8	20.42
LINE1	516	462.1	16.89
LINE2	315	88.2	3.22
LINE3	37	84	0.31
LTR elements	443	227.0	8.29
ERV-class I	112	79.2	2.89
ERV(K)-class II	8	8.5	0.31
ERV(L)-class III	83	39.5	1.44
MaLR	240	99.8	3.65
DNA elements	294	77.6	2.84
hAT group			
MER1-Charlie	182	38.1	1.39
Zaphod	13	4.3	0.16
Tc-1 group			
MER2-Tigger	57	28.0	1.02
Tc2	4	0.9	0.03
Mariner	14	2.6	0.10
PiggyBac-like	2	0.5	0.02
Unclassified	22	3.2	0.12
Unclassified	3	3.8	0.14
Total interspersed repeats		1,226.8	44.83

Table 1. Number of copies and fraction of the genome made of mobile elements. *Adapted from Lander et al, 2001.*

The study of these sequences has led to the identification of different types of mobile elements which can be classified as autonomous (i.e., *cis* mobilization) or non-autonomous (i.e., *trans* mobilization), depending on their coding capability (Kazazian, 2004; Singh *et al*, 2014). Additionally, TEs can be classified according to their mobilization mechanism (Lander *et al*, 2001), and the two main types of TEs are:

- **Class 1:** Some mobile elements that mobilize using a "**cut-and-paste**" mechanism and a **DNA intermediate,** which involves that these DNA sequences physically move within genomes, changing their location as a result. These elements are known as **DNA transposons**, and in general are highly abundant in prokaryotes (Beck *et al*, 2011; Lander *et al*, 2001; Bourque *et al*, 2018).

- Class 2: Mobile elements that mobilize using a "**copy and paste**" mechanism and an **intermediate RNA;** this mechanism allows not only mobilization of a DNA sequence, but also the generation of new copies within genomes. These mobile elements are known as **retrotransposons**, and they use a specialized enzyme known as **Reverse Transcriptase (RT)** to convert the intermediate RNA into DNA. There are several subclasses of retrotransposons according to whether they encode the enzymatic machinery required for their mobilization. On this basis, we distinguish between **autonomous** and **non-autonomous** retrotransposons. Moreover, retrotransposons can be subdivided into two main branches, depending on the presence (or absence) of Long Terminal Repeats (**LTRs**) in their structure

(Beck *et al,* 2011*;* Lander *et al,* 2001*;* Bourque *et al,* 2018*)*:

- LTR-type retrotransposons:
  - **Autonomous:** Endogenous Retroviruses (**ERVs**).
- Non-LTR type retrotransposons:
  - **Autonomous:** Long INterspersed Elements (**LINEs**).
  - **Non-autonomous:** Short INterspersed Elements (**SINEs**).

Below, I will describe the main TEs found in the human genome, with an emphasis on those relevant to this Thesis.

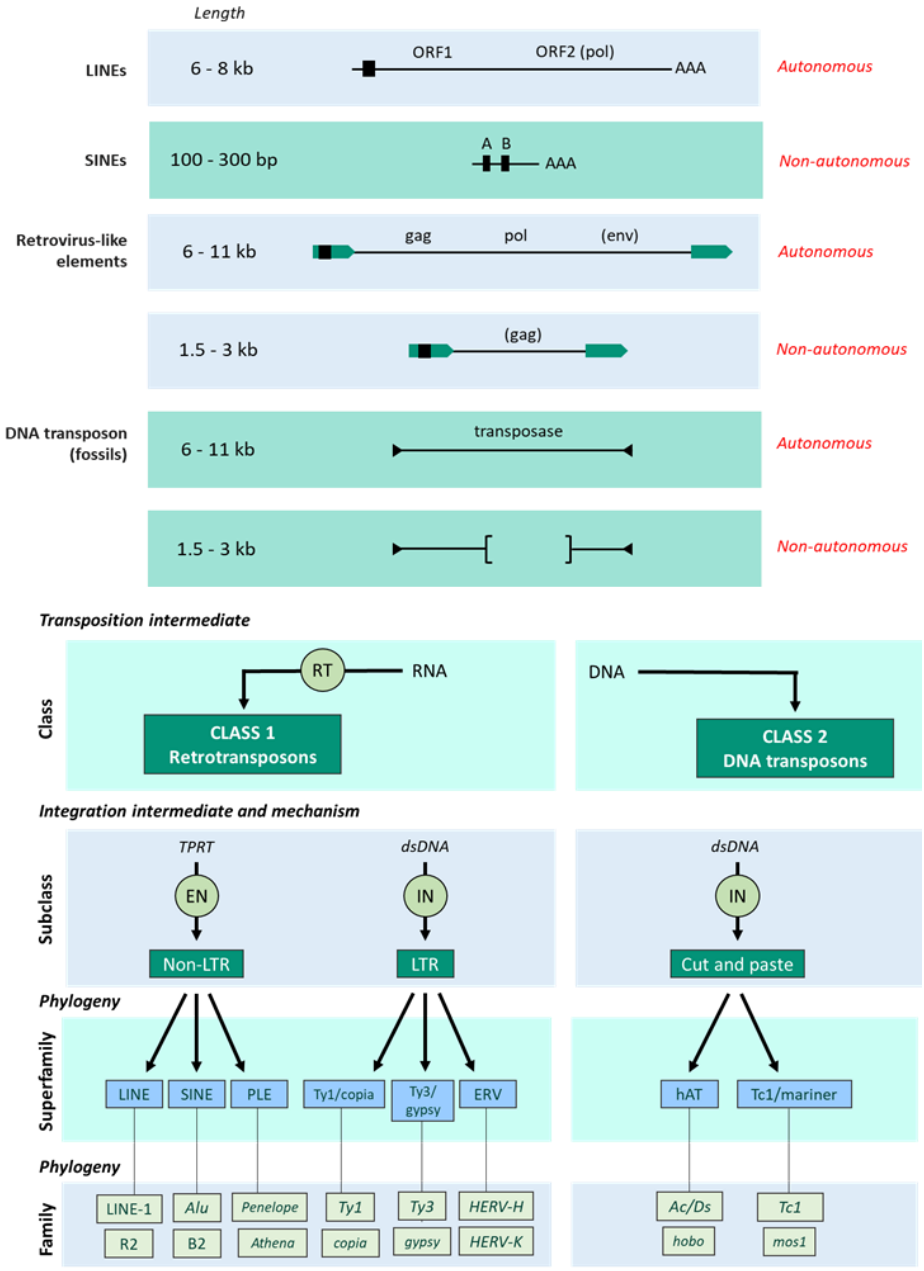


Figure 1. Classes of mobile elements in the human genome. *Shown are TEs relevant for this Thesis which are classified based on their mobilization mechanism. Adapted from Lander et al, (2001) and Bourque et al, (2018).*

1.2.1. DNA transposons

DNA transposons in the human genome are currently inactive; however, they constitute as much as 3% of the human genome (Lander *et al,* 2001*;* Craig, 2002). The human genome contains at least seven major classes of DNA transposons, which can be subdivided into many families with independent evolutionary origins (Smit, 1996). Human DNA transposons resemble bacterial transposons and are flanked by short Terminal Inverted Repeats (TIRs). "Autonomous" DNA-Transposons encode a specialized enzyme, known as Transposase, that bind both TIRs and mediates their mobilization through a "cut-and-paste" mechanism. In contrast, "Non-autonomous" DNA-Transposons, known as MITEs (for Miniature Inverted-repeat Transposable Elements (Jiang *et al,* 2004*)*), rely on the enzymatic machinery of autonomous DNA-Transposons for their mobilization. DNA transposons never use RNA intermediates in their mobilization mechanism, and are only capable of cleaving/inserting their genome at different genomic locations (Munoz-Lopez & Garcia-Perez, 2010).

In a typical round of DNA transposition, transcription of the mRNA occurs in the nucleus, followed by translation of the transposase in the cytoplasm. The transposase is then relocated to the nucleus, where it will recognize and bind the TIR sequences (normally the last 26 base pairs (bp) of the transposon sequence), cleaving the transposon DNA, and finally inserting it into a different genomic location. Due to the staggered cut of the transposase in the genome, the newly inserted DNA-Transposon is usually flanked by short Target Site Duplications (TSDs). Using this mobilization mechanism, there is no amplification in the copy number of DNA-Transposons; however, DNA-Transposons can increase their copy number using an alternative mechanism, known as replicative transposition (Derbyshire & Grindley, 1986).

To note, when the transposase access the nucleus, it cannot distinguish active from inactive DNA-Transposons elements, and can mobilize both (i.e., complete transposon sequences and sequences that are defective and inactive (Munoz-Lopez & Garcia-Perez, 2010)). As a result, inactive copies accumulate in the genome over evolution and the transposition process becomes less efficient with time. This occurs in all families of DNA transposons and, eventually, causes them to become extinct. To survive, DNA transposons must move by horizontal transfer into virgin genomes, and there is considerable evidence for such transfers over evolution (Lander *et al,* 2001*;* Simmons, 1992). Indeed, there is solid evidence indicating that a type of mammalian DNA transposon, known as Mariner, reached the human genome by horizontal transfer from insects (Smit, 1996).

While inactive in humans and most mammals, DNA transposons are widely distributed in nature and are often present and active in fungi, fish, plants, rotifers. Within mammals, several studies have demonstrated recent DNA transposon activity in some bat species, indicating that DNA transposons continue to generate genetic diversity in some mammals (Ray *et al*, 2008).

However, a DNA transposon was "re-built" using recombinant DNA technology, employing several copies of a DNA transposon that had been recently inserted into the salmon genome (Ivics *et al*, 1997). The transposase was resurrected from "fossil" sequences of transposases belonging to the Tc1/mariner class transposons present in the salmonid genome (Radice *et al*, 1994). In this way, it has been possible to generate a DNA transposon-like element active in mammalian cells, known as Sleeping Beauty (**SB**) transposon (Ivics *et al*, 1997). In this Thesis, I used the SB system as a control in some experiments (see below).

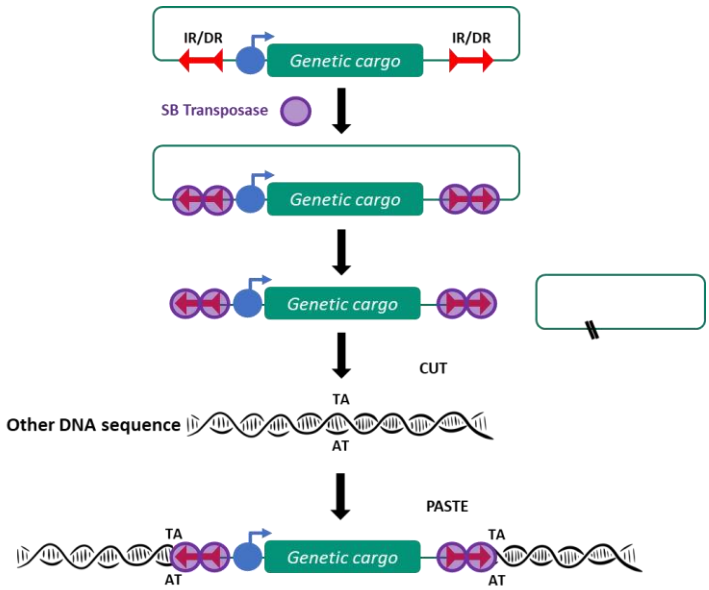


Figure 2. Mechanism of SB-mediated transposition*. The upper part of the image shows a transposon, defined by the presence of inverted IR/DR or TIR repeats, which would be contained within another DNA molecule (i.e., plasmid). The transposon would contain an expression cassette called "Genetic cargo", capable of transcribing itself if a promoter is present upstream. The Sleeping Beauty transposase (whose sequence would be encoded in another region of the genome or plasmid, not shown), would bind to the TIR regions, as shown in the image, and would excise the transposon. Then, the insertion of the transposon would be carried out in new genomic loci, on a TA sequence. In the process, the TA sequence of the insertion site is duplicated (i.e., TSDs).*

Due to its high mobilization rate in mammalian cells, the SB transposon has been used as an effective DNA editing tool (**Figure 2**) (Geurts *et al*, 2003). Thus, SB represents a synthetic transposon that has been designed to introduce sequences into vertebrate chromosomes. This system is composed of the Sleeping Beauty transposase and a sequence of interest that is inserted into the genome, that needs to be flanked by two TIRs that will be recognized by the transposase (**Figure 2**).

1.2.2. Long Terminal Repeats (LTR) containing retrotransposons or Endogenous Retro Viruses (ERVs)

Endogenous retroviruses (**ERVs**) or LTR-Retroelements are a type of retrotransposons that contain Long Terminal Repeat (**LTRs**) sequences at their flanking ends (Wilkinson *et al*, 1994; Mager & Stoye, *2015*; Medstrand & Mager, 1998).

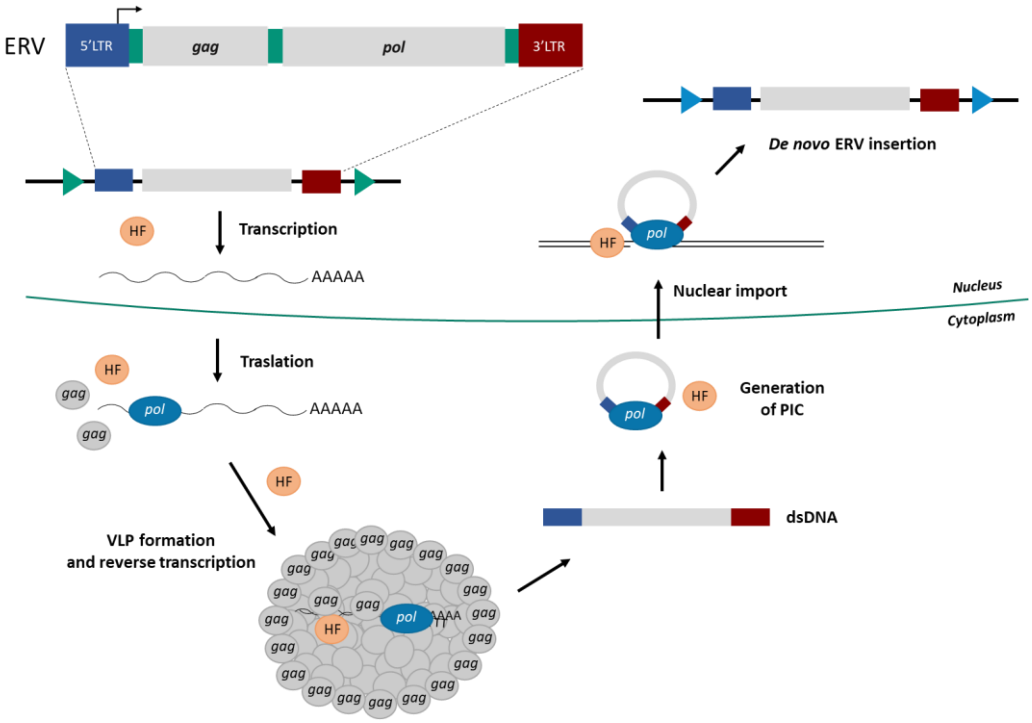
They represent the largest branch of elements that move through an RNA intermediate, and are present in plants, fungi and humans (Lander *et al*, 2001). Although multiple types of LTR retrotransposons have been characterized in nature, only vertebrate-specific ERVs seem to have been active in the mammalian genome. Indeed, ERVs constitute 8% and 10% of the human and mouse genomes, respectively, ranging from evolutionary ancient sequences (previous to the divergence of mammals), to evolutionary active young elements (Lander *et al*, 2001; Waterston *et al*, 2002). It has been demonstrated that ERVs can act as genetic agents for the development of disease and cancer (i.e., introducing a transduced oncogene or by insertional activation of a host proto-oncogene), as reported in mice (Howard *et al*, 2008)); in contrast, ERV-derived proteins have also been domesticated to perform functions in the host (i.e., in placental development, where the envelope gene (*env)* from ERVs is expressed and show fusogenic properties (Blaise *et al*, 2003)). Thus, and over the course of evolution, many ERVs have disappeared as a result of natural selection, while others have been able to survive by becoming part of the host genome and being harnessed to play beneficial roles.

LTR sequences, necessary for transcriptional regulation, are found in all ERVs, and ERVs can be further subdivided depending on their coding capability. Within ERVs, full-length autonomous elements code for **gag** and **pol** genes (Wilkinson *et al*, 1994). The **gag** gene encodes for a protein that can bind nucleic acids and is involved in the retrotransposition process of ERV RNAs (Garfinkel *et al*, 1985; Cristofari *et al*, 2000). The **pol** gene encodes for a protein with several conserved enzymatic domains/activities: Protease (PR), Reverse Transcriptase (RT), RNase H and INTegrase (INT) (Kim *et al*, 1998). Due to sequence/structure similarities, it has been proposed that infecting "exogenous retroviruses" might have arisen from ERVs, through the acquisition of a cell envelope **(env)** gene (Malik, 2000). Indeed, retroviruses can also be considered as LTR retrotransposons, and both use the same insertion mechanism.

Retrotransposition of LTR elements, as infection by "exogenous retroviruses", (**Figure 3**) requires the formation of a cytoplasmic Viral Like Particle (VLP) where

the intermediate RNA is reverse transcribed into double stranded DNA (dsDNA), also using a tRNA to prime first cDNA synthesis.

Figure 3. LTR-retrotransposon mobilization cycle. *A full-length LTR element is transcribed*



from its 5'LTR region in the nucleus. Transcribed RNAs are then transported to the cytoplasm, where gag and pol genes are translated and processed giving rise to mature proteins with PR, RT, RNase H and INT activity. The resulting proteins are packaged with ERV RNAs into virus-like particles (VLPs). ERV RNAs are reverse transcribed into dsDNA, which is then processed by the INT activity to generate the pre-integration complex (PIC) which is imported into the nucleus. Once in the nucleus, INT activity inserts the ERV dsDNA into a new genomic location, generating a de novo ERV insertion flanked by TSDs (arrowheads). During ERV mobilization, different host factors (HF) are involved and regulate their integration. Adapted from Garcia-Perez et al, 2016.

Mammalian retroviruses are divided into three main classes, each of them having many families with independent origins (Lander *et al*, 2001; Bourque *et al*, 2018). In genomes, the majority of ERVs (85%) are found as "fossil" solo LTRs, where the internal sequence of the ERV has been lost by homologous recombination (i.e., using flanking LTRs (Mager & Stoye, 2015; Garcia-Perez *et al*, 2016)).

In general, ERVs are not considered to be active in humans, at least in the heritable genome. However, the human genome contains a few presumably active

ERVs, found as rare variants of the HERV-K family. To note, they can have a complete ERV sequence or only the "LTR" (Wildschutte et al, 2016). Thus, these rare active HERV-K alleles present at a very low allelic frequency in some human populations (Lenz, 2016) could retrotranspose in selected humans and individuals, and could impact human biology/health.

At difference with humans, ERVs are active in other vertebrates, including mice. As a representative example, there are two active ERV families in the mouse genome, known as Intracisternal-A-Particle (**IAP)** and Endogenous type D murine (**MusD)** retrotransposons (Doolittle *et al,* 1989; Mager & Freeman, 2000). The MusD family comprises about 100 copies of 7.5-kb-long elements containing **gag** and **pol** genes (**Figure 4**), with similarities to beta retrovirus. The **gag** gene contains a structural protein that is processed by a protease (**pro**), also encoded by the retrotransposon in order to give rise to the capsid and nucleocapsid proteins during the maturation of the particle; the **pol** gene encodes the other enzymes necessary for reverse transcription/integration (i.e., RT, RNase H, and INT). In addition, the MusD genome also contains: i) a binding region for tRNA (PBS, for "primer binding site"), ii) a region rich in purines that cannot be degraded by the RNase H activity that encodes the retroelement but that is key for its mobility (PPT, for "poly purine tract") and iii) a potential retroviral assembly signal. To note, I used the mouse LTR element MusD in this Thesis as a control. Briefly, Dewannieux and colleagues (Ribet *et al,* 2004) cloned and sequenced active MusD elements from the mouse genome, and they further showed that they could retrotranspose in mammalian cells, both *in cis* and *in trans*.

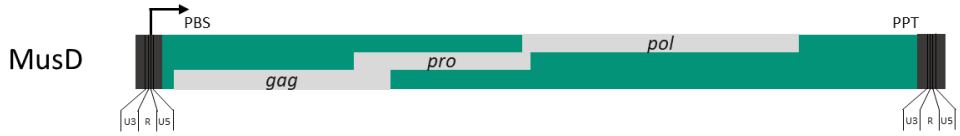


Figure 4. MusD structure. *Dark gray, LTRs with U3-R-U5 organization; PBS, primer binding site; PPT, poly purine tract. Adapted from Ribet et al, 2004.*

1.2.3. Non-LTR retrotransposons

Currently, non-LTR retrotransposons are the only mobile elements active in humans. As retrotransposons, they use an intermediate RNA and RT activity to mediate their mobility but lack LTR sequences and their mobilization mechanism is radically different from that used by ERVs. The human genome contains two main types of non-LTR elements: LINEs and SINEs[1] (Lander *et al*, 2001; Richardson *et al*, 2015).

- Long INterspersed Elements (LINEs)

LINEs are the TE class most abundant in the human genome, comprising 20% of our genome, and can be subdivided in class 3 LINEs (i.e., LINE-3s or CR1), class 2 LINEs (LINE-2s), and class 1 LINEs (LINE-1s). LINE-1 or L1 elements are autonomous retrotransposons that code the enzymatic machinery involved in their mobilization (reviewed in Macia *et al*, 2015; Richardson *et al*, 2015). LINE-1 elements comprise more than 17% of our genome and there are more than 600000 copies of this element per genome (Lander *et al*, 2001; Macia *et al*, 2015; Richardson *et al*, 2015). Despite their abundance, a reference human genome only contains between **80-100 active LINE-1s**, all belonging to the L1Hs subfamily (Brouha *et al*, 2003). Please note that a full description of active human LINE-1 elements will be discussed in detail in section 2.

Active L1Hs retrotransposons are approximately 6 kb in length, and from 5´to 3´, they contain:

i) a 900bp long **5´untranslated region (UTR)** with internal sense RNA polymerase II promoter. Notably, the same 5'UTR also contains an antisense RNA polymerase II promoter, giving rise to a protein known as **L1-ORF0** when antisense-derived L1 RNAs are translated.

ii) two non-overlapping open reading frames (ORFs), known as **L1-ORF1** and **L1-ORF2**, separated by a 63 bp long intergenic region.

iii) a 250 bp long **3´UTR** containing a polyadenylation signal and ending in a **polyA tail**.

During L1 retrotransposition, at difference with ERVs, reverse transcription of

[1] To note, there are many more characteristics and details of SINEs worth to be discussed here; however, I mainly discuss LINE-1 retrotransposons as they are the focus of this Thesis.

the intermediate L1 mRNA occurs in the nucleus (see below), and initiates at the 3´end polyA tail of the L1 mRNA. However, reverse transcription does not often occur completely, leading to 5´truncated L1 insertions; indeed, most L1 insertions are 5'truncated and non-functional. As a result, most LINE-derived repeats found in our genome are short, with an average size of 900 bp for all copies of LINE-1, and an average size of 1,070 bp for copies of the currently active LINE-1 element in humans (L1Hs). As other transposable elements, new L1 insertions are often flanked by short TSDs of 7 ± 20 bp, due to the mechanism of retrotransposition (see below).

Notably, LINE-1 is responsible for most of the ongoing reverse transcription in the genome, including the retrotransposition of non-autonomous SINEs and the creation of processed pseudogenes (Esnault *et al,* 2000*;* Wei *et al,* 2001*)*.

LINEs are also present in the genome of other organisms, including the model organism zebrafish *(Danio rerio*). However, at difference with humans, both LINE-2 and LINE-1s seem to be currently active in zebrafish (Howe *et al,* 2013*)*. Furthermore, there are 3 and 17 potentially active LINE-2 and LINE-1 subfamilies in the zebrafish genome, respectively, at difference with our genome, where only a single subfamily of LINE-1 is currently active (L1Hs elements). While zebrafish LINE-1s are most similar to human LINE-1s, no active LINE-1 copy has been identified so far in the zebrafish genome. However, pioneer work from the Okada lab identified potentially active LINE-2 copies from zebrafish, and some were able to retrotranspose in cultured human HeLa cells (Sugano *et al,* 2006*)* and are also suspected to be active in fish (Howe *et al,* 2013*)*.

One of the potentially active LINE-2 copies with activity in human cells are ZfL2-2 elements. A prototype ZfL2-2 element has a simpler structure, with a length of 4.2-Kb, and coding for a single ORF (Figure 5). Notably, the single ORF from ZfL2-2 elements has high homology with human L1-ORF2p (and with other mammalian LINE-1 elements) and contains reverse transcriptase (RT) and endonuclease (EN) domains, further highlining similarities in their retrotransposition mechanism. In this Thesis, I used zebrafish Zfl2-2 elements as controls.

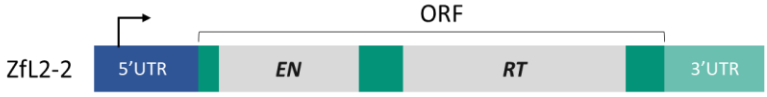


Figure 5. Zebrafish LINE-2 element (ZfL2-2)*. Adapted from Sugano et al, 2006.*

- Short INterspersed Elements (SINEs)

SINEs are active non-autonomous non-LTR retrotransposons; in fact, SINEs require the proteins encoded by active LINEs to mediate their mobilization *in trans* (Dewannieux *et al*, 2003; Hancks *et al*, 2011; Hancks & Kazazian, 2012). Nevertheless, they represent 13% of our genome, and have accumulated millions of copies in our genome (Lander *et al*, 2001; Macia *et al*, 2015; Richardson *et al*, 2015). As LINE-1s, SINEs are currently active in the human genome, and there are two types of SINEs: Alu and composite SINE-VNTR-Alus (SVA) elements.

Alu elements are the most abundant among SINEs, and are also the mobile element with the highest copy number in our genome (there are 1,000,000 copies of them). Alu elements are derived from the 7SL RNA (Ullu & Melli, 1982), which is an integral part of the ribosome. SVA elements are much less abundant in the genome, and there are less than 3000 copies per genome. SVAs are characterized by being composed of portions of other retrotransposons, and are exclusive to higher primates (Dewannieux *et al*, 2003; Hancks *et al*, 2011; Hancks & Kazazian, 2012).

1.3. Impact of mobile elements in genomes

Due to the myriad of processes associated with their mobilization, it is undeniable that mobile elements have had a major role in the evolution of different organisms, driving genome evolution in a variety of ways (Kazazian & Moran, 1998; Belancio *et al*, 2008; Kazazian, 2004; Goodier & Kazazian, 2008). New mobile element insertions can lead to changes in genome function and/or regulation, changes that can be beneficial, neutral or mutagenic. The eukaryotic genome evolution has been driven by different processes, including the breakage and joining of different chromosomes (translocations), the combination of functional domains into exons, gene conversion, as well as gene and segment duplication (**Figure 6**).

During evolution, mobile elements have contributed to the creation of important cellular functions, such as adaptative immune responses and Telomerase. In mammals, non-LTR retrotransposons have had a significant impact on the evolution of genomes, and for about 500-600 million years have been a major force in evolution (Kazazian, 2004). There are several examples of functions that seem to have been acquired from mobile elements, and are functions where RTs play a major role. Currently, in the mitochondrial and chloroplastic genomes of fungi and plants, as well as in some bacteria, we can find what are the likely ancestors of LINE elements: **mobile group II introns**. Group II introns encode a

RT very similar to that encoded by other non-LTR retrotransposons (Malik *et al, 1999;* Belfort *et al,* 2002*)*, and use a similar mobilization mechanism (i.e., retrohoming). Notably, similar reverse transcriptase-containing elements also inhabit some yeast genomes, including *Candida albicans* (Goodwin *et al,* 2001*)*. Thus, RT activity represent a major mechanism driving genome evolution.

It has been suggested that non-LTR retrotransposons could be the **ancestors of telomerases** because of similarities between telomerase and the RTs of non-LTR retrotransposons (Eickbush, 1997; Kopera *et al,* 2011*)*. Favoring this hypothesis are *Drosophila spp.* and other insects, which lack conventional telomeres and telomerase but use two non-LTR retrotransposons, **TART** and **HeT-A**, to maintain telomeres, by accumulating *de novo* insertions at chromosome ends (Biessmann *et al,* 1992*;* Levis *et al,* 1993*)*. Further consistent with this hypothesis, some rotifers exploit a mix of **Telomerase and LINE-like retrotransposons** to maintain telomeres (Curcio & Belfort, 2007). In fact, even currently active human LINE-1s could insert at telomeres when these are unprotected (Morrish *et al,* 2007*)*, further supporting that Telomerase is likely a LINE element that has been domesticated to perform a cell-specific function.

But not all acquired functions are related to RTs; strong evidence suggests that DNA Transposons contributed to the **V(D)J recombination system** of the immune system, which is responsible for generating antibody variability for antigen recognition. It has been demonstrated that mobile elements have given rise to RAG endonucleases, which function as transposases *in vitro*, having probably evolved from DNA transposons (Kapitonov & Jurka, 2005).

While the above are classical examples of "domestication" processes, the main activity of mobile elements, that is accumulate more copies in genomes, can also drive genome evolution. **Exon shuffling** (Moran *et al,* 1999*)* is an example of such processes, which occurs when coding sequences adjacent to active non-LTR retrotransposons are mobilized to new genomic locations, which can lead to the generation of new cellular functions (Sayah *et al,* 2004*)*.

While the above are examples of how mobile elements can positively impact genome evolution, there are also numerous reports documenting negative impacts associated with mobile elements. Mobile elements can **create genetic instability** through non-allelic homologous recombination and by introducing sequences that alter gene expression (Figure 6). Additionally, due to their activity, mobile elements have potential to cause DNA double-strand breaks or to modulate the epigenetic state of chromosomes. To avoid these negative impacts, and acting to regulate the activity of mobile elements, the cell has evolved multiple mechanisms to restrict the activity of mobile elements. In section 4, I will discuss specific mechanisms that have evolved to control the activity of LINE-1 retrotransposons, currently active in humans *(4. Mechanisms regulating LINE-1 retrotransposition)*.

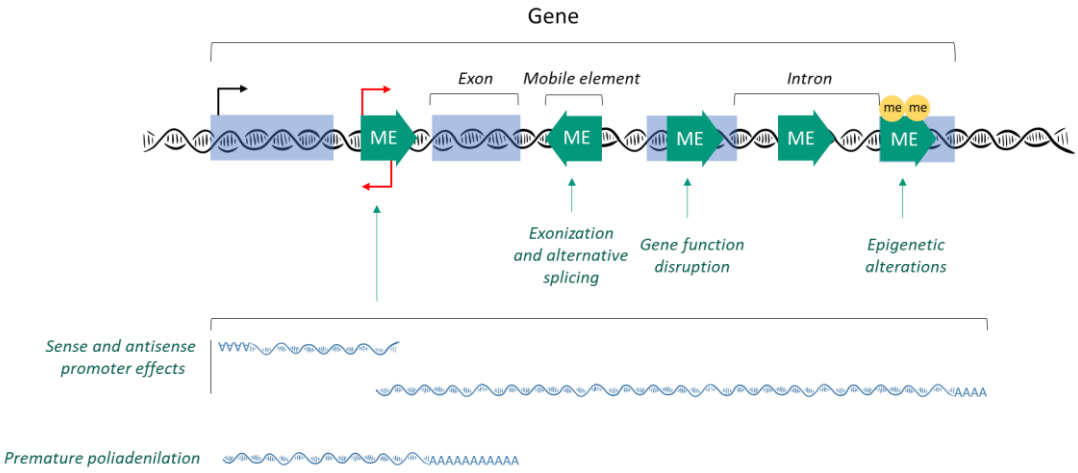


Figure 6. The impact of Mobile Elements (ME). *A* de novo *mobile element insertion can impact genes in a myriad of ways (some depicted in this figure). The sense and antisense promoter effects are characteristic of LINEs. Adapted from Macia et al, (2015) and Garcia-Perez et al, (2016).*

2. Active LINE-1 retrotransposons

As described above, LINE-1 or L1 elements are the only autonomous mobile element class currently active in our genome (Macia et al, 2015; Richardson *et al*, 2015), and their activity over evolution has created a large proportion of our genome. Indeed, because they are responsible for the mobilization of SINEs, at least one third of our genome has been generated by the activity of a single mobile element: LINE-1s. Despite their prevalence and impact, the basic mechanisms that regulate their mobility are largely unknown.

2.1. Structure of retrotransposition-competent L1s (RC-L1s)

Despite their abundance, only a small fraction of LINE-1s are currently active, and are known as retrotransposition competent L1s, or RC-L1s. A prototypical **RC-L1** (Macia *et al*, 2015; Richardson *et al*, 2015) has a length of ~6,000bp and, from 5'to 3', contains:

- A 900 bp long **5' UTR**. This sequence contains a SENSE internal RNA polymerase II promoter (Swergold, 1990) which is required to generate the intermediary mRNA (**Figure 7**). In addition, this region harbors RNA polymerase II ANTISENSE promoter activity (Speek, 2001; Macia *et al*, 2011), generating antisense transcripts that when translated give rise to **L1-ORF0p** (Denli *et al*, 2015; Macia *et al*, 2015)**.**

- **L1-ORF1p,** coding for a 40 kDa protein containing an RNA Recognition motif (RRM) that confers RNA binding capability (Khazina & Weichenrieder, 2009), and a leucine zipper domain (lz). Furthermore, L1-ORF1p is endowed with nucleic acid chaperone activity (Hohjoh & Singer, 1997; Martin & Bushman, 2001). The biochemical activities of L1-ORF1p are strictly required for LINE-1 mobilization (Moran *et al*, 1996) (**Figure 7**).

- A 63 bp long intergenic sequence, separating both non-overlapping ORFs.

- **L1-ORF2p**, coding for a 150 kDa multidomain protein; several domains have been identified within L1-ORF2p, which from N- to C-terminus include: i) an APE-like endonuclease (**EN**, Feng *et al*, 1996); ii) a PCNA-Interaction Protein **(PIP) motif**, allowing to interact with PCNA (proliferating cell nuclear antigen) (Taylor *et al*, 2013); iii) a **RT**, (Mathias *et al*, 1991; Richardson *et al*, 2015); and iv) a **C-terminal domain** with a **zinc knuckle domain (rich in cysteines)** (**Figure 7**). All enzymatic activities and domains of L1-ORF2p are strictly required for LINE-1 mobilization

(Moran *et al,* 1996*)*.

- A ~250 bp long **3' UTR,** containing a weak polyadenylation signal (Moran *et al,* 1999*)* and a **polyadenine tail (polyA),** that is required for LINE-1 mobilization (Doucet *et al,* 2015*)*.

As described above, due to characteristics of their mobilization process, most LINE-1s in our genome are flanked by short **TSDs** of 2 to 20 bp.

Figure 7. Long INterspersed Element class 1 (LINE-1) structure. *UTR, untranslated region; CC, coiled coil; RRM, RNA recognition motif; CTD, carboxyl-terminal domain; EN, endonuclease; RT, reverse transcriptase; C, cysteine-rich domain. Adapted from Beck et al, 2011.*

In 2003, upon completion of the first human genome reference sequence, all human LINE-1s meeting the above criteria were assayed for activity in cultured cells. These analyses revealed that an average human genome contains **80-100 RC-L1s**, that is LINE-1 elements with the ability to mobilize in our genome (Brouha *et al,* 2003*;* Beck *et al,* 2010*)*. However, the same study found that within the 80-100 RC-L1s, there is a small group of active LINE-1s that could retrotranspose at the highest level, and these are known as "hot" RC-L1s.

2.2. LINE-1 retrotransposition mechanism

LINE-1 mobilization is known as retrotransposition, which takes place by a mechanism known as ***Target Primed Reverse Transcription*** or **TPRT** (Luan *et al,* 1993; Cost *et al,* 2002**),** originally discovered using a non-LTR retroelement from the silkworm *Bombyx mori*.

The LINE-1 (L1) retrotransposition cycle begins with the transcription of a full-length sense L1 mRNA from any RC-L1 in our genome, generating a polyadenylated L1-mRNA. The L1-mRNA is exported to the cytoplasm, where despite its unconventional bicistronic nature, it is translated (Alisch, 2006; Dmitriev *et al,* 2007*)*, generating hundreds of L1-ORF1p molecules and as little as one molecule of L1-ORF2p per L1-mRNA. This is because L1-ORF1p is translated using an efficient and canonical cap-dependent process, while L1-ORF2p translation occurs at reduced levels using an unconventional and inefficient termination/reinitation mechanism (Alisch, 2006).

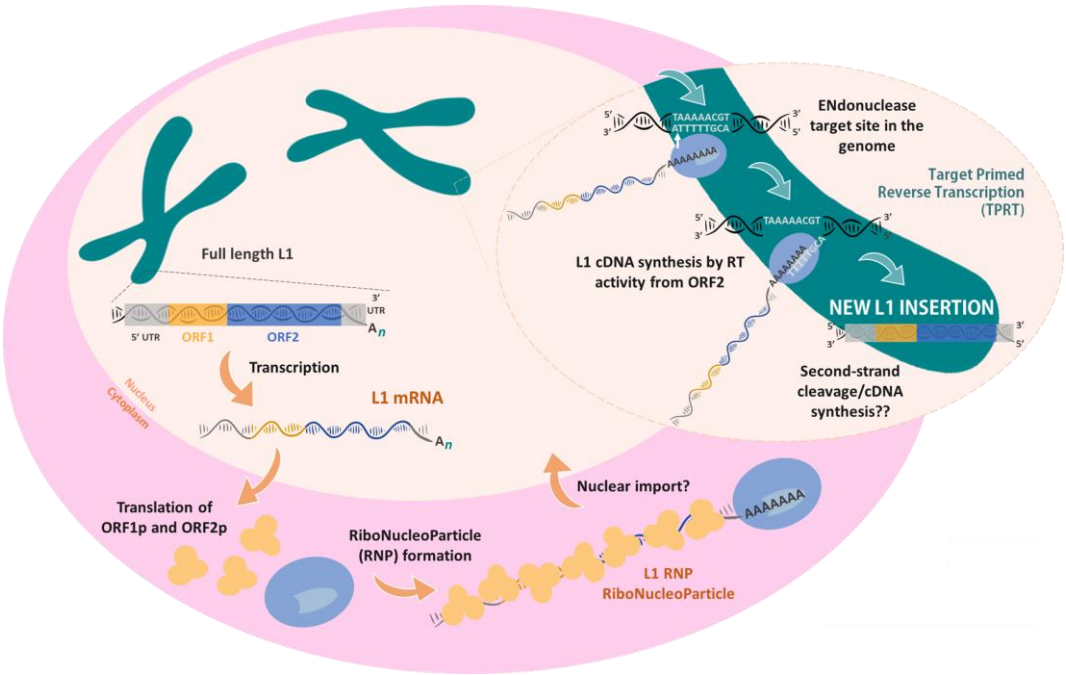


Figure 8. The LINE-1 retrotransposition cycle. *See text for details. Adapted from Beck et al, 2011.*

Upon transcription and translation (Swergold, 1990; Alisch, 2006; Dmitriev *et al*, 2007), both L1-encoded proteins bind to their encoding L1-mRNA, in a process known as cis-preference (Wei *et al*, 2001), generating a L1 RiboNucleoprotein Particle **(L1-RNP)** (Doucet *et al*, 2010; Kulpa & Moran, 2005), considered as the intermediary of LINE-1 retrotransposition. L1-RNPs subsequently enters the nucleus by a process that does not require cell division (Kubo *et al*, 2006; Macia *et al*, 2017). Once in the nucleus, through TPRT, a fully mature L1-RNP can generate a *de novo* insertion in the genome; TPRT starts when the EN activity of L1-ORF2p recognizes a consensus sequence (5'TTTTT/AA and variants (Jurka, 1997; Flasch *et al*, 2019) in the bottom strand of genomic DNA, generating a single cut that expose a free 3' OH; the exposed 3' OH is subsequently used as a primer by the L1 RT activity to initiate synthesis of the first cDNA of the new L1 copy; to note, pairing of the L1 mRNA polyA tail with the exposed T-rich stretch at the site of cleaving is thought to facilitate initiation of reverse transcription, as snap-velcro (Monot *et al*, 2013). As a result, the first cDNA copy of the new L1 element (L1-cDNA) is anchored to the genome (**Figure 8**).

The initial steps of TPRT generate a L1-mRNA:L1-cDNA hybrid covalently linked to the genome (i.e., a Y-branch intermediate). According to how the RNA might be removed from the L1-mRNA:L1-cDNA hybrid, there are two main classes of LINEs: elements encoding a functional RNase H domain (mainly present in plants and lower eukaryotes) and elements lacking this domain (most mammals, including human LINE-1s); (Malik *et al*, 1999; Olivares *et al*, 2002; Piskareva *et al*, 2003). While the RNase H activity encoded by LINEs would degrade the RNA strand from the L1-mRNA:L1-cDNA heteroduplex, **it is currently unknown how LINE elements without a functional RNase H domain would accomplish this**. In this context, the RT of some non-LTR retroelements can displace RNA strands annealed to ssDNAs without losing processivity (Kurzynska-Kokorniak *et al*, 2007), but such activity is very rare in polymerases.

Although how synthesis of the second strand of the new L1 insertion occur is not fully understood, it requires cleaving of the top strand of genomic DNA, exposing a new 3´OH that ultimately will prime synthesis of the second-strand of the new L1 insertion. *In vitro*, L1-ORF2p could cleave the top strand of genomic DNAs, and as other RTs, the L1 RT can generate dsDNAs using cDNA as a template. Thus, it is likely that L1-ORF2p is directly involved in the generation of the second strand of the new L1 insertion (Cost *et al*, 2002; Piskareva *et al*, 2003; Piskareva & Schmatchenko, 2006; Richardson *et al*, 2015). The result of TPRT is the generation of a new LINE-1 insertion, usually flanked by short TSDs, and often 5′ truncated (Richardson *et al*, 2015).

Despite 5´truncation during TPRT, a process not fully understood but where DNA repair proteins seem to play a role (Coufal *et al*, 2011*)*, there are other mechanisms that can lead to the insertion of "dead on arrival" L1 copies, such as splicing of the L1 mRNA (Larson *et al*, 2018*)* or premature polyadenylation (Perepelitsa-Belancio & Deininger, 2003).

Additionally, while the majority of *de novo* L1 insertions are generated via the aforementioned endonuclease-dependent mechanism, previous studies (Sen *et al*, 2007*)* have revealed that in the absence of non-homologous end joining DNA repair proteins, L1 elements can use an alternative endonuclease-independent insertion pathway. Endonuclease-independent (**ENi**) L1 insertions are thought to occur in sites of DNA disrepair, such as single or double-strand breaks. In these insertions, the polyA tail of the L1 RNA would anneal with exposed free ends (i.e., 3´OH) at sites of DNA disrepair, allowing the RT activity to generate an L1-cDNA. How second strand synthesis would occur is completely unknown, but it seems to rely on enzymatic activities encoded within L1. Consistently, ENi-L1 insertions are structurally atypical and distinguishable from canonical L1 insertions, and are often associated with alterations at the insertion site (deletions, inversions and even interchromosomal translocations of genomic DNA). However, it is not known whether ENi-retrotransposition might operate *in vivo* or only in cells where DNA repair mechanisms are dysfunctional.

2.2.1. A system to study retrotransposition: The retrotransposition assay

Using a clever design originally developed by Boeke and colleagues to study mobilization of LTR retrotransposons from yeast (Boeke *et al*, 1985*)*, in 1996 Moran and colleagues developed the first engineered assay to study human L1 retrotransposition in cultured cells (Moran *et al*, 1996*)*. The LINE-1 retrotransposition assay exploits vectors carrying a **full-length human RC-L1**, normally a **"hot" RC-L1**, tagged with a retrotransposition cassette containing a reporter gene whose expression can only be activated after a round of bona fide retrotransposition. Several **reporters** have been used to follow retrotransposition, including Enhanced Green Fluorescent Protein (EGFP), several genes conferring resistance to mammalian antibiotics (Neomycin, Blastidicin) and more recently luciferase. The rationale of the retrotransposition assay is based on cloning a **retrotransposition cassette in the opposite orientation of L1 transcription** (i.e., antisense); furthermore, the antisense reporter gene is interrupted by an intron, cloned in the sense orientation of L1 transcription (**Figure 9**). With this configuration, **a functional reporter can only be expressed after the completion**

of a round of LINE-1 retrotransposition, as the intron is removed from the intermediate L1-mRNA prior to reverse transcription/integration in genomic DNA. Thus, the new insertion will contain the reporter cassette without the intron, activating expression of a functional reporter, which can be used as a proxy to quantify retrotransposition efficiency in cells. While developed >20 years ago, the reporter based retrotransposition assay is still the gold standard in the field, and our knowledge of L1 biology has dramatically increased since its original description.

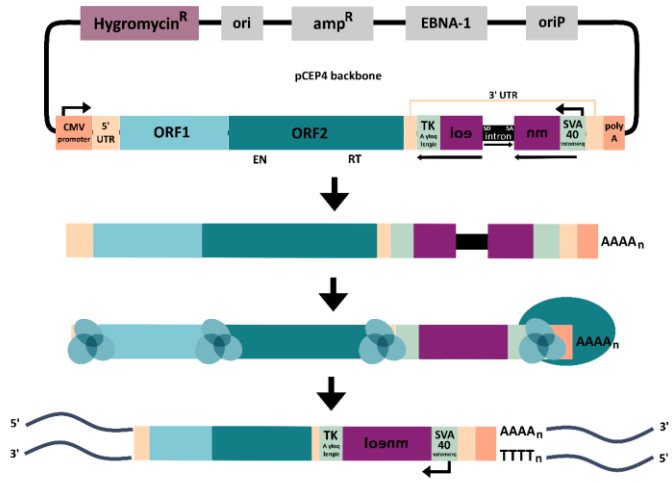


Figure 9. The retrotransposition assay. *Upon transfection, the LINE-1 element is transcribed from the upstream CMV promoter, and transcription is terminated by a SV40 polyadenylation signal (denoted as polyA) located downstream of the tagged L1. Once transcribed, the intron undergoes splicing and the intermediate RNA is exported to the cytoplasm, where L1-ORF1p (light blue circles) and L1-ORF2p (green circle) are translated from the chimeric L1-mRNA. Only after reverse transcription and integration into a genomic locus, the reporter gene (antibiotic resistance gene* mneoI *in this example) can be expressed to confer antibiotic resistance to cells. An example of a retrotransposon-competent LINE-1 vector is plasmid pJM101/L1.3 (a plasmid used in this Thesis), which contains the mneoI reporter cassette (purple) at the 3' UTR region in the opposite orientation to the transcription orientation of LINE-1. The reporter gene, neomycin phosphotransferase, is interrupted by an intron (in black), which is in the same orientation with respect to the transcription direction of LINE-1. The pCEP4 plasmid backbone encodes for the viral EBNA-1 protein (EBNA-1) and contains an origin of viral replication (oriP) and the hygromycin B resistance gene (hygR) for plasmid replication and hygromycin selection in human cells. The plasmid backbone has also a bacterial origin of replication (ori) and ampicillin resistance gene (ampR) for replication and ampicillin selection, respectively, in E. coli.*

3. LINE-1 retrotransposition in pluripotent cells

Mobile elements are the prototype of "selfish DNA", whose only function is to accumulate more copies of themselves in cells able to transmit genetic information to the next generation. LINE-1 elements are not an exception and accumulate new L1 copies in the heritable genome, to ensure their evolutionary success. In mammals, there are two cellular niches where the accumulation of new L1 insertions will allow their passing to the next generation: **mature germ cells** and **early embryogenesis**. Thus, and over the years, several labs have analyzed LINE-1 expression and activity in these two cellular niches.

In 2007, Van den Hurk *et al, demo*nstrated that LINE-1 retrotransposition could occur during **early stages of human embryonic development**; while characterizing a *de novo* mutagenic L1 insertion in a patient affected with choroideremia (an X-linked eye disease), these authors demonstrated that the L1 insertion actually occurred during the mother's early embryogenesis, before germ line segregation. In a back-to-back study, Garcia-Perez *et al,* 2007 used human embryonic stem cells (hESCs) as a model of early human embryogenesis and reported overexpression of LINE-1 RNAs, including RNAs from presumably active L1Hs copies, as well as expression of retrotransposition intermediates (L1-RNPs). Furthermore, using the engineered retrotransposition assay, these authors reported a low level of retrotransposition in a panel of hESC lines. More recently, the mobilization of endogenous LINE-1s and SINEs has been reported in hESCs and induced pluripotent stem cells (hiPSCs), using Next Generation DNA sequencing (NGS) approaches (Wissing *et al,* 2012*;* Klawitter *et al,* 2016*;* Muñoz-Lopez *et al,* 2019*)*.

Complementing the above studies, Kano *et al,* 2009 analyzed LINE-1 expression and retrotransposition in a mouse model of human L1 retrotransposition, which used the same rationale of the engineered L1 retrotransposition assay. These authors reported expression of LINE-1 RNAs in developing germ cells and meiotic cells. Intriguingly, Kano et al. demonstrated that **L1-mRNAs could be deposited by both male and female germ cells**, and that could be integrated in the genome during late embryogenesis. Kano and colleagues also explored the frequency of retrotransposition in germ cells and early embryogenesis, and reported that most LINE-1 integrations seemed to occur after fertilization, and rarely in germ cells. Interestingly, Kano *et al.* (2009) also found that some animals with mosaic *de novo* L1 insertions couldn't transmit them to their offspring. Thus, the frequency and timing of heritable retrotransposition of L1 *in vivo* was unclear.

Later in 2017, Richardson and colleagues adapted NGS methods to capture-and-sequence active retrotransposons from the mouse genome (mouse RC-seq) (Baillie *et al*, 2011; Shukla *et al*, 2013). Briefly, mouse RC-seq (mRC-seq) was employed to study *de novo* insertion accumulation in two and three-generation mouse pedigrees. By applying this technology to pedigrees of wild-type C57BL/6J mice, they investigated the rate and timing of heritable L1 insertions. For this purpose, mRC-seq and whole genome sequencing were used to identify retrotransposon insertions absent in the reference genome and not previously identified in transposable element polymorphism analyses. Insertions characteristic of the parental generation were identified and their origin investigated. At the same time, sequencing of different tissues of the specimens of each generation of the pedigrees was carried out. Thus, homozygous, heterozygous, mosaic specimens for *de novo* insertions or those presenting the insertion in the germline were characterized, analyzing the origin of the new insertions and how they were transmitted to the progeny. Richardson and colleagues demonstrated that, although **germ cells could accumulate *de novo* L1 insertions, most *de novo* LINE-1 and SINE insertions accumulate during early embryogenesis** (Richardson *et al*, 2017). Richardson and colleagues also demonstrated that *de novo* L1 insertions occurring in primordial germ cells, pluripotent cells from the early embryo or in adult germ cells, during gamete development, could be transmitted to new generations.

Consistently, Freeman and colleagues exploited NGS to analyze *de novo* L1 retrotransposition in human sperm, and reported a frequency of retrotransposition of **<1 insertion per 400 haploid genomes** (Freeman *et al*, 2011). Finally, Feusier *et al*, (2009) exploited NGS of human pedigrees to analyze the ongoing frequency of retrotransposition in the human genome, reporting rates of **1 insertion in 40 births and of 1 in 63 births for Alu and LINE-1**, respectively, after analyzing a panel of >600 sequenced human genomes (Feusier *et al*, 2019). Remarkably, in agreement with previous studies, Feusier and colleagues demonstrated that the majority of these insertions occur during early embryogenesis.

Thus, in mammals the load of heritable retrotransposition occurs during early embryogenesis. Because research with human embryos implies a significant number of ethical considerations, and it is strictly prohibited in many countries around the world, hESCs and other pluripotent cells mimicking early human embryogenesis [i.e., hiPSCs and human embryonic carcinoma cells (hECs)] have become an excellent model to study L1 biology under physiological conditions.

Indeed, the number of L1 biology studies using pluripotent cells has increased over the past years. In this context, Flasch *et al, (*2019) characterized >88,000 engineered LINE-1 insertions in different cell lines, including hESCs; Flasch and colleagues compared the hallmarks associated with *de novo* L1 retrotransposition events in several human cell lines that mimic cellular niches where endogenous L1s are known to mobilize, such as cancer (using HeLa cells), pluripotent human cells [using hECs (PA-1) and hESCs (H9)] and brain cells (see below, using neural progenitor cells (NPCs) derived from hESCs). Remarkably, despite previous assumptions, Flasch and colleagues demonstrated that the **EN activity of L1-ORF2p is the main determinant to direct L1 integration** in the human genome, and that the **epigenetic status of preintegration sites is irrelevant for L1 integration.**

In summary, a number of studies have analyzed L1 expression in pluripotent cells, concluding that **hESC, hiPSC and hECs lines naturally overexpress a wide range of L1 and Alu RNAs** (from active and inactive subfamilies), as well as LINE-1 retrotransposition intermediates **(L1-RNPs)** (Hohjoh & Singer, 1997; Kulpa & Moran, 2005; Martin, 1991; Garcia-Perez *et al,* 2007; Holmes *et al,* 1992; Skowronski *et al,* 1988; Macia *et al*, 2011, 2017). Similarly, our laboratory and others have shown that new **L1 insertions accumulate in early stages of human embryonic development**, specifically in **our PLURIPOTENT genome**. Collectively, these studies demonstrate that pluripotent models, are the best models for understanding how our genome regulates LINE-1 mobility.

3.1. Impact of L1 retrotransposition in genomes

As discussed above, evidence from different contexts such as cell lines, animal models and patient characterization has demonstrated that, in humans, most *de novo* L1 retrotransposition events accumulate during early stages of human embryonic development (Van den Hurk *et al,* 2007; Garcia-Perez *et al,* 2007, 2010; Klawitter *et al*, 2016; Kano *et al*, 2009; Wissing *et al*, 2012*)*. Retrotransposition during early embryonic development has an evolutionary meaning for LINE-1s, as a fraction of these insertions would pass to the next generation, ensuring the evolutionary success of these mobile elements (Garcia-Perez *et al*, 2016*)*.

However, several studies have shown that **L1 activity and its impact are not limited to early embryogenesis or germ cells**. Indeed, it is clear now that L1 is expressed and frequently retrotranspose in two other cellular niches: **cancer cells** (Miki *et al*, 1992; Moran *et al*, 1996; Feng *et al*, 1996*)* and the **brain**, two non-heritable cellular niches where the potential benefits to L1, evolutionary or not, are less clear. In 1992, a seminal study by Miki *et al, demo*nstrated ongoing L1 retrotransposition

in colon cancer. Since then, and partly driven by the Genomics Revolution and the use of NGS methods, other studies have conclusively demonstrated that **L1 is overexpressed in most human cancer types**, and that **L1 retrotransposition in cancer patients is very common**, especially in tumors of an epithelial origin (Moran *et al*, 1996; Feng *et al*, 1996; Burns, 2017; Iskow *et al*, 2010; Helman *et al*, 2014; Tubio *et al*, 2014; Scott & Devine, 2017; Rodriguez-Martin *et al*, 2020). While several studies have also demonstrated that the impact of L1 insertions can influence tumor progression (Miki *et al*, 1992; Shukla *et al*, 2013; Rodriguez-Martin *et al*, 2020), the overall importance of L1 expression/retrotransposition in cancer remains to be analyzed in further detail, and from different angles (i.e., could L1 expression be used as a biomarker to stratify cancer patients during diagnosis and treatment? Would L1 inhibition lead to a better response to current cancer therapies? Can L1 retrotransposition drive chemotherapy resistance?).

Similar to cancer cells, it is well established now that L1 is expressed and can retrotranspose in selected cells from the mammalian brain. In 2005, Muotri *et al*, *firs*t found that **L1 RNAs are expressed in Neuronal Progenitor Cells (NPCs)** from the brain of rats and mice (Muotri *et al*, 2005). Muotri and colleagues also demonstrated that human L1s could retrotranspose at a high frequency in rodent NPCs, suggesting that the genome of brain cells might be more variable than envisioned. Later on, other studies demonstrated that human **L1s are also expressed and retrotranspose at a high level in human NPCs**, derived from hESCs or isolated from human fetuses (Coufal *et al*, 2009, 2011), and also in mature non-dividing neurons (Macia *et al*, 2017). Furthermore, and although with significant variability in deduced retrotransposition rates, NGS methods have conclusively demonstrated that endogenous L1s retrotranspose in NPCs and mature neurons in our brain. As with cancer cells, much research remains to be conducted to learn whether L1 expression/retrotransposition in the brain has any functional significance for brain biology and/or L1 biology.

Similarly, it has been observed that **during cellular senescence L1 elements are no longer repressed** and have been linked to the development of inflammation associated with aging. However, the extent to which retrotransposons may contribute to aging remains unknown. In 2019, De Cecco and colleagues described how during cellular senescence LINE-1 was no longer transcriptionally repressed, resulting in the activation of a type-I interferon (INF-I) response. This response is characteristic of late senescence and is triggered by cytoplasmic L1-cDNAs. The authors of this study observed how treatment of aged mice with L1 reverse transcriptase inhibitors resulted in a decrease in the activation of the INF-I response and a reduction in age-associated inflammation (De Cecco *et al*, 2019). In

a similar manner, Simon and colleagues (2019), exploited mice deficient for SIRT6, which exhibit reduced life expectancy, growth retardation, and elevated LINE-1 activity, to further explore L1 activity in aged cells. By treatment with L1 reverse transcriptase inhibitors they were able to observe an improvement in the health and life expectancy of these mice, with an overall reduction in the type I interferon response (Simon *et al*, 2019). Thus, modulation of L1 activity could be a strategy to reduce the impact of age-associated pathologies.

Due to their ongoing activity in humans, **new L1 insertions can sporadically cause an array of human disorders since birth**. Mobile element insertions within coding sequences can lead to its inactivation or interfere with its regulation, ultimately inducing a phenotype or disease (Macia *et al*, 2015; Richardson *et al*, 2015). Similarly, new insertions in introns and/or regulatory sequences (promoters, 5 and 3 UTRs, etc) can cause changes in gene regulation processes through various mechanisms (see section *1.3. Impact of mobile elements in genomes*, **Figure 6**). In addition, L1 insertions (in 10-30% of the cases) can be accompanied by severe genome modifications at the insertion site, ranging from simple deletions of different size to potential chromosomal translocations. Thus, due to ongoing L1 activity, and since LINE-1 insertions are inserted randomly in our genome, the range of diseases produced by L1 is very wide, ranging from cancer to hemophilia (Macia *et al*, 2015; Garcia-Perez *et al*, 2016; Hancks & Kazazian, 2016). Notably, not only active mobile elements can generate human disorders, as fixed/resident insertions can induce recombination processes due to their inherent repetitive nature, and by several mechanisms (allelic recombination, non-allelic recombination, Single Strand Annealing recombination (SSA), etc).

To avoid the deleterious impact of LINE-1 activity, cells have evolved different mechanisms to control their activity, acting at distinct steps of the LINE-1 cycle, to ultimately prevent their uncontrolled spreading and to confer stability to our genome. Notably, these mechanisms might be critical in our pluripotent genome, as LINE-1s are naturally expressed and mobilize during early embryonic development, in our heritable genome. The aim of this Thesis is to better understand how active LINE-1s are regulated in pluripotent cells, and I will next discuss what is known about LINE-1 regulation.

4. Mechanisms regulating LINE-1 retrotransposition

As discussed above, and over the course of evolution, the mutagenic potential of mobile elements has driven the human genome to develop mechanisms to reduce their activity, but much remains to be discovered about these mechanisms controlling LINE-1 activity. Indeed, **it has been demonstrated that L1 activity is influenced by the combined action of repressors and activators**. Over the years, several studies have reported different mechanisms acting to regulate expression and/or retrotransposition of LINE-1s, and this is an area of intense research in the field of mobile element biology (recently reviewed in Garcia-Perez *et al*, 2016; Heras *et al*, 2014; Pizarro & Cristofari, 2016). Due to their evolutionary implications, those mechanisms acting to restrict LINE-1 expression and mobilization in germ cells and embryonic pluripotent cells are particularly important (Crichton *et al*, 2014). Below I will discuss known mechanisms that control L1 expression/retrotransposition, with emphasis on those relevant in pluripotent cells.

4.1. Epigenetic control of L1 expression: regulation at the chromatin level

To mobilize, RC-L1s must be first transcribed using their internal promoter, which is located in their 5'UTRs. Therefore, a major mechanism acting to regulate L1 mobility is to prevent their transcription, exploiting several epigenetic pathways. Below I will describe epigenetic mechanisms that keep mobile element transcriptionally inactive, further preventing their mobility:

4.1.1. DNA methylation of the L1 promoter

The LINE-1 promoter contains a canonical CpG island with >20 CpGs, which when methylated inhibit LINE-1 expression (Thayer *et al*, 1993; Bourc'his & Bestor, 2004; Coufal *et al*, 2009). Methylation occurs through the action of a mammalian metabolic cascade involving DNA methyltransferases (DNMTs) that modify CpG dinucleotide-rich islands, inducing cytosine methylation, which hinders the binding of transcription factors and RNA II polymerase. Remarkably, the methylation of CpG islands in promoters is a repressive mechanism exploited by mobile elements but also by cellular genes. Indeed, it has been proposed that regulation by DNA methylation might be another case of domestication, where a mechanism arising to combat mobile elements has been recycled to regulate expression of genes (Martienssen, 1998).

It has been demonstrated that DNMT3L drives *de novo* methylation of retroelement promoter sequences in primordial germ cells (Bourc'his & Bestor, 2004). Similarly, DNMT1 has been shown to regulate expression of RC-L1s in human NPCs (Jönsson *et al,* 2019*)*. To note, genomes are naturally hypomethylated during early embryogenesis. However, and despite this apparent relaxation of DNA methylation, it has been demonstrated that interfering with the activity of DNMTs in hESCs results in de-repression of LINE-1 RNAs, mostly affecting evolutionary young L1s (L1Hs and L1PA2)(Castro-Diaz *et al,* 2014*)*. Similar to embryonic cells, two recent studies in mouse germ cells (Bourc'his & Bestor, 2004; Barau *et al,* 2016*)* reported how the loss of different DNMTs result in increased LINE-1 RNA levels. In sum, **DNA methylation is a major mechanism controlling expression of evolutionary young LINE-1s, in somatic and pluripotent cells.**

While DNA methylation is key to regulate L1, how the promoters of young L1Hs are recognized, by DNMTs or by accessory proteins, is not completely known. Remarkably, a recent study uncovered that a conserved binding site for Ying/Yang1 (YY1), located in the first 30 nucleotides of the 5´UTR of young and active L1s (i.e., L1Hs elements), is required to methylate the L1 promoter in differentiated cells. Indeed, the study by Sanchez-Luque et al. further demonstrated that truncated L1s lacking the binding site for YY1 are not methylated and are expressed in differentiated cells. Despite the critical role of YY1, it is very likely that other factors are involved in targeting the promoter of young L1s for methylation (Sanchez-Luque *et al,* 2019*)*.

Another modification associated with DNA methylation is the transformation of 5-methylcytosine (5-mC) into 5'hydroxymethylcytosine (5-OHmC), which is carried out by the TET (Ten-eleven translocation) family of proteins. This modification has been observed in LINE-1 promoters of pluripotent cells (Branco *et al,* 2012*)*, as well as within LINE-1 sequences, but its epigenetic role is unclear.

The methylation level of mobile element promoters is not constant in every cell type (Faulkner *et al,* 2009*)*. Additionally, beside DNA methylation levels, there are additional epigenetic factors that influence mobile element expression in human cells. In fact, several epigenetic mechanisms involved in DNA methylation and neurodevelopmental diseases have been found to be specific to certain cell types; these include histone deacetylase 1 (HDAC1), the transcription factor Sox2, and the methyl-CpG-binding protein 2 (MeCP2), which form a repressor complex acting on the LINE-1 promoter controlling neuronal transcription of LINE-1s, specifically in NPCs and neural stem cells (NSCs) (Muotri *et al,* 2010*)*.

There are additional epigenetic cofactors influencing L1 expression that do not directly involve methylation, but that can act in a cell type or locus specific manner: KRAB proteins. In fact, a mechanism which acts at the epigenetic level, specifically on chromatin and repressing LINE-1 expression is mediated by the KRAB family of proteins. KAP1 (KRAB-associated protein 1) mediates heterochromatin formation and is recruited at LINE-1s by zinc finger proteins containing KRAB motifs (KRAB-ZFPs). In hESCs, KRAB-ZFPs often bind the 5' end of full-length evolutionary old LINE-1 elements, and mediate their transcriptional repression. Thus, in hESCs this mechanism of repression operates only in currently inactive old L1 subfamilies (L1PA3, L1PA4 and L1PA5) (Castro-Diaz *et al*, 2014; Pezic *et al*, 2014). Further analyses generated a model fitting the red queen hypothesis, in which KRAB-ZNFs co-evolve with mobile elements in parallel, getting adapted to their modifications in order to continue repressing their expression, likely explaining why evolutionary young active L1s are not targeted, yet, by repressive KRAB-ZNFs (Imbeault *et al*, 2017; Ecco *et al*, 2017). In addition, the HUSH complex has been suggested to mediate chromatin remodeling and transcriptional silencing of young retrotransposition competent LINE-1s in cooperation with KAP1, and in hESCs (Liu *et al*, 2018).

In summary, DNA methylation of mobile element promoters is a major mechanism to control the mobility of LINE-1s in our heritable genome, together with KRAB-ZNFs and other epigenetic cofactors. Notably, the epigenetic repression of L1 extend early embryogenesis and also operates in somatic cells, preventing the accumulation of insertions in somatic cells, which has no evolutionary benefit for L1 and that it could be detrimental for the host (i.e., new L1 insertions can create mutations) (Bourc'his & Bestor, 2004).

4.1.2. Histone deacetylation as a possible L1 restriction mechanism.

As discussed above, **genomes are naturally hypomethylated during the development of primordial germ cells (PGCs) and shortly after fertilization, during early embryogenesis** (Zeng & Chen, 2019). After fertilization, there is a global loss of methylation in genomes, which is key for embryo preimplantation and for gastrulation/organogenesis (as embryonic cells need to regulate their differentiation into several lineages, restricts differentiation and prevents going back to an undifferentiated state (Zeng & Chen, 2019; Messerschmidt *et al*, 2014)). Notably, this window of hypomethylation is exploited by mobile elements to ensure their evolutionary success. This might explain why LINE-1 insertions accumulate at this level, and since this occurs shortly after fertilization, most of the new insertions will be present in the majority of germ cells of the neonate, ensuring

the evolutionary success of L1. In other words, the "defect" of cell reprogramming is used by LINE-1 elements to amplify and perpetuate in genomes. Thus, it stands to reason that **additional mechanisms, different from DNA methylation, might operate in pluripotent embryonic cells to regulate expression/activity of mobile elements**, since frequent and uncontrollable L1 mobilization would induce a high number of mutations that might be incompatible with life.

Despite the potential of hESCs/hiPSCs to study early human development processes, these cells are notoriously difficult to grow, they transfect poorly, and tend to lose their pluripotent status due to spontaneous differentiation (Dakhore *et al*, 2018). Thus, in order to better understand L1 accumulation in pluripotent cells, in 2010 Garcia-Perez *et al,* analyzed whether hEC lines could be more useful as models of early embryogenesis. Briefly, hECs have a transcription profile very similar to hESCs (Sperger *et al,* 2003*),* and prior to the first derivation of hESCs in 1998, hECs were the only available model to study early human embryogenesis. Over the years, several human hEC lines have been derived from teratomas, which are benign germ cell tumors, and as hESCs, some hEC lines retain their pluripotency but can differentiate into the three main germ layers (ecto, meso and endoderm) (Przyborski *et al,* 2004*)*. However, at difference with hESCs/hiPSCs, hEC lines are easy to culture, transfect very well, and can be easily maintained as pluripotent cells. Thus, Garcia-Perez *et al,* first analyzed L1 expression in a panel of pluripotent male and female hEC lines, and as in hESCs, they reported natural expression of L1 RNAs and of L1-RNPs. Consistently, pluripotent hECs are characterized by having hypomethylated genomes, including L1 promoters. Thus, LINE-1 elements are naturally expressed in pluripotent cells (PCs)(Garcia-Perez *et al,* 2007, 2010; Klawitter *et al,* 2016*;* Macia *et al,* 2011*;* Wissing *et al,* 2012*)*.

Next, Garcia-Perez *et al,* explored engineered L1 retrotransposition in the same panel of PCs. Previously, a low level of engineered L1 retrotransposition was reported in several hESC lines (Garcia-Perez *et al,* 2007*)*; the low level of retrotransposition reported in hESCs was likely related to how poorly these cells transfect. Surprisingly, despite transfecting at a much higher rate, Garcia-Perez *et al,* detected even lower retrotransposition rates in hECs when using the same engineered L1 mobility assay. These results suggested that additional mechanisms might operate in PCs to regulate retrotransposition. Indeed, using inhibitors of histone deacetylases (IHDACs), Garcia-Perez *et al,* demonstrated that new **L1 insertions accumulate at a high level in human PCs, but that *de novo* L1-insertions are strongly silenced through the action of histone deacetylases** (Garcia-Perez *et al,* 2007, 2010; Wissing *et al,* 2012*)*. In contrast, using isogenic differentiated cells (DCs), Garcia-Perez *et al,* demonstrated that silencing of new

L1 insertions is specific to PCs. Indeed, L1 silencing was also observed in other embryonic pluripotent models tested (hECs, hESCs, and hiPSCs). Thus, Garcia-Perez et al. discovered a **novel epigenetic mechanism, specific to PCs, involved in the silencing of *de novo* LINE-1 insertions** where chromatin modifying agents such as histone deacetylases seem to be involved (Garcia-Perez *et al*, 2010*)*. Further experiments allowed Garcia-Perez *et al*, to *co*nclude that:

- Silencing occurs during/shortly after insertion.
- The silencing process is independent of the sequence of the retrotransposon inserted.
- Silencing likely recognizes the TPRT reaction, and does not target viral integrations.
- The silencing response disappears rapidly when cell differentiation begins.

Despite this knowledge, **the mechanism of silencing**, both for initiation and maintenance, **or how the cell detects new L1 insertions are currently unknown**.

From a biological and functional angle, retrotransposon silencing in PCs reflects a mechanism that guarantees the integrity of our heritable genome, avoiding possible mutations due to overaccumulation of L1 insertions during embryogenesis. That is, in the absence of L1-silencing, new L1 insertions in PCs could produce secondary, tertiary, etc. insertions in an exponential manner, leading to an amplification of L1 that is possibly incompatible with life. Furthermore, this study revealed how L1 regulation is directly influenced by the pluripotent status of cells, and confirmed that hECs PCs are an excellent model to better understand L1 regulation in humans.

4.2. Control mechanisms against the L1-mRNA: Regulation at the transcriptional and post-transcriptional levels.

Research from several groups have uncovered different **mechanisms acting against the L1-mRNA** (Heras *et al*, 2014*)*, which might be **relevant when L1 evades transcriptional control**. Among these mechanisms, we find **Piwi proteins** and **piwi interacting small RNAs (piRNAs)**. Piwi proteins are highly conserved during evolution and are known to interact with piwi-interacting RNAs (piRNAs, 26-32 nt), which are RNAs transcribed from clusters of mobile elements. The interaction of Piwi proteins with piRNAs against L1 results in L1 RNA degradation. This silencing mechanism seems to operate in hiPSCs (Marchetto *et al*, 2013*)*, and likely in other PCs. It is worth noting that there is evidence of

interaction between piRNAs and DNMT3L, as the piRNA pathway seems to be acting by establishing methylation patterns for DNMT3L in mouse mobile elements (Aravin *et al,* 2008*)*. Thus, mechanism controlling L1 at the RNA level seem to interact with L1 epigenetic control mechanisms.

Beside piRNAs, another mechanism involving small RNAs occurs through the action of selected microRNAs (**miRNAs)**, specifically miRNAs miR-128 (Hamdorf *et al,* 2015*)* and Let-7 (Tristán-Ramos *et al,* 2020*)*. miRNAs are known to regulate many biological pathways and processes in cells. Additionally, some miRNA-related proteins, such as DROSHA/DGCR8 (Heras *et al,* 2013*)* and Dicer (Bodak *et al,* 2017*)*, have also been involved in targeting the LINE-1 RNA.

Other **factors regulating L1 retrotransposition that target the L1-mRNA** are the RNA helicase MOV10 (Moloney leukemia virus 10), a component of the RISC (RNA-induced silencing complex) complex, which inhibits retrovirus replication, and induces L1-mRNA degradation (Goodier *et al,* 2012*;* Li *et al,* 2013*)*. Notably MOV10 acts with TUT4/7 (Terminal uridyltransferases 4/7) to restrict LINE-1 retrotransposition, by uridylating the 3' end of L1-RNAs (Warkocki *et al,* 2018*)*.

To note, there are clear **similarities between regulation of LINE-1 retroelements and DNA/RNA viruses**. In relation with these viruses, mammals have evolved the **type I interferon (INF) response**, and this is the **main innate immune defense mechanism against viral nucleic acids**. There are numerous cell sensors able to recognize the typical characteristics of viral nucleic acids, triggering signaling pathways which ultimately would activate type I INFs and pro-inflammatory cytokine responses, leading to the activation of interferon-stimulated genes (ISGs), which stablish an antiviral cellular state (Schneider *et al*, 2014; Ivashkiv & Donlin, 2014; Gazquez-Gutierrez *et al*, 2021*)*. In this context, several mechanisms able to sense viral nucleic acids were also found to restrict mobile elements, being able to sense mobile element derived nucleic acids and triggering a type I INF response.

The INF-stimulated nucleic acid binding proteins APOBEC, ADAR1 and RNase L were found to restrict mobile elements as follows:

- The **adenosine deaminase acting on RNA 1 (ADAR1)** converts adenosines to inosines in dsRNA molecules, destabilizing dsRNA structures and preventing the activation of the type I INF response (Mannion *et al,* 2014*;* Orecchini *et al,* 2017*;* Rice *et al,* 2014*)*. Alteration or mutation of ADAR1 has been observed to be associated with the development of Aicardi Goutières Syndrome (AGS), a type I interferonopathy (Rice *et al,* 2012*)*.

- The **apolipoprotein B mRNA-editing enzyme, catalytic polypeptide-like 3 (APOBEC3)** is able to sense the presence of cytoplasmic DNA, which catalyzes the conversion of cytidines to uridines in RNA and ssDNA. Its activity is able to mutate the cDNA of retrotransposons (LINEs, SINEs and ERVs) as well as HIV-1 cDNAs (Bogerd, 2006; Muckenfuss *et al*, 2006; Richardson *et al*, 2014).

- **RNase L** is an endonuclease activated by dsRNA able to degrade cytoplasmic ssRNA, promoting the translational arrest of viral infected cells (Zhang *et al*, 2014; Burke *et al*, 2019). It is a component of the INF-stimulated oligoadenylate synthetase (OAS)/RNase L system, involved in the regulation of LINE-1 and IAP retrotransposition levels (Zhang *et al*, 2014).

Other proteins involved in type I INF response and in regulating mobile elements are TREX1, SAMHD1, the antiviral protein ZAP, the hnRNPL protein, MAVS, BST2, GAIT and condensing II.

- The **three-prime repair exonuclease 1 (TREX1)** is a 3'-5' DNA exonuclease that targets single-stranded and double-stranded DNAs (ssDNA and dsDNA). Its expression is controlled by INFs (ISG) and the loss of its function or its absence (dominant negative mutations) have been reported to be related with several autoimmune disorders as AGS or Systemic Lupus Erythematous (Crow *et al*, 2006a; Stetson *et al*, 2008; Lee-Kirsch *et al*, 2007; Thomas *et al*, 2017). TREX1 is an inhibitor of LINE-1s mobility, as in TREX1 knock-out cells LINE-1 retrotransposition increases, whereas it decreases upon TREX1 overexpression (Stetson *et al*, 2008).

- The **dNTPase SAM domain and HD domain-containing protein 1 (SAMHD1)** is a dNTP triphosphohydrolase triphosphatase which plays an important role in the regulation of the cytosolic deoxynucleotide pool (dNTPs), but may also have biologically relevant ribonuclease activity. Its activity reduces the cytosolic dNTPs pool, preventing viral replication. SAMHD1 regulates mobile elements and has also been associated with AGS. SAMHD1 is also a potent suppressor of LINE-1s mobility, as well as of other non-autonomous mobile retroelements such as Alu and SVA SINEs. Similar to TREX1, SAMHD1 knock-out cells exhibited increased retrotransposition of LINE-1, Alu and SVA, while overexpression of SAMHD1 resulted in decreased retrotransposition of LINE-1, Alu and SVA (Zhao et al, 2013).

- The **Heterogenous Nuclear Ribonucleoprotein L RNA binding protein (hnRNPL)** binds to the Internal Ribosome Entry Site (IRES) signal of mouse LINE-1s, acting as a retrotransposon restricting factor that is able to inhibit retrotransposition *in vitro* (Peddigari *et al*, 2013; Moldovan & Moran, 2015).

- Other proteins involved in the regulation of mobile elements are the Zinc Finger CCCH-type antiviral protein **(ZAP)** (Goodier *et al*, 2015; Moldovan & Moran, 2015), the Mitochondrial antiviral-signaling protein **(MAVS)** and the Bone Marrow Stromal Cell Antigen 2 **(BST2)**, all able to restrict retrotransposition *in vitro* (Goodier *et al*, 2015). The Gamma-Interferon Activated Inhibitor of Translation **(GAIT)** protein and condensing II have also been observed to restrict retrotransposition in epithelial cells, presumably by interfering with LINE-1 transcription and/or translation (Ward *et al*, 2017).

5. Aicardi Goutières Syndrome (AGS)

Aicardi-Goutières syndrome (AGS) is a rare human disorder connected with deregulation of LINE-1 retrotransposition. This rare genetic disorder mainly affects the brain and skin, being characterized, in its most severe form, by the presence of cerebral atrophy, leukodystrophy, intracranial calcifications, chronic cerebrospinal fluid (CSF) lymphocytosis, increased interferon alpha in CSF, and negative serological results for common prenatal infections (Ali *et al,* 2006*)*. Neurological dysfunction becomes clinically evident during infancy, manifesting as progressive microcephaly, spasticity, dystonic posture, profound psychomotor retardation, and often death in the early infancy. Outside the nervous system, those affected also suffer from thrombocytopenia, elevated liver transaminase levels, hepatosplenomegaly and intermittent fever, which may erroneously suggest the presence of an infectious process or autoimmune response (Crow *et al*, 2006b).

AGS exhibits a phenotype similar to that observed during *in utero* viral infections. A permanent increase in INF activity levels in CSF and serum (Lebon *et al,* 1988*)*, together with increased expression of ISGs in peripheral blood (Rice *et al,* 2013*)*, is also observed in AGS patients. On the basis of these observations, **AGS can be identified as an inflammatory disorder associated with the induction of a type I interferon-mediated innate immune response**, probably driven by nucleic acids of endogenous origin (Crow & Rehwinkel, 2009). However, **the nature of the nucleic acid that induces the immune response in these patients is currently unknown**.

In 2006, four genes associated with AGS were identified; **TREX1** (Crow *et al,* 2006a), **RNASEH2A, RNASEH2B, and RNASEH2C** (Crow *et al,* 2006b). Subsequently, mutations in three additional genes were also described in patients showing a phenotype compatible with AGS: **SAMHD1** (Rice *et al,* 2009*)*, **ADAR1** (Rice *et al,* 2012*)* and **IFIH1** (Rice *et al,* 2014*)*. Most of these proteins have also been found to be involved in the innate immune response to viral infections and endogenous mobile elements (Volkman & Stetson, 2014).

Interestingly, the alteration of just one of the aforementioned genes is capable of triggering AGS. Since no evidence of viral infections has been observed in AGS patients, the role of endogenous mobile elements becomes important for explaining the accumulation of these nucleic acids. For this reason, several studies were carried out to determine the role of AGS proteins in regulating the mobilization of endogenous mobile elements.

In relation with these studies, **TREX1** (Stetson *et al*, 2008), **SAMHD1** (Zhao *et al*, 2013) and **ADAR1** (Orecchini *et al*, 2017), cell factors involved in type I interferon response, have been found to **restrict LINE-1 retrotransposition**. Mutations in these genes could compromise their ability to restrict LINE-1 activity and be related with AGS development (Stetson *et al*, 2008). Within AGS related genes, **RNase H2 was suggested to also control LINE-1 retrotransposition (Volkman & Stetson, 2014), although this possibility has not been studied in deep until this Thesis**.

RNase H2 is a heterotrimeric complex formed by three subunits (**RNASEH2A, RNASEH2B and RNASEH2C** (Crow *et al*, 2006b)), and is the predominant nuclear enzyme responsible for the degradation of DNA:RNA hybrids (reviewed by Reijns & Jackson (2014)), although this might depend on the cell type. RNase H2 is essential for genome stability, and in addition to degrading the RNA of DNA:RNA hybrids, RNase H2 is also able of cleaving the 5'-phosphodiester bond of miss incorporated ribonucleotides embedded in dsDNAs (McElhinny *et al*, 2010; Hiller *et al*, 2012; Reijns *et al*, 2012), participating in a process called Ribonucleotide Excision Repair (RER) (Sparks *et al*, 2012).

The catalytic core of RNase H2 is present in the RNASEH2A subunit; however, all three subunits are required for its enzymatic activity (Reijns & Jackson, 2014). The precise role of the RNASEH2B subunit is not known; however, the **RNASEH2B subunit features a functional PIP motif that interacts with PCNA** (Chon *et al*, 2009). In fact, **PCNA directs RNase H2 activity to replication and repair foci** (Bubeck *et al*, 2011; Kind *et al*, 2014).

Due to its function, RNase H2 may play a regulatory role in LINE-1 retrotransposition, and this was a main objective in my Thesis.

6. The LINE-1 interactome

Although LINE-1s are classified as autonomous mobile elements, **L1 activity is influenced and regulated by the combined action of several cellular host factors that act to promote** (i.e., activators) **or repress** (i.e., repressors) **L1 expression and retrotransposition**. Collectively, the pool of cellular host factors acting to activate or repress L1 expression/retrotransposition is known as the **"LINE-1 interactome"**.

Through the use of ultracentrifugation, immunofluorescence (IF), immunoprecipitation (IP) and cell culture assays, it has been determined that the basal L1-RNP contains the L1 mRNA, L1-ORF1p, L1-ORF2p, and a number of interacting factors (Wei et al, 2001; Goodier *et al*, 2010; Doucet *et al*, 2010). These studies provided critical insight into the composition of L1-RNP complexes, and determined that **L1-RNPs are the essential unit for retrotransposition**. Thus, cellular **factors that modulate retrotransposition might probably be associated with the basal L1-RNP**. However, which cellular factors are involved in the retrotransposition process, in the formation and transport of L1-RNPs, or in the regulation of L1 expression, are largely unknown.

In the last 20 years, the improvements in protein purification and mass spectrometry have led to several projects focused on studying and characterizing the LINE-1 interactome (Goodier *et al*, 2012, 2013; Mandal *et al*, 2013; Taylor *et al*, 2013). Below, I will discuss major findings in this area of research.

6.1. Proteomic study of the LINE-1 interactome

To better understand the L1 interactome, several attempts to co-immunoprecipitate L1-RNPs, together with its interactors, have been made. These approaches used established cellular lines, the majority of them transformed cell lines, and often using overexpression of engineered L1s (see section *2.2.1. A system to study retrotransposition: The retrotransposition assay*). The cell lines that have been routinely used for L1 research (such as HeLa, 143B, HEK293T, U2OS...) are all transformed cell lines, and might reproduce a cellular state similar to that found in human cancer (Goodier *et al*, 2012, 2013; Mandal *et al*, 2013; Taylor *et al*, 2013). As discussed above, under physiological conditions L1 is expressed and mobilize during embryogenesis and in brain, while L1 expression is effectively repressed in most somatic cells (Muotri *et al*, 2005). Thus, **a major limitation of the majority of L1- interactome studies conducted so far is the use of non-physiological transformed cellular models**.

Beside limitations imposed by their transformed status, it has been demonstrated that the expression of endogenous L1-RNPs is highly variable in these cell lines, and L1 expression levels are inversely correlated with genome methylation levels (Ergün *et al*, 2004; Dai *et al*, 2014; Faulkner *et al*, 2009). Thus, to study L1 activity and the L1-interactome in cell lines, most studies have used engineered L1 constructs where a strong exogenous viral promoter is used to achieve a high-expression level of L1 RNAs and encoded proteins (Moran *et al, 1996)*. This is **another major limitation of most L1-interactome studies, as overexpression of proteins can lead to stoichiometric imbalance in complexes, the appearance of promiscuous interactions, or the activation of viral restriction pathways (type I INF and others) that will likely influence the deduced L1-interactome.**

In sum, and while information gained using transformed cell lines and overexpression L1 vectors has proven useful in the past, it is very likely that the composition of the L1-interactome and L1 regulation in physiological cellular models will be fundamentally different.

Pioneer studies aimed to characterize the L1 interactome used sucrose cushion velocity sedimentation of cell lysates to concentrate L1-RNPs (Kulpa & Moran, 2005, 2006b). This approach provided valuable information about the mechanism of L1 retrotransposition and the role of L1-ORF1p, but suffered from low purity, which limited its usefulness for determining the L1 interactome. To solve these limitations, following studies used engineered L1 constructs containing different epitope tags in L1-ORF1p, L1-ORF2p or both (V5, HA, FLAG, T7, etc; Goodier *et al*, 2013*;* Mandal *et al,* 2013*;* Taylor *et al,* 2013*)*. The use of these constructs has made possible to co-immunoprecipitate L1-ORF1p or L1-ORF2p, together with their interactors, to subsequently identify proteins associated with L1 ORF1p and/or L1 ORF2p by mass spectrometry. This approach has identified many interactors, including RNA transport proteins, regulators of gene expression, post-translational modifiers, helicases and splicing factors. Below, I will discuss seminal proteomic studies on L1-RNPs:

- Goodier *et al, (*2013) used overexpression vectors where L1-ORF1p was epitope-tagged (FLAG-tagged), allowing authors to use immunoprecipitation to purify L1-ORF1p and its associated interactors [using anti-FLAG-M2 affinity gel (Sigma) and transformed human cells (HEK293T, 2102Ep, HeLa)], which were identified by mass spectrometry (MS). These experiments found important inhibitory proteins, including some controlling HIV and other retroviruses (Goodier *et al,* 2013*)*. In fact, Goodier *et al,* validated that the helicase MOV10 interacts with L1-RNPs and can inhibit L1 retrotransposition

(Goodier *et al*, 2012*)*. Thus, this study (Goodier *et al*, 2013*)* represents the first detailed screening of proteins interacting with L1-ORF1p (i.e., the L1-ORF1p interactome) in the context of L1-RNPs. The interaction with identified proteins was confirmed by direct co-immunoprecipitation and subcellular co-localization in cytoplasmic granules. Furthermore, by overexpressing many of these proteins, they confirmed that many acted as repressors of L1 activity, as most reduced L1 retrotransposition in cell culture (Table 2). These authors also described how L1-ORF1p co-localized with proteins present in stress granules, such as LARP1, PCBP2, HNRNPA1 or PABPC1.

The results obtained in this pioneer study paved the way for future research on the life cycle of L1, from transcription to insertion in the genome, and of repressive strategies used by cells to coexist with L1. The utility of the approach developed by Goodier and colleagues was explored on a second study from the same lab (Mandal *et al*, 2013*)*. However, even when different interactors were found (Goodier *et al*, 2013*;* Mandal *et al*, 2013*)*, technical limitations prevented the isolation of pure and active L1-RNP complexes in analytically manipulable quantities, and this is a major limitation of the approach developed by Goodier and colleagues. In fact, while traditional L1-ORF1p tagging using epitopes (HA, FLAG, V5, T7, etc) has been used to study proteins that co-purify with L1-RNPs (Doucet *et al*, 2010; Goodier *et al*, 2012, 2013), the use of standard agarose beads (anti-FLAG M2 affinity gel (Sigma)) for co-immunoprecipitation (Goodier et al. 2013 Mandal *et al*, 2013*)* provides a higher background signal due to non-specific binding of lysate components to the beads and results in an insufficient yield and purity to obtain integral proteomics and proteins with high specific activity.

In addition (as indicated in the section on *III. INTRODUCTION, 2.2*.), the translation of L1-ORF1p is highly efficient, whereas L1-ORF2p is translated by an unconventional system (Alisch, 2006) resulting in low levels of L1-ORF2p expression, making mechanistic investigation of L1-ORF2p very challenging.

To solve these limitations, the development of new methods incorporating cryomilling, together with rapid affinity capture, allowed achieving an excellent purification, significantly improving past L1-interactome studies (Cristea *et al*, 2005*;* Domanski *et al*, 2012*;* Oeffinger *et al*, 2007*)*. Briefly, using these new methods, intact cells are flash frozen in liquid nitrogen and cryomilled in solid phase *(VI. MATERIALS AND METHODS, 17. Cryomilling of frozen cells)*. In doing that, the native intermolecular interactions are preserved during milling, producing a very fine powder suitable for subsequent rapid affinity capture of protein complexes. For such affinity capture, magnetic beads conjugated with antibodies are used. Magnetic beads exhibit a faster protein binding rate than agarose beads, are

significantly smaller (1-4 μm versus 50-150 μm agarose beads), and are better at detecting weak (labile) interactors of immunoprecipitated proteins, since the physical stress to which agarose beads are subjected during successive rounds of centrifugation can cause loss of interactions (Domanski et al, 2012; Trinkle-Mulcahy *et al*, 2008). In consequence, the use of magnetic beads allows rapid purification with binding times as short as 5 mins, and facilitates binding of large complexes that would be excluded by the pores of traditional matrices (Oeffinger *et al*, 2007) (*VI. MATERIALS AND METHODS, 19. Affinity capture of protein complexes from cell powder)*. To note, **this improved methodology has been used in the proteomic studies carried out in this Thesis**.

- In 2013, Taylor *et al*, conducted a large-scale proteomic study of the L1 interactome, combining cryomilling and rapid affinity capture with isotopic differentiation of interactions as random or directed (I-DIRT), allowing them to distinguish stable interactions formed *in vivo* from post-lysis artifacts. Furthermore, Taylor and colleagues created a system to overexpress and purify active L1-RNP complexes from suspension cultures of human HEK293T$_{LD}$ cells, in an attempt to solve low levels of L1 expression in cultured cells. HEK293T$_{LD}$ (Dai *et al*, 2012) is an inducible cell line using the Tet-On system, a system that allows the activation and overexpression of those vectors regulated by a specific promoter (TRE-Tight element). In addition, HEK293T cells express the SV40 virus T antigen (Lin *et al*, 2014), which allows achieving a high copy number of vectors containing the SV40 ori, resulting in higher levels of protein expression in cells transiently transfected with plasmids. For these reasons, the HEK293T cell line was used for protein overexpression using expression vectors, and the fact that HEK293T$_{LD}$ cells have an integrated inducible Tet-On expression system allows for further increased expression levels. Taylor and colleagues constructed a series of inducible L1-expression vectors to be expressed in HEK293T$_{LD}$ cells, employing both codon-optimized L1 elements (ORFeus-Hs) (An *et al*, 2011) and its native parental L1 (L1RP) (Kimberland *et al*, 1999), both containing epitope tags in L1-ORF1p and L1-ORF2p.

Upon overexpressing and purifying engineered L1-RNPs in HEK293T$_{LD}$ cells, Taylor and colleagues used an *in vitro* L1 RT assay, termed LEAP for LINE-1 Element Amplification Protocol, to analyze the activity of purified L1-RNPs; briefly, the LEAP assay test the ability of purified L1-RNPs to reverse transcribe the L1-mRNA *in vitro* (Kulpa & Moran, 2006a; Huira C. Kopera Doucet & Moran, 2016). Remarkably, Taylor and colleagues reported significant higher specific activity of L1-RNPs (70-fold higher) when compared to L1-RNPs obtained by sucrose cushion. Consistently, RNA sequencing (RNA-

seq) revealed that L1-mRNAs are abundant components of purified L1-RNPs (Taylor *et al,* 2013*)*. Using this improved purification method, Taylor and colleagues characterized 37 high-confidence interactors, some already known such as PABPC1 and MOV10, and novel interactors such as UPF1 and PCNA.

The interaction with novel interactors was confirmed using co-IP and retrotransposition assays. In fact, these authors validated that UPF1, a key factor of the non-sense mediated decay pathway, is an L1-ORF1p interactor, which reduces L1 RNA levels and retrotransposition. Similarly, Taylor *et al,* validated that PCNA interacts with L1-ORF2p through a PIP-box motif located between the EN and RT domains, and studies on the mechanism of regulation suggested that the interaction with PCNA occurs during or immediately after TPRT. Interestingly, the results of this study also suggested the existence of at least three types of L1-RNPs, which differs in composition and functionality. Undeniable, this study represents a high-quality proteomic study of the L1-interactome, and for L1-ORF1p and L1-ORF2p, even if authors used transformed cell lines and L1 overexpression vectors.

In Goodier *et al.* 2013 only FLAG-tagged L1-ORF1p was captured, while in Taylor *et al.* 2013 expression vectors with both L1-ORF1p and L1-ORF2p tagging were used, and represents the first proteomic study where the interactome of L1-ORF2p was studied by immunoprecipitation. Goodier *et al,* 2013 found a total of 96 L1-interactors, of which 69 were identified in Taylor *et al,* 2013, but only 7 were found as significant according to this study: MOV10, MEPCE, PABPC1, PABPC4, PURA, TOP1, and TROVE2 (**Table 2**).

Regarding the L1-ORF2p interactome, Goodier *et al,* 2013 were able to co-immunoprecipitate L1-ORF2p together with L1-ORF1p in their experiments; however, among all the factors found in Goodier et al. and subsequently detected again in Taylor *et al,* 2013, only 5 were identified as significant interactors that co-immunoprecipitated with L1-ORF2p: PABPC1, PABPC4, PURA, TOP1, and MOV10. Thus, few L1-ORF2p-specific interactors are known.

In addition, Goodier *et al,* 2013 analyzed the effect of the overexpression of 69 factors on LINE-1 retrotransposition, and found 23 factors that reduced LINE-1 retrotransposition levels between 50 and 100% (**Table 2**). Some of these factors were identified in Taylor *et al,* 2013 although not as significant (shown in **Table 2**).

Table 2. Significant proteins found in Goodier *el at*, 2013 and Taylor *et al, 2013.*

	Gene names	Protein names	Nuclear (Nu)/cytoplasmic (C) extracts (Goodier *et al,* 2013*)*	Co-IPs with L1-ORF1p (Taylor *et al,* 2013*)*	Co-IPs with L1-ORF2p (Taylor *et al,* 2013*)*	Found using L1RP or ORFeus-Hs	L1 + cDNA RTSN (%, see Goodier *et al,* 2013)
1	CSDA	Cold shock domain protein A	Nu/C	Yes	Yes	Both	~ 0
2	DDX39A	DEAD (Asp-Glu-Ala-Asp) box polypeptide 39	Nu	Yes	Yes	Both	~ 15
3	ELAVL1	Embryonic lethal abnormal vision n, Drosophila like 1/Hu antigen R	Nu	Yes	Yes	Both	~ 10
4	HIST1H1C	Histone cluster 1, H1e	Nu	Yes	Yes	Both	~ 30
5	HNRNPA1	Heterogeneous nuclear ribonucleoprotein A1	Nu	Yes	Yes	Both	~ 25
6	HNRNPL	Heterogeneous nuclear ribonucleoprotein L	Nu	Yes	Yes	Both	~ 5
7	HNRNPU	Heterogeneous nuclear ribonucleoprotein U (scaffold attachment factor A)	Nu	Yes	Yes	Both	~ 20
8	ILF2	Interleukin enhancer-binding factor 2	Nu	Yes	Yes	Both	~ 35
9	MEPCE	7SK snRNA methylphosphate capping enzyme	C	Yes*	Yes*	Both*	~ 60
10	MOV10	Helicase MOV-10 (Moloney leukemia virus 10)	Nu/C	Yes*	Yes*	Both*	~ 0
11	PABPC1	Poly(A)-binding protein, cytoplasmic 1	Nu/C	Yes*	Yes*	Both*	~45
12	PABPC4	Poly(A)-binding protein, cytoplasmic 4	Nu/C	Yes*	Yes*	Both*	Not tested
13	PCBP2	Poly(rC)-binding protein 2	Nu/C	Yes	Yes	Both	~ 20
14	PURA	Purine-rich element-binding protein A	Nu	Yes (L1RP)	Yes* (ORFeus-Hs)	ORFeus-Hs*	< 5
15	RALY	RNA-binding protein, autoantigenic (hnRNP-associated with lethal yellow homolog (mouse))	Nu	Yes	Yes	Both	~ 20
16	RBMX	RNA-binding motif protein, X-linked	Nu	Yes	No	ORFeus-Hs	~ 40

Table 2. Significant proteins found in Goodier *et al,* 2013 and Taylor *et al, 2013.*

	Gene names	Protein names	Nuclear (Nu)/cytoplasmic (C) extracts (Goodier *et al*, 2013*)*	Co-IPs with L1-ORF1p (Taylor *et al*, 2013*)*	Co-IPs with L1-ORF2p (Taylor *et al*, 2013*)*	Found using L1RP or ORFeus-Hs	L1 + cDNA RTSN (%, see Goodier *et al*, 2013)
17	SNRNP70	U1 Small nuclear ribonucleoprotein 70 kDa	C	Yes	Yes	Both	~ 20
18	TOP1	DNA topoisomerase I	Nu	No	Yes*	ORFeus-Hs*	Not tested
19	TRA2B	Transformer 2 b homolog (Drosophila)	Nu	Yes	Yes	Both	~ 25
20	TROVE2	60kDa Ro protein, Sjogren syndrome antigen A2	Nu/C	Yes*	Yes (ORFeus-Hs)	Both*	~ 75
21	YBX1	Y box-binding protein 1	Nu/C	Yes	Yes	Both	~ 15

Column 4 (Nuclear (Nu)/cytoplasmic (C) extracts) indicates whether the proteins were found in the cytosolic or nuclear fraction from the cell lysates in Goodier et al, 2013. Column 5 (Co-IPs with L1-ORF1p), proteins co-immunoprecipitated when capturing L1-ORF1p. Column 6 /Co-IPs with L1-ORF2p), proteins co-immunoprecipitation when capturing L1-ORF2p. Column 7 (Found using L1RP or ORFeus-HS), indicates whether the proteins were found when using codon-optimized L1 elements (ORFeus-Hs) (An et al, 2011) and its native parental L1 (L1RP) (Kimberland et al, 1999). Asterisks indicate that the factor was significant in that condition. "Yes" in columns 5–6 indicates that the protein was identified in the study, but did not reach the significance threshold established by the authors. "Yes" in columns 5-6 means that the protein was identified as a significant L1-interacting protein by statistical methods in Taylor et al, 2013. "No" means that the protein was not identified in that condition.*

- In 2015, a study by Moldovan and Moran used a method similar to that employed by Goodier *et al*, in 2013, that is using conventional purification techniques, and identified 39 cellular proteins as L1-ORF1p interactors (**Table 3**). Moldovan and Moran used transformed HeLa cells and engineered L1 vectors where L1-ORF1p was tagged with a FLAG epitope-tag (in its C-terminus, pJM101/L1.3FLAG). After transfection, L1-RNPs were isolated by immunoprecipitation from standard whole cell lysates, followed by liquid chromatography coupled to mass spectrometry (LC MS/MS). Of the 39 interactors, 13 were validated by biochemical assays, showing an RNA dependent interaction with L1-ORF1p, via an RNA bridge (**Table 3**, noted as "Yes*"). Notably, most proteins interacting with L1-ORF1p were previously found in previous proteomic studies (33 out of 39, see **Table 3**) (Taylor *et al*, 2013; Peddigari *et al*, 2013; Goodier *et al*, 2013). Moldovan and Moran analyzed in detail a novel L1-ORF1p interactor, ZAP, and demonstrated that ZAP could inhibit L1 and Alu retrotransposition (as discussed in previous section *4. Mechanisms regulating LINE-1 retrotransposition*).

- Building from their previous study, in 2018, Taylor *et al,* attempted to further characterize the composition of the different types of L1-RNPs described in their 2013 study (Taylor *et al*, 2013). They again used HEK293T$_{LD}$ cells (Dai *et al*, 2012; Taylor *et al*, 2013) expressing allelic L1 constructs containing LINE-1 mutations in conserved domains, to identify how these domains affect the L1-interactome. Using differential affinity purifications, quantitative mass spectrometry, and RNA-seq, Taylor and colleagues characterized proteins and nucleic acids associated with enzymatically active L1-RNP macromolecular complexes (**Table 3**). These results revealed that LINE-1 can indeed form at least two different macromolecular complexes, containing different host factors. Notably, a canonical L1-RNP complex containing L1-ORF1p, L1-ORF2p and the L1-mRNA, together with several other proteins (MOV10, ZCCHC3, PABPC1, PABPC4), was found in the cytoplasm, and authors hypothesized that it will likely correspond to **"a canonical L1-RNP".** In addition, they also identified a novel macromolecular complex in the nucleus, containing L1-ORF2p, the L1-mRNA, and proteins involved in DNA replication and DNA repair (PCNA, PURA/B, TOP1 and PARP1), but lacking L1-ORF1p. Authors hypothesized that this second complex likely correspond to **"TPRT complexes",** allowing to identify factors that specifically might regulate TPRT. Other factors, like UPF1, were found interacting with both complexes (see **Table 3**, column 9).

Multiple host genome-encoded proliferation and restriction factors have been identified and validated from these studies, but it is notable that little overlap has been observed between the L1-linked proteins described by the different studies. However, although this is likely to be largely explained by technical differences in experimental approaches, it raises questions regarding the uniformity of L1 interactomes and the potential for context specificity, both technical and biological. As indicated above, all proteomic studies discussed above exploited transformed cell lines and exogenously expressed L1s, and it is currently unknown whether endogenous L1-RNPs might interact with the same proteins in non-transformed cell lines.

- Importantly, complementing the above studies, Vuong *et al, (*2019) used affinity capture and MS techniques to examine the interactome of endogenous L1-RNPs in non-transformed cells, using hESCs, which as described above, naturally overexpress L1-RNPs. Remarkably, this study identified previously validated L1-ORF1p interactors, although for the first-time reported interactions with endogenous L1-RNP complexes. Thus, this study is the first addressing an important question in L1 biology: is the LINE-1 interactome different in cell types where L1 mobilize under physiological conditions? However, a major limitation of this study is that none of the previously mentioned advanced proteomic techniques were employed (cryomilling, I-DIRT, etc). Thus, although relevant, further research is needed to define the L1-interactome in PCs, and to learn whether L1 is regulated in a different manner depending on the pluripotent status of cells. In fact, a major goal of this Thesis is to analyze the L1-interactome of PCs, using advanced proteomic techniques, and to compare the interactome of isogenic PCs and DCs.

Table 3. Common identified proteins among different studies

Gene names	(Goodier *et al*, 2013)	(Taylor *et al*, 2013)	(Moldovan and Moran, 2015)	(Taylor *et al*, 2018)	(Vuong *et al*, 2019)	Co-IP with L1-ORF1p/L1-ORF2p (Taylor *et al*, 2013)	Co-IP with L1-ORF1p/L1-ORF2p (Taylor *et al*, 2018)	(Liu *et al*, 2018)	(Mita *et al*, 2020)
MOV10	Yes*	Yes*	Yes*	Yes*	-	Both*	Both*	Suppressor	Middle
PURA	Yes*	Yes*	Yes*	Yes*	Yes*	Both (L1-ORF2p*)	L1-ORF2p*	Middle	Activator
UPF1	No	Yes*	Yes*	Yes*	-	Both	Both*	Activators	-
DHX9	Yes*	Yes	Yes*	No	-	Both	-	Activators	Middle
FAM120A	Yes*	Yes	Yes*	No	-	L1-ORF1p	-	-	Middle
hnRNPL	Yes*	Yes	Yes*	No	-	Both	-	-	-
IGF2BP3	Yes*	Yes	Yes*	No	-	Both	-	-	Activator
LARP1	Yes*	Yes	Yes*	No	-	Both	-	-	Activator
MATR3	Yes*	Yes	Yes*	No	-	Both	-	-	Middle
NCL	Yes*	Yes	Yes*	No	-	Both	-	-	Activator
ZAP	No	Yes	Yes*	No	-	L1-ORF1p	-	-	-
IFL3	Yes*	No	Yes*	No	-	-	-	-	-
CDK9	No	No	Yes*	No	-	-	-	-	Middle
TROVE2	Yes*	Yes*	Yes	Yes*	-	L1-ORF1p*	L1-ORF1p*	-	Middle
DDX21	Yes*	Yes	Yes	No	-	Both	-	-	Middle
hnRNPA2B1	Yes*	Yes	Yes	No	-	Both	-	-	Activator
hnRNPC	Yes*	Yes	Yes	No	-	Both	-	-	Middle
IGF2BP2	Yes*	Yes	Yes	No	-	L1-ORF1p	-	-	Middle
SFRS1	Yes*	Yes	Yes	No	-	Both	-	Activators	Middle

Gene names	(Goodier *et al*, 2013*)*	(Taylor *et al*, 2013*)*	(Moldovan and Moran, 2015)	(Taylor *et al*, 2018*)*	(Vuong *et al*, 2019)	Co-IP with L1-ORF1p/L1-ORF2p (Taylor *et al*, 2013*)*	Co-IP with L1-ORF1p/L1-ORF2p (Taylor *et al*, 2018*)*	(Liu *et al*, 2017*)*	(Mita *et al*, 2020*)*
ATXN2	No	Yes	Yes	No	-	Both	-	-	Middle
DHX15	No	Yes	Yes	No	-	Both	-	-	Activator
FUBP3	No	Yes	Yes	No	-	Both	-	-	Middle
GNB2L1	No	Yes	Yes	No	-	Both	-	-	Middle
HIST1H1D	No	Yes	Yes	No	-	Both	-	-	Middle
hnRPDL	No	Yes	Yes	No	-	L1-ORF1p	-	-	Middle
KHSRP	No	Yes	Yes	No	-	Both	-	-	Middle
KIF5B	No	Yes	Yes	No	-	-	-	-	Middle
NCBP1	No	Yes	Yes	No	-	Both	-	-	Middle
NCBP2	No	Yes	Yes	No	-	L1-ORF1p	-	-	Activator
PAR-4	No	Yes	Yes	No	-	-	-	-	-
SFRS7	No	Yes	Yes	No	-	Both	-	-	Middle
SRRM2	No	Yes	Yes	No	-	L1-ORF1p	-	-	-
SYNCRIP	Yes*	No	Yes	No	-	-	-	Activators	Activator
ANP32A	No	No	Yes	No	-	-	-	-	Middle
C21orf70	No	No	Yes	No	-	-	-	-	-
hnRPA0	No	No	Yes	No	-	-	-	-	-
LARP5	No	No	Yes	No	-	-	-	-	-
MPO	No	No	Yes	No	-	Both	-	-	Middle
USP10	No	No	Yes	No	-	-	-	-	Middle
MEPCE	Yes*	Yes*	No	Yes*	-	L1-ORF1p	L1-ORF1p*	-	-

Gene names	(Goodier *et al*, 2013)	(Taylor *et al*, 2013)	(Moldovan and Moran, 2015)	(Taylor *et al*, 2018)	(Vuong *et al*, 2019)	Co-IP with L1-ORF1p/L1-ORF2p (Taylor *et al*, 2013)	Co-IP with L1-ORF1p/L1-ORF2p (Taylor *et al*, 2018)	(Liu *et al*, 2017)	(Mita *et al*, 2020)
PABPC4	Yes*	Yes*	No	Yes*	-	Both*	Both*	-	Middle
TOP1	Yes*	Yes*	No	Yes*	-	Both (L1-ORF2p*)	L1-ORF2p*	-	Middle
DDX6	No	Yes*	No	Yes*	-	L1-ORF1p*	L1-ORF1p*	-	Middle
HIST1H2BO	No	Yes*	No	Yes*	-	Both*	L1-ORF1p*	-	Activator
HSPA1A	No	Yes*	No	Yes*	-	Both	-	-	Middle
HSPA8	No	Yes*	No	Yes*	-	Both (L1-ORF2p*)	L1-ORF2p*	-	-
PCNA	No	Yes*	No	Yes*	-	L1-ORF2p*	L1-ORF2p*	Activators	-
PURB	No	Yes*	No	Yes*	-	Both*	L1-ORF2p*	Middle	Middle
RPS27A	No	Yes*	No	Yes*	-	Both	L1-ORF2p*	-	-
YME1L	No	Yes*	No	Yes*	-	L1-ORF2p*	L1-ORF2p*	-	Middle
ZCCHC3	No	Yes*	No	Yes*	-	Both*	Both*	-	Middle
TUBB	No	Yes	No	Yes*	-	Both	L1-ORF2p*	-	-
PABPC1	Yes*	No	No	Yes*	-	-	Both*	Activators	Activator
CORO1B	No	No	No	Yes*	-	-	L1-ORF1p*	-	Middle
ERAL1	No	No	No	Yes*	-	-	L1-ORF1p*	-	Middle
FKBP4	No	No	No	Yes*	-	-	L1-ORF2p*	-	Middle
HAX1	No	No	No	Yes*	-	-	L1-ORF2p*	-	Middle
HMCES	No	No	No	Yes*	-	-	L1-ORF2p*	-	-
HSP90AA1	No	No	No	Yes*	-	-	L1-ORF2p*	-	-
IPO7	No	No	No	Yes*	-	-	L1-ORF2p*	-	Middle
LARP7	No	No	No	Yes*	-	-	L1-ORF1p*	-	Middle

Gene names	(Goodier *et al*, 2013*)*	(Taylor *et al*, 2013*)*	(Moldovan and Moran, 2015)	(Taylor *et al*, 2018*)*	(Vuong *et al*, 2019)	Co-IP with L1-ORF1p/L1-ORF2p (Taylor *et al*, 2013*)*	Co-IP with L1-ORF1p/L1-ORF2p (Taylor *et al*, 2018*)*	(Liu *et al*, 2017*)*	(Mita *et al*, 2020*)*
NAP1L1	No	No	No	Yes*	-	-	L1-ORF2p*	-	Middle
PABPC4L	No	No	No	Yes*	-	-	L1-ORF1p*	-	Middle
PARP1	No	No	No	Yes*	-	-	L1-ORF2p*	Middle	Middle
TIMM13	No	No	No	Yes*	-	-	L1-ORF2p*	-	Activator
TOMM40	No	No	No	Yes*	-	-	L1-ORF2p*	-	-
TUBB4B	No	No	No	Yes*	-	-	L1-ORF2p*	-	-
YARS2	No	No	No	Yes*	-	-	L1-ORF1p*	-	Middle

Columns 3-7, proteins co-immunoprecipitated in the indicated study. Column 8 (Co-IP with L1-ORF1p/L1-ORF2p, Taylor et al, 2013), proteins co-immunoprecipitated when capturing L1-ORF1p, L1-ORF2p or in both conditions. Asterisks indicate that the factor was significant in that condition. "Yes" in columns 3–7 indicates that the protein was identified in the study, but did not reach the significance threshold established by the authors. "Yes" in columns 3-7 means that the protein was identified as a significant L1-interacting protein by statistical methods in Taylor et al, 2013. "No" means that the protein was not identified in that condition. Columns 10-11, factors found as activators (increase L1 retrotransposition), middle (doesn't affect) or suppressors (inhibit L1 retrotransposition).*

6.2. Genomic studies of the LINE-1 interactome

Just as proteomic approaches have been used to better understand how L1 activity is regulated, the development of small interfering RNA (siRNA) and CRISPR-Cas9 genome editing systems have allowed researchers to analyze how L1 expression/activity is regulated on a genome wide manner.

In 2018, Liu *et al,* exploited CRISPR-Cas9 screening strategies in two different human cell lines, both transformed, to identify genes regulating L1 retrotransposition. This study identified many genes involved in several biological pathways that either promote or restrict engineered L1 retrotransposition (**Table 3**). These genes, some linked to human diseases, were able to regulate L1 at the transcriptional or post-transcriptional levels. Notably, Liu *et al.* exploited native/natural RC-L1s, but also codon optimized RC-L1s, revealing that some genes were able to regulate retrotransposition in a sequence dependent manner, highlighting the complexity of L1 regulation.

Among identified genes, the repressor of L1 retrotransposition MORC2 and the human silencing complex (HUSH) subunits, MPP8 and TASOR8, were investigated in depth. Notably, Liu and colleagues demonstrated that HUSH and MORC2 could selectively bind full-length, evolutionarily young L1s (i.e., L1Hs). These young L1s tend to be located within transcriptionally euchromatic environments but were silenced by trimethylation of histone H3 Lys9 (H3K9me3), promoted by HUSH and MORC2. In fact, it was further demonstrated that the expression of genes transcriptionally active but containing silenced L1s in their introns was influenced by the repressive marks deposited by HUSH and MORC2 in these intronic L1s, uncovering a new mechanism to fine-tune gene expression, mediated by the accumulation of young L1s in introns. In summary, beside elucidating a novel restriction pathway for L1 and illustrating how epigenetic silencing of L1 influences host gene expression, this study represents one of the larger host-factor study conducted to date, and has become an important resource for L1 researchers.

Later, in a study conducted by Mita and colleagues (2020) studying the role of the BRCA1 gene in relation with L1, authors carried out a whole genome siRNA screening using image based retrotransposition assays in HeLa-M2 cells. In this study, several cellular factors regulating L1 retrotransposition were reported: 220 "inhibitors" and 2,681 "supporters". Cluster analysis of inhibitors identified "Fanconi anemia pathway" and "DNA repair" as the most enriched clusters of proteins. Among the identified factors, BRCA1 was analyzed in detail, revealing how it affects the frequency of retrotransposition and the structure of L1,

controlling the translation of L1-ORF2p by binding to L1-mRNA. This study unveiled the role of BRCA1 and shed new light on how TPRT occurs, where BRCA1 and its associated proteins may block the generation of a second upstream cleavage by the EN L1-ORF2p activity. However, more research is needed to confirm this hypothesis.

IV. HYPOTHESIS

Theoretical background supporting the hypothesis

The following are key findings in the field of L1 biology relevant to this Thesis, which have been discussed in the introduction section *(III. INTRODUCTION):*

- The majority of **new and heritable L1 retrotransposition events in humans accumulate during early stages of human embryonic development** *(3. LINE-1 retrotransposition in pluripotent cells)* (Van den Hurk et al, 2007; Garcia-Perez *et al*, 2007, 2010; Klawitter *et al*, 2016; Kano *et al*, 2009; Wissing *et al*, 2012).

- **Epigenetic silencing of L1 expression, by methylation of CpG islands located in L1 promoters, is the main mechanism used to regulate LINE-1 expression** *(4.1.1. DNA methylation of the L1 promoter)* (Thayer *et al*, 1993; Bourc'his & Bestor, 2004; Coufal *et al*, 2009). However, L1 seems to take advantage of the genome wide hypomethylation that occurs shortly after fertilization, during early human embryonic development (Zeng & Chen, 2019), to accumulate *de novo* insertions that could pass to the next generation.

- A **Y-like structure containing a L1-cDNA:L1-mRNA hybrid covalently linked to genomic DNA is generated during TPRT** *(2.2. The LINE-1 retrotransposition mechanism)*. Notably, it has been observed that the formation of cDNA:RNA hybrids on CpG islands can prevent methylation of the underlying DNA sequence.

- A **new epigenetic mechanism that mediates silencing of *de novo* LINE-1 insertions**, in a sequence independent manner and likely targeting TPRT intermediates, **operates specifically in PCs**. Chromatin modifying factors such as histone deacetylases are involved in the maintenance of L1 silencing in PCs (Garcia-Perez *et al*, 2007, 2010; Wissing *et al*, 2012*) (4.1.2. Histone deacetylation as a possible L1 restriction mechanism).* L1-silencing is rapidly attenuated upon induction of cellular differentiation. However, **how PCs recognize *de novo* L1 insertions and the mechanism of silencing are unknown** (Garcia-Perez *et al*, 2010*)*.

- Previous **L1 interactome studies, conducted in transformed cell lines** which to some level are more similar to somatic differentiated cells, **and using overexpressed epitope-tagged L1 proteins** (Goodier *et al*, 2013; Mandal *et al*, 2013; Taylor *et al*, 2013; Moldovan & Moran, 2015; Taylor *et al*, 2018), have **provided an ample overview of host factors interacting with L1-RNPs in this cellular context**.

- A single proteomic study of L1-interactors in hESCs identified common interactors with transformed cell lines (Vuong *et al*, 2019*) (6.1. Proteomic study of the LINE-1 interactome).*

- Several **factors associated with type I interferon response** [SAMHD1 (Zhao *et al*, 2013*)*, ADAR1 (Orecchini *et al*, 2017*)*, and TREX1 (Stetson *et al*, 2008*)*], which **when**

mutated lead to the development of AGS *(5. Aicardi Goutières Syndrome (AGS)*)**, can inhibit LINE-1 retrotransposition** when overexpressed in cells. However, **RNase H2**, **is another AGS gene whose impact on L1 regulation is currently unknown** (Crow *et al*, 2006b).

- **PCNA has been observed to interact with L1-ORF2p**, via a PIP motif located between the EN and RT domains (Taylor *et al*, 2013*)*. The interaction of L1-ORF2p with PCNA is required for retrotransposition, and probably occurs during or immediately after TPRT.

Hypothesis

Based on the above L1 knowledge, I propose several hypotheses that will be explored in this Thesis:

- Considering that during TPRT a Y-like intermediate (L1-cDNA:L1-mRNA attached to genomic DNA) is generated, with the involvement of PCNA, and as the L1-mRNA of the hybrid must be removed to complete a round of retrotransposition, **I propose that RNase H2 may be a possible factor involved in L1-silencing, or in the control of L1 retrotransposition**. To note, the interaction of L1-ORF2p with PCNA requires a PIP motif located between the EN and RT domains, and it is worth noting that **RNase H2 also contains a PIP motif** (in the B subunit).

- RC-L1s naturally retrotranspose in PCs, and new insertions are epigenetically silenced by host factors that might recognize the Y-like structure generated during TPRT. As L1 silencing is attenuated in DCs, **I propose that the factor/s involved in recognizing the Y-like structure might be intimately linked to human pluripotency**.

- Proteomic studies uncovered several factors regulating retrotransposition, and the background of cells used in these studies influence the nature of L1 interactors. Thus, **I propose that comparing the L1-interactome of isogenic PCs and DCs could identify pluripotent-associated interactors involved in L1-silencing**. While these experiments will ultimately reveal whether proteomic studies could be useful to identify factors involved in L1-silencing, these experiments will also reveal how dynamic the L1-interactome is during cellular differentiation, which can provide further insight in L1 regulation.

V. OBJECTIVES

Thesis Objectives

- **General Objectives:**

1. Determine the LINE-1 interactome in pluripotent (PCs) and differentiated (DCs) cells and generate a list of L1-interactors involved in L1 silencing.

2. Study the role of RNase H2 on LINE-1 retrotransposition.

- **Specific Objectives:**

1. Generation of a new L1 mobility system that allows the isolation of LINE-1 mobility intermediary complexes (RNPs) in PCs and DCs cells.

2. Establish a list of genes that interact with L1-RNPs in PCs and DCs.

3. Analysis of cellular factors possibly associated with epigenetic silencing of L1 mobility in PCs.

    - Sub-objective 3.1. Role of RNase H2.

VI. MATERIALS AND METHODS

1. Materials

Culture plates
100 cm^2 culture plates
T75 culture flasks
6-well culture plates
500 cm^2 culture plates

Cell culture and subculture
DPBS 1X
TRYPSIN-EDTA (1X), 0.05%, 500ML (Gibco, Cat. 25300-062)
TrypLE Express Enzyme™ (1X), sin rojo fenol (Gibco, Cat. 12604021
Matrigel (basement membrane matrix 10 ml * without LDEV) (BD, Cat. 354234)

Serums and media
Dulbecco's modified Eagle's medium (DMEM; Gibco, Cat no 41965-039)
McCoy's 5A medium (Gibco; Cat no 26600-023)
Fetal bovine serum (FBS)
Minimum Essential Medium, HEPES, GlutaMAX™ Supplement (Gibco, Cat. 42360032)
Heat inactivated Fetal bovine serum (heat inactivated FBS)
MEM Non-Essential Amino Acids Solution (100X) (Gibco, Cat. 11140050)
KnockOut Serum replacement medium (Invitrogen, Cat. 10828028)
Retinoic Acid, trans-only (Sigma, Cat. R2625-100MG)
ROCK inhibitor (StemCell, Cat. Y-27632)
Penicilin/Streptomicin 100 ml (100X) (Gibco, 15140122)
Fetal bovine serum (FBS) with 10% DMSO
Fetal bovine serum heat inactivated (heat inactivated FBS) with 10% DMSO

Cell transfection
X-tremeGENE ™ 9 - DNA transfection reagent (Roche, Cat. XTG9-RO)
FuGENE 6 ® (Promega, Cat. E2691)
Optimem (GIBCO, 31985-047)

Gel electrophoresis and western blot
Micro BCA™ Protein Assay Kit (Thermo Fisher, Cat. 23235)
Bradford assay Protein Assay Kit (Bio-Rad, Cat. 5000001)
HEPES, EDTA
RIPA Buffer (Sigma-Aldrich, Cat. R0278)
cOmplete™ EDTA-free Protease Inhibitor Cocktail (Roche, Cat. 11873580001)
β-mercaptoethanol (Sigma-Aldrich, Cat. M6250)

PMSF (Sigma-Aldrich, Cat. P7626)
Tris, Glycine, SDS, Methanol (Sigma-Aldrich)
Precision Plus Protein™ Dual Color Standard (Bio-Rad, Cat. 1610374)
Precision Plus Protein™ All Blue Standards (Bio-Rad, Cat. 1610373)
NuPAGE™ 4-12% Bis-Tris Protein Gels (ThermoFisher, Cat. NP0321PK2)
NuPAGE™ MOPS SDS Buffer Kit (for Bis-Tris Gels) (ThermoFisher, Cat. NP0050)
PVDF
Nitrocellulose
Ponceau S Staining Solution (Sigma, Cat. P7170)
Sypro Ruby, SYPRO® Ruby Protein Gel Stain reagent (Invitrogen, Cat. S12000)
TBS, Tween, Milk powder, BSA.
Clarity™ Western ECL Kit (Bio-Rad, Cat.

Antibodies

Sheep anti-RNase H2 antibody (raised against human recombinant RNase H2, Reijns *et al,* (2012), 1:1. 000)
Rabbit anti-RNASEH2A antibody (Origene, Cat. TA306706, 1:1,000)
Mouse anti-RNASEH2A antibody (Santa Cruz, Cat. sc-515475, 1:1,000)
Mouse anti-α-tubulin B512 (Sigma, Cat. T6074, 1:5,000)
Mouse anti-vinculin antibody (Sigma, Cat. V9264, 1:1,000)
Rabbit anti-L1Hs-ORF1p antibody (provided by Dr Oliver Weichenrieder, Max-Planck, Germany, Macia *et al,* (2017), 1:5. 000)
Mouse anti-V5 antibody (clone V5-10, Sigma, Cat. V8012, 1:10,000)
Mouse monoclonal anti-LINE-1 ORF1p Antibody, clone 4H1 (Sigma-Aldrich, Cat. MABC1152)
Dynabeads M-270 Epoxy (Invitrogen, Cat. 143.02D)
Mouse monoclonal Anti-β-Actin antibody (Sigma-Aldrich, Cat. A1978)
Anti-mouse IgG, HRP-linked Antibody (Cell Signaling, Cat. 7076)
Anti-rabbit IgG, HRP-linked Antibody (Cell Signaling, Cat. 7074)

2. Cell line culture.

HeLa cells were a gift from G. Stewart (Birmingham), originally obtained from ATCC and were cultured in Dulbecco's modified Eagle's medium (DMEM; Gibco, Cat no 41965-039) supplemented with 10% fetal bovine serum (FBS), 50 U/ml penicillin and 50 μg/ml streptomycin.

U2OS cells were purchased from the European Collection of Authenticated Cell Cultures (ECACC, Cat no. 92022711) and HCT116 p53-/- cells (Bunz, 1998; Dornan *et al*, 2004) were a gift from K. Ball (Edinburgh). They were grown in modified McCoy's 5A medium (Gibco; Cat no 26600-023) supplemented with 10% fetal bovine serum (FBS), 50 U/ml penicillin and 50 μg/ml streptomycin. All cell lines were cultured at 37°C, 5% CO_2 and atmospheric O_2.

PA-1 cells were a gift from J. V. Moran (Michigan), originally obtained from ATCC. They were grown in MEM - GLUTAMAX (Gibco) supplemented with 10% of heat inactivated fetal bovine serum (FBS), 50 U/ml penicillin and 50 μg/ml streptomycin and 1X of MEM-Minimum Essential Aminoacids (Gibco). Fetal bovine serum was heat inactivated by heating at 56°C for 20 minutes. PA-1 cells were differentiated through its culture in in MEM - GLUTAMAX (Gibco) supplemented with 10% of KnockOut Serum Replacement medium (Invitrogen), 10 μM of retinoic acid (Sigma), 50 U/ml penicillin and 50 μg/ml streptomycin and 1X of MEM-Minimum Essential Aminoacids (Gibco). In both cases PA-1 were cultured 37°C, 5% CO_2 and atmospheric O_2.

HEK293T$_{LD}$ cells were a gift from John LaCava (New York), and were grown in the conditions indicated in Taylor et al, 2010 and Taylor *et al*, 2018.

Cells were checked at least once a month using Lonza-Mycoalert Mycoplasma Detection Kit in order to asset they were mycoplasma free. In addition, the identity of the cell lines was confirmed by STRs analysis at least once a year (Lorgen, Granada, Spain).

2.1. Cell thawing

In order to thaw the cell lines, the following protocol was used:

1. Prepare a 15 ml falcon tube with 5 ml of tempered culture medium (37 °C).
2. Take the cryotube with the cells from LN_2 or -80°C and thaw in a bath at 37°C (until only a small ice ball is visible). Transfer the volume to the 15 ml falcon tube. Take 1 ml and transfer it back to the cryotube to collect the remaining cells, then return this volume to the 15 ml tube (to recover as many cells as possible). This step should be done carefully but as quickly as possible as the cells are frozen in fetal bovine serum (FBS) with a 10% of DMSO and DMSO is toxic to the cells.
3. Centrifuge at 1200 rpm for 4 minutes.
4. Check the cell pellet and carefully aspirate the culture media.
5. Resuspend in 10 ml of tempered culture medium and transfer to a T75 flask.
6. Move the flask to distribute the cells homogeneously.

2.2. Subculture routine

Cells were routinely maintained in T75 culture flasks. For a T75 culture flask with cells at 80-90% confluency, the subculture procedure detailed below was followed:

1. Aspirate the culture medium.
2. Wash off any remaining medium by adding 3 ml of 1X PBS to the flask and rotating the flask so that it spreads over the entire surface.
3. Aspirate the 1X PBS. Add 2.5 ml of tempered 0.05% TRIPSIN-EDTA (1X) or TrypLE™ Express Enzyme (1X) and incubate at 37 °C for 3-5 minutes (time depends on cell confluency, it is ready when cells detach). To seed the cells for an experiment or if they are undergoing differentiation, use TrypLE™ Express (1X).
4. Neutralize with 2.5 ml of culture medium and resuspend the cells.
5. Dilute the cells depending on when confluent cells will be needed, recommended dilutions are between 1:5 (confluent in one to two days) and 1:10-1:15 (this dilution allows the cells to hold for 3-4 days. PA-1 cells should be around 5-10% confluent when seeded (otherwise they may not grow).
6. Add 10 ml of warm culture medium and resuspend the cell pellet. Transfer to a new T75 flask (or keep in the same one if preferred in case it has few uses).
7. Move the flask to distribute the cells homogeneously.

2.3. Cell freezing

In order to freeze cells, the following procedure was followed. Cells were frozen once they were 85-95% confluent. Three to four cryotubes were prepared for a confluent T75 culture flask.

1. Aspirate the culture medium.
2. Wash off any remaining medium by adding 3 ml of 1X PBS to the flask and rotating the flask so that it spreads over the entire surface.
3. Aspirate the 1X PBS.
4. Add 2.5 ml of tempered 0.05% TRIPSIN-EDTA (1X) or TrypLE ™ Express Enzyme (1X) and incubate at 37 °C for 3-5 minutes (time depends on cell confluency, it is ready when cells detach). To seed the cells for an experiment or if they are undergoing differentiation, use TrypLE™ Express (1X).
5. Neutralize with 2.5 ml of culture medium and resuspend the cells.
6. Transfer the cells with the medium to a 15 ml tube and centrifuge at 1000 rpm for 5 minutes.
7. Aspirate the culture medium and add 3-4 ml of COLD FBS (heat inactivated for PA-1 cell line) with a 10% DMSO (should be as cold as possible as DMSO is toxic to the cells).
8. Carefully resuspend the cells and transfer 0.5-1 ml of cells to each cryotube (as quickly as possible).
9. Freeze the cells at -80°C. Cells can be viable at this temperature for approximately 4-6 months if the process has been followed as indicated. For prolonged storage, move the cells to LN2 after 1-2 days.

3. Cell transfection

The following procedure was followed to perform cell transfections making use of X-tremeGENE™ 9 or FuGENE 6 ® (Promega):

The day before transfection, the cells to be transfected were trypsinized, counted by Neubauer chamber and seeded. Depending on the assay, between $2 \cdot 10^4$ and $1 \cdot 10^5$ cells were seeded. The seeded wells were kept at 37°C for at least 16-18 h before transfection (never more than 20 h). This incubation time is necessary for the cells to adhere to the plate, but should not be exceeded, as this reduces transfection efficiency.

For one well of a 6-well plate:

- In a sterile tube, add 485 ul of room temperature (RT) tempered Opti-MEM.
- Add 15 µl of X-tremeGENE™ 9 to the tube (dropwise, directly onto Opti-MEM) and incubate at RT for 5 minutes (up to 20 minutes).
- In another sterile tube, add 5 µg of plasmid DNA.
- Once incubation is complete, add the Opti-MEM / X-tremeGENE™ 9 mix to the tube with the DNA.
- Move the tube to make sure everything is mixed and incubate at room temperature for 20 minutes (up to 60 minutes).
- Add the mixture dropwise over the cells.
- Incubate overnight at 37 °C.
- Change the culture media the next day.

4. Gel electrophoresis, staining and western blotting.

Acrylamide gel electrophoresis

For electrophoresis, either homemade gels or NuPAGE™ 4-12% Bis-Tris Protein Gels (ThermoFisher) were used.

Homemade gels of 1.5 mm thickness were prepared with a 15% of polyacrylamide for the separator part when the protein to be observed was ORF1, and 5% of polyacrylamide for the concentrator gel. Both the separator and concentrator gels were left to polymerize for half an hour and later, stored at 4°C wrapped in paper moistened with double distilled water when not immediately used.

The electrophoresis system was set up, filled with electrophoresis buffer (0.25 M Tris, 1.92 M Glycine, 1% SDS) or NuPAGE™ MOPS SDS Running Buffer (20X) (ThermoFisher) and 5-20 μl of sample (10 to 50 μg of protein) in loading buffer (homemade loading buffer [4X, 8% SDS, 24% glycerol, 240 mM Tris-HCl pH 6.8, 4.8 % β-mercaptoethanol and 0.4-0.8 % Bromophenol Blue] or NuPAGE LDS Sample Buffer (4X)(ThermoFisher)) per well were loaded, depending on the gel capacity and the particular experiment. As a marker, 5 μl/well of Precision Plus ProteinTM All Blue Standards molecular weight marker (BioRad) was used. The gel was then run at 200 V for 40-60 minutes.

In order to stain the gels with Sypro Ruby, SYPRO® Ruby Protein Gel Stain reagent (Invitrogen, Cat. S12000) was used following the manufacturer's instructions for the Basic Protocol.

Transference to nitrocellulose or PVDF membrane

Following electrophoresis, proteins were transferred to a nitrocellulose or PVDF membrane using the wet transfer system (BioRad). Gel, membrane, sponges and Whatman paper were equilibrated in transfer buffer (5.8 g Tris, 2.9 glycine and 20 ml methanol in 1 liter preparation) before assembling the system. In case of a PVDF membrane, it was previously activated by wetting it in methanol for 10 seconds, and later in transfer buffer for a couple of minutes. After equilibration, the transference sandwich, consisting of the support, sponges and Whatman paper arranged on both sides of the gel, was assembled and the system was run at a fixed amperage of 70 V for two hours.

Subsequently, in order to confirm a correct transference, the membrane was stained with Ponceau S Solution (Sigma) to ensure the presence of the proteins.

Immunoblotting

Following transference, the membrane blocked. Blocking was performed by using TBS-Tween (0.1%) with 5% milk and blocked for 2 hours under agitation. Later, the membrane was incubated at 4 °C with the primary antibody of interest in 0.1% TBS-Tween - 5% BSA overnight with agitation.

The next day three 10-minute washes with 0.1% TBS-Tween under agitation were carried out and then the membrane was incubated with an anti-Rabbit or anti-Mouse IgG secondary antibody coupled to Horse Raddish Peroxidase (HRP) (Cell Signaling) at a dilution of 1:1000/1:10000 for 1 hour at room temperature. After this time, the previous washing steps were repeated.

Finally, the membrane was developed using Bio-Rad's Clarity™ Western ECL Kit using a 1:1 ratio per membrane as indicated in the manufacturer's instructions and using the Image Quant device, which detects chemiluminescence, to obtain a digital image.

In order to have a loading control against which protein expression could be quantified, a second round of incubation, washing and developing was carried out, this time using HRP-conjugated antibodies to actin or tubulin (1:1000), when necessary. Finally, protein expression was quantified using ImageJ software.

The following primary antibodies were used for immunoblotting (at the indicated dilutions): sheep anti-RNase H2 antibody (raised against human recombinant RNase H2, Reijns *et al, (2012*), 1:1000); rabbit anti-RNASEH2A antibody (Origene TA306706, 1:1000) mouse anti-RNASEH2A antibody (Santa Cruz sc-515475, 1:1,000); mouse anti-a-tubulin B512 (Sigma T6074, 1:5000); mouse anti-vinculin antibody (Sigma V9264, 1:1000); rabbit anti-L1Hs-ORF1p antibody (provided by Dr Oliver Weichenrieder, Max-Planck, Germany, 1:5000); mouse anti-L1RE1-ORF1p antibody (clone 4H1, Sigma MABC1152, 1:5000); mouse anti-β-actin antibody (1:20000; Sigma); mouse anti-V5 antibody (clone V5-10, Sigma V8012, 1:10000). In quantitative westerns, goat anti-rabbit and anti-mouse fluorescent secondary antibodies were used at a dilution of 1: 20000. As secondary antibodies we used horseradish peroxidase (HRP)-linked antibodies (Cell Signaling).

5. Plasmids

5.1. Plasmid production

All plasmids were grown in DH5α-competent *E. coli* DH5α (F- Φ80lacZΔM15 Δ (lacZYA-argF) U169 recA1 endA1 hsdR17 (rK-, mK +) supE44 phoA λ-thi-1 gyrA96 relA1).

Competent bacteria were prepared as described in Inoue *et al, (1990*). DH5α-competent *E.coli* were transformed to obtain transformant colonies expressing each of the plasmids by the following procedure. First, competent bacteria were thawed on ice and then 10 μl of bacteria were added to 1 μl of plasmid and incubated on ice for 30 minutes. After 30 minutes, a heat shock was given at 42 °C for 45 seconds by shaking the tube for the first 25 seconds. Finally, bacteria were incubated on ice for 2 minutes and then 100 μl of pre-warmed LB was added and streaked onto a 100 μg/ml LB-ampicillin agar plate.

Once the colonies were obtained, plasmid preparations were prepared using the Qiagen Plasmid Midi Kit (Qiagen, Germany) according to the manufacturer's protocol. For this, 1-2 colonies were initially seeded in flasks with 50 ml LB-ampicillin 100 μg/ml and grown in an incubator at 37 °C and 220 rpm for 14-16 hours.

Plasmid DNA was analyzed by electrophoresis (0.7% agarose-ethidium bromide gels) and only highly "supercoiled" plasmid preparations were used for transfection experiments.

5.2. Characteristics and list of the plasmids employed

The plasmids used for retrotransposition/transposition assays involving new insertions selection by Geneticin (or G418) or Blasticidin (Blast) are characterized by having cassettes with the genes conferring resistance to these antibiotics in such a configuration that only after an insertion process their expression can be activated (**Figure 9**). In the case of retrotransposon cassettes, these genes are located antisense to the transcription direction of the mobile element to be studied. In addition, they are interrupted by an intron (often the γ-globin intron) that goes in the direction of the transcript and when the mobile element is transcribed, it undergoes splicing allowing the resistance gene to be uninterrupted. Thus, if the mobile element is transcribed and inserted, the resistance gene (*mneoI* or *mblastI*) is expressed and confers resistance to the antibiotic (G418 or blast) (**Figure 10**).

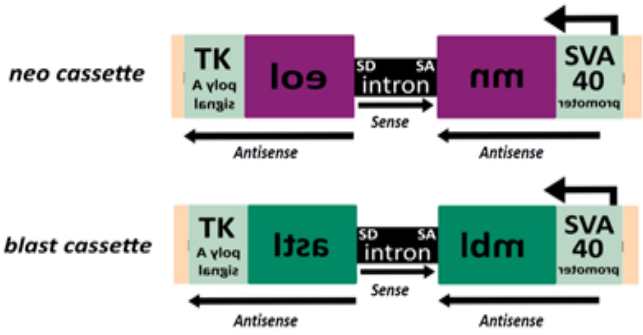


Figure 10. Antibiotic resistance genes. *Resistance genes have their own promoter and polyadenylation signal. The mneol gene codes for the neomycin phosphotransferase (active against geneticin) and blastI for blasticidin S deaminase (active against blasticidin S).*

Other plasmids were used for the generation of Knock-Out cell lines, carrying the Cas9 nuclease and the guide RNAs (sgRNAs) necessary for its function. Moreover, another kind of plasmids carried the cDNA sequence of a gene of interest, allowing the overexpression of the selected gene.

5.3. Plasmids employed

pSpCas9n(RNASEH2A-g1)-2A-GFP and pSpCas9n(RNASEH2A- g2)-2A-Puro express guide RNAs (gRNAs) designed against exon 1 (TGCCCGCCTCATCGACGCCC) and intron 1 (CCCGTGCTGGGT GCGCCCCT) of the RNASEH2A gene, encoding the Cas9n (nickase mutant D10A) fused to the cDNA of the enhanced green fluorescent protein (EGFP) and the puromycin N-acetiltransferase (puro), respectively. The gRNAs were generated by alignment of the DNA oligonucleotides y were cloned in the BbsI site of the vectors pSpCas9n(BB)-2A-GFP and pSpCas9n (BB)-2A-Puro (Addgene, Cat. 48140 and 48141, respectively; a gift from Feng Zhang) as previously described in (Ran *et al, 2013)*.

pMSCVpuro-RNASEH2A-WT, this plasmid contains the coding sequence of human RNASEH2A (NM_006397.2) cloned into the pMSCVpuro-Dest vector, a Gateway-compatible version of the pMSCVpuro vector (Clontech).

pMSCVpuro-RNASEH2A-SoF, is a vector derived from the pMSCVpuro-RNASEH2A-WT plasmid containing the human RNASAH2A sequence with two miss-sense mutations (P40D/Y210A) resulting in a separation of function mutant (Chon *et al, 2013)*.

PMSCVpuro-RNASEH2A-CD, a plasmid derived from the pMSCVpuro-RNASEH2A-WT vector containing the human RNASAH2A sequence with two miss-sense mutations (D34A/D169A) resulting in the loss of the catalytic activity (catalytic dead; Reijns *et al, (2011*)).

pMSCVpuro-RNaseH1, plasmid encoding for the nuclear isoform of the human RNASEH1 gene (NM_002936.5, aa27-286), cloned into the pMSCVpuro-Dest vector.

pcDNA3.1/nV5-RNASEH2A, this plasmid encodes the human RNASEH2A gene cloned into the pcDNA3.1/nV5-DEST vector (contains a V5 epitope at the N-terminal position).

cDNA3.1/nV5-RNASEH2A-SoF, encodes the human RNASEH2A gene with the miss-sense mutations P40D/Y210A cloned into the vector pcDNA3.1/nV5-DEST (contains a V5 epitope at the N-terminal position).

pcDNA3.1/nV5-RNASEH2B, plasmid encoding the sequence of the human RNASEH2B (NM_024570.3) cloned into the pcDNA3.1/nV5-DEST vector (contains a V5 epitope at the N-terminal position).

pcDNA3.1/nV5-RNASEH2C, plasmid encoding the sequence of the human RNASEH2C (NM_032193.3) cloned into the pcDNA3.1/nV5-DEST vector (contains a V5 epitope at the N-terminal position).

pK-barr, previously described in (Bogerd et al, 2006), expresses human β-arrestin tagged with the HA epitope at its C-terminal end.

pK-A3A, previously described in (Bogerd *et al*, 2006), expresses human APOBEC3A tagged with the HA epitope at its C-terminal end.

pU6ineo, previously described in (Richardson *et al*, 2014). This plasmid encodes an expression cassette for the neomycin phosphotransferase (NEO) from the vector pEGFP-N1 (Clontech) cloned into a pBSKS-II (+) vector (Stratagene) which contains a U6 promoter in its multi cloning site.

pcDNA6.1, (Invitrogen) encodes an expression cassette with the blasticidin S deaminase.

JM101/L1.3, previously described in (Sassaman *et al*, 1997), contains a complete copy of a human L1.3 element tagged with the *mneoI* reporter cassette (Freeman *et al*, 1994; Moran *et al*, 1996), and is cloned in pCEP4 (Life Technologies).

JM101/L1.3-D205A, previously described in (Wei *et al*, 2001), is a JM101/L1.3-derived plasmid containing a miss-sense mutation (D205A) in the EN domain of L1-ORF2p.

JM101/L1.3-D702A, previously described in (Wei *et al*, 2001), is derived from JM101/L1.3 and contains a miss-sense mutation in the RT domain of L1-ORF2p (D702A).

JJ101/L1.3, was previously described in (Kopera *et al*, 2011). It encodes for a full copy of a human L1.3 element (Sassaman *et al*, 1997) tagged with the *mblastI* indicator cassette (Morrish *et al*, 2002; Goodier *et al*, 2007) and is cloned in the pCEP4 vector (Life Technologies).

JJ101/L1.3-D205A, previously described in (Kopera *et al*, 2011); is a plasmid derived from JJ101/L1.3 containing a miss-sense mutation (D205A) in the EN domain of L1-ORF2p.

JJ101/L1.3-D702A, previously described in (Kopera *et al, 2011)*; derived from JJ101/L1.3 and contains a miss-sense mutation in the RT domain of L1-ORF2p (D702A).

JJ101/L1.3-PIP6, is derived from JJ101/L1.3 and contains two miss-sense mutations

in the PIP domain of L1-ORF2p (Y414A, Y415A).

pXY014, was previously described in (Xie *et al, 2011)* and encodes for a full copy of the human L1RP element (Kimberland *et al, 1999)* tagged with the *mflucI* indicator cassette (Xie *et al, 2011)* and is cloned into a modified pCEP4 vector (Life Technologies) encoding a Renilla luciferase expression cassette.

pXY017, previously described in (Xie *et al, 2011)*; is derived from pXY014 containing two miss-sense mutations in the RNA-binding domain of L1-ORF1p (RR261/62AA). This plasmid was used as a negative control for luciferase-based retrotransposon assays.

pCMVMusD-6neoTNF, previously described in (Ribet *et al, 2004)*; contains a complete copy of the mouse MusD element (AC124426, positions 9,078-16,569 (+)) tagged with the neo^{TNF} indicator cassette (Esnault, 2002) and is cloned into the pCMVVbeta vector (Clontech).

Zfl2-2mneoI, previously described in (Sugano *et al, 2006;* Garcia-Perez *et al, 2010)*; contains a complete copy of the zebrafish Zfl2-2 element tagged with the *mneoI* indicator cassette at the 3'UTR end of LINE-1 (Freeman *et al, 1994;* Sugano *et al, 2006)* and is cloned into pCEP4 (Life Technologies).

pT2neo, previously described in (Mátés *et al, 2009)*; contains the cDNA of the neomycin phosphotransferase directed by SV40 and flanked by SB TIR.

pCMV-SB100x, previously described in (Mátés *et al, 2009)*; contains a hyper-reactive SB Transposase targeted by the CMV promoter.

pCEP-EGFP, previously described in (Alisch, 2006); contains the coding sequence of the humanized GFP protein cloned into pCEP4 (Invitrogen).

pGEX6P1-hsRNASEH2BCA, pGEX6P1-hsRNASEH2BCA(D34A/D169A), pGEX6P1-hsRNASEH2B(A177T)CA and pGEX6P1-hsRNA-SEH2BC(R69W)A, previously described in (Reijns *et al, 2011)*, enable the expression of the GST-tagged human RNASEH2B subunits and the untagged RNASEH2C and A subunits in *Escherichia coli*. The amino acid substitutions indicated in brackets were introduced into the relevant subunits by targeted mutagenesis.

pGEX6P1-hsRNASEH2BCA(P40D/Y210A), pGEX6P1-hsRNA-SEH2BCA(G37S) and pGEX6P1-hsRNASEH2BCA(E225G) have the P40D/Y210A separation of function mutation and the G37S and E225G AGS mutations respectively, introduced into RNASEH2A by targeted mutagenesis.

pLD401, codon optimized human full-length “hot” RC-L1 (L1.3, known as

ORFeus-H) with 3xFLAG tagged L1-ORF2p, described in (Taylor *et al, 2013)*.

pM302, native human full-length "hot" RC-L1 (L1.3) with 3xFLAG tagged L1-ORF2p, previously described in (Dai *et al, 2012;* Taylor *et al, 2013)*.

pJM101/L1.3 – ORF1-T7 / ORF2-3xFLAG contains a copy of the "hot" human RC-L1 (L1.3), where the C-term of L1-ORF1 is tagged with a T7 tag and L1-ORF2p contains a triple FLAG epitope in the C-term. A retrotransposition indicator cassette, mneoI, is cloned in the 3´UTR of L1.3. The vector backbone contains a hygromycin resistance cassette as well as OriP and EBNA-1, allowing its selection in cultured cells as a non-integrating episome.

pDK101. contains a copy of the "hot" human RC-L1 (L1.3), where the C-term of L1-ORF1 is tagged with a T7 tag. A retrotransposition indicator cassette, mneoI, is cloned in the 3´UTR of L1.3, and is cloned in pCEP4 (Life Technologies).

LD208/PM65, plasmid carrying the rtTA Advanced from Clontech and a neomycin resistance gene, described in Urlinger *et al*, 2000.

pLD215/PM66, plasmid carrying the rtTA advanced from Clontech and a blasticidin resistance gene (BSD). Replacement of the neomycin resistance gene with the BSD gene was done by Lixin Dai.

6. Retrotransposition assays.

The retrotransposition assay consists on a LINE-1 based mobilization assay using an active human LINE-1 element (L1.3, Sassaman et al, 1997) labeled with a reporter gene that can only be activated after a round of retrotransposition (Moran *et al*, 1996).

In this assay, a transient transfection of a cell line with a LINE-1 construct (or any other mobile element) containing a reporter cassette (reporter gene) is performed. The expression of the reporter cassette, which occurs only after a successful round of retrotransposition, allows the detection and quantification of the rate of retrotransposition.

The rate of retrotransposition evaluated does not belong to an endogenous LINE-1 element, but to a construct, and its evaluation provides information about the mechanism of LINE-1 retrotransposition, its regulation by the cell and the impact that retrotransposition generates in the genome. The process that takes place in the transfected cells and that finally allows colonies to originate from those cells in which retrotransposition events occurred is shown in **Figure 9** (*III. INTRODUCTION, 2.2.1. A system to study retrotransposition: The retrotransposition assay*).

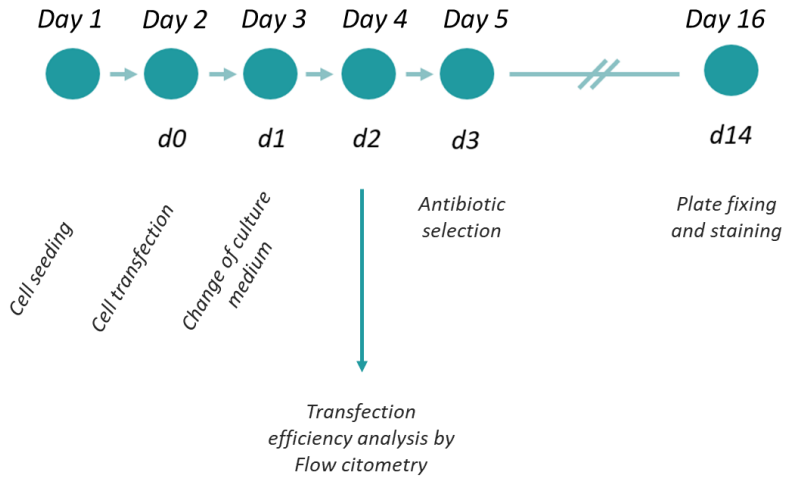


Figure 11. Retrotransposition assay procedure. *Description of the assay in the text below.*

The retrotransposition assay lasts for 16 days (**Figure 11**). The following procedure is followed:

Day 1 - Cell seeding. The first day, cell seeding is carried out. Generally, 2×10^4 HeLa/U2OS/HCT116 cells were plated in each well of a 6-well culture plate with the usual culture medium. To measure transfection efficiency, 1×10^5 cells per well were plated in the designated wells. The cells were cultured in a humidified incubator at 37 °C with 7% CO2.

Day 2 - Transfection. On the second day, transfection was performed 16-18 h after cell seeding (d0) (**Figure 11**). FuGENE 6 ® transfection reagent (Promega) is used according to the manufacturer's instructions. For a 6-well plate, for each well, a mixture of 97 µl of Optimen (Invitrogen) at room temperature and 3 µl of FuGENE 6 ® (Promega) is prepared in a sterile tube and incubated for 5 min at room temperature. In another sterile tube, 1 µg of plasmid DNA is added in a separate sterile tube. In cases of co-transfection, 0.5 µg of each plasmid (for 2 vectors) or 0.33 µg of each plasmid (3 vectors) were added. Subsequently, 100 µl of the previously incubated mixture is added on top of the DNA and incubated for 20 minutes. After this time, the mixture is added dropwise over the cells and incubated overnight at 37 °C. Two technical replicates of each condition are performed. In all the cases, a drug toxicity control condition, consisting of a vector directly expressing the resistance reporter gene (pU6i neo or pCDNA 6.1), is included as a control for drug toxicity. The wells used to measure transfection efficiency are transfected with the pcepGFP vector.

Day 3 - Change of culture medium (d1). The day after transfection, a culture medium change is carried out by aspirating the medium with the transfection mixture and adding fresh medium, in order to stop the transfection.

Day 4 - Determination of the transfection efficiency (d2). Two days after transfection, transfection efficiency is measured. The wells are trypsinized (neutralized with a PBS - 3% of FBS solution) and the cells are collected in cytometer tubes and centrifuged at 1200 rpm for 5 minutes. The supernatant is discarded and the cells are resuspended in PBS. Transfection efficiency measurements have been performed by flow cytometry on a BD FACS CantoTM II cytometer. Two populations are defined according to whether they exhibit green fluorescence or not, and the percentage of live cells expressing GFP is measured. This percentage is an indicator of the percentage of cells successfully transfected with the plasmids, in other words, the transfection efficiency.

Days 5-16 - Selection of antibiotic resistant cells with retrotransposon events. Three days after transfection (d3), selection by antibiotics is started and extended for up to 14 days after transfection (d14) for assays with *mneoI* as a reporter cassette (**Figure 11**). For the neomycin reporter cassette, 400 μg/ml G418 were added to the culture medium from a stock of 50 mg/ml G418. The culture medium with G418 was changed every other day up to 14 days after transfection (d14). For *mblastI* assays, blasticidin S-HCl at a 10 μg/ml concentration of from a stock of 10 mg/ml was added 5-6 days after transfection (d6-d7).

Day 16 - Quantification of the LINE-1 retrotransposition assay. For both mneoI and mblastI cassette assays, 14 days after transfection (d14), colony fixation was performed using a 0.2% formaldehyde and 0.2% glutaraldehyde fixing solution in PBS. Colonies were fixed from 30 minutes to 1 hour at room temperature. Subsequently, colonies were stained using a 0.1% crystal violet solution for 30 minutes to 1 hour. Finally, the colonies were rinsed with water and air dried.

In order to estimate the retrotransposition rate, the colonies in each well of the same condition were counted and the mean and standard deviation of the replicates were calculated. The means were adjusted according to transfection efficiency by dividing the retrotransposition mean and standard deviation by the transfection efficiency. To express the adjusted retrotransposition rate as a percentage, in cases where the toxicity control (pU6i neo or pCDNA 6.1) showed significant differences between the different lines, the adjusted retrotransposition mean was divided by the adjusted mean of the toxicity control and multiplied by 100. Otherwise, the percentage was adjusted against the condition with the highest retrotransposition rate (considered as 100 %).

7. Transposition assays.

In order to analyze the transposition of the Sleeping Beauty (SB) transposon in HeLa and U2OS parental cells and the RNASEH2A-KO and control clones, about 1 x 10^5 cells per well were seeded in a 6-well plate. Eighteen hours later, plasmid DNA transfection of the SB system vectors was performed using FuGENE 6® (Promega) and Opti-MEM (Gibco) following the previously indicated protocol *(V. MATERIALS AND METHODS, 3. Cell transfection)*. In each well, cells were transfected with 1 μg of pT2neo plasmid (control) or co-transfected with 1 μg of pT2neo plasmid and 0.5 μg of pCMV-SB100x plasmid. Seventy-two hours after transfection, the cells were trypsinized and counted; subsequently, 10% of the transfected cells were seeded on a 10 cm plate. Forty-eight hours later, selection with G418 (400 μg/ml) was started. Selection with G418 was extended for 10 days,

and G418-resistant colonies were fixed with 2% paraformaldehyde/0.4% glutaraldehyde and stained with 0.1% crystal violet solution. As in the retrotransposition experiments, co-transfection with the pCEP-EGFP plasmid was performed in parallel to control for differences in transfection efficiency.

8. Generation of KO cell lines and RNase H2 mutants.

In order to generate the RNASEH2A-KO cell lines, the parental cell line was seeded in 6-well plates and transfected with the two vectors encoding sgRNA and Cas9n using Lipofectamine 2000 (Thermo Fisher Scientific) following the manufacturer indications. Forty-eight hours after transfection, individual cells expressing EGFP were distributed into wells of 96-well plates using the BD FACSJazz kit (BD Biosciences) and cultured to generate cell lines. RNASEH2A-KO clones were selected based on the size of the PCR products of the target region, and deletions/insertions were subsequently confirmed by Sanger DNA sequencing.

The oligonucleotides (5'-3') used for PCR amplification and sequencing of the RNASEH2A loci were 5'-ACCCGCTCCTCCTGCAGCAGTATTAG and 5'-TCCCTTGGTGGTGCAGTGCAATC.

The lack of functional RNASEH2A was confirmed by Western blotting, RNase H2 activity assay and using alkaline gel electrophoresis as described below. Clones validated as functionally null were selected as KO clones. Some KO clones retained very low levels of RNASEH2A protein expression after long exposure of Western blot membranes upon development. For these clones, Sanger sequencing showed the presence of deletions in the reading frame, in each case removing essential residues from the catalytic site of the enzyme, including Asp34, making them enzymatically nonfunctional. Clones expressing wild-type RNASEH2A protein were identified in parallel and used as controls.

9. Complementation of RNase H2 KO cell lines.

To complement the RNASEH2A-KO HeLa clones, cells were infected with retroviral supernatant produced in Amphotropic Phoenix pack-aging cells (Swift *et al, 2001)* using pMSCVpuro-based vectors in the presence of 4 μg/ml polybrene and were selected using 2 μg/ml puromycin to achieve stable integration.

10. Cell extracts preparation for RNase H2 experiments.

Whole cell extracts (WCE) for RNase H activity assays and for determining protein levels of the RNase H2 subunits were prepared by incubating the cells in lysis buffer 1 [Tris-HCl 50 mM pH 8, 0, 280 mM NaCl, 0.5% NP-40, 0.2 mM EDTA, 0.2 mM EGTA, 10% (vol/vol) glycerol, 1 mM DTT and 1 mM phenylmethylsulfonyl fluoride (PMSF)] for 10 min on ice. An equal volume of [HEPES 20 mM pH 7.9, KCl 10 mM, EDTA 1 mM, 10% (vol/vol) glycerol, DTT 1 mM and PMSF 1 mM] was added and subsequently incubated for an additional 10 minutes. Cell debris were removed by centrifugation (17,000 g for 10 min at 4°C), and protein concentration was determined by Bradford assay Protein Assay Kit (Bio-Rad).

In the overexpression assays performed with HeLa and U2OS cells, cells were transfected with the overexpression plasmids and subsequently harvested by trypsinization. Once the cells were detached, they were centrifuged at 1200 rpm for 4 min to obtain the cell pellet.

Extracts were prepared in lysis buffer 2 [RIPA (Sigma) supplemented 1× Complete Mini EDTA-free Protease Inhibitor cocktail (Roche), 0.1% Phosphatase Inhibitor 1&2 (Sigma), 1 mM PMSF (Sigma) and 0.25% β-mercaptoethanol (Sigma)]. Lysis buffer was added at a 1/4 wt/volume ratio (100 μL of buffer were added to a cell pellet of about 25 mg) and the cells were incubated on ice for 10 min. Cell debris were removed by centrifugation (1,000 g for 5 min at 4°C) and total protein concentration was determined using the Micro BCA kit (Thermo) following standard procedure.

Whole cell extracts employed in the section *2. Study of the LINE-1 interactome in pluripotent (PCs) and differentiated (DCs) PA-1 cells* were prepared following the same procedure used for the overexpression assays but using lysis buffer 3 [20 mM HEPES, pH7.4, 1% Tritonx-100, 500 mM NaCl, Protease Inhibitor Cocktail Tablets (complete EDTA-free) (ROCHE)] instead of buffer 2.

11. RNase H2 activity assay.

In order to assess RNase H2 activity in whole cell extracts, a fluorescent substrate release assay based on FRET (Fluorescence Resonance Energy Transfer) was performed as previously described in (Reijns *et al, 2011)*.

The specific activity of RNase H2 was determined by measuring the cleavage of a double-stranded DNA substrate containing a single embedded ribonucleotide (DRD:DNA). As a control to correct for background activity not derived from

RNase H2, the activity against a DNA:DNA substrate with the same sequence (a substrate without embedded ribonucleotide) was used as a control.

Substrates were formed by hybridizing a fluorescein-labeled oligonucleotide at its 3' end (5'GATCTGAGCCTGGCTGGGaGCT or 5'GATCTGAGCCTGGGAGCT; uppercase DNA, lowercase RNA) with a complementary DNA oligonucleotide labeled with DABCYL at its 5' end (Eurogen-tec). Reactions were performed in 100 µL of reaction buffer (KCl 60 mM, Tris-HCl 50 mM pH 8.0, MgCl2 10 mM, BSA 0.01%, Triton X-100 0.01) with a substrate concentration of 250 nM in 96-well plates with flat bottom at 24 °C. Whole cell lysates were prepared as described in the previous section, and a final protein concentration of 100 ng/µl was used. Fluorescence was read (100 ms) every 5 min for up to 90 min using a VICTOR2 1420 multilabel counter (Perkin Elmer), with a 480 nm excitation filter and a 535 nm emission filter.

To evaluate RNase H2 activity in whole cell extracts using the gel-based assay, different reactions were prepared with a range of protein concentrations (50-400 ng/µl) and incubated at 37 ° C in a 5 µl reaction volume with 2 µM of the substrate described above (5'gatctgagc-ctgggagct for RNA:DNA) for 30 min or 1 h. Reactions were stopped by adding an equal volume of 96% formamide, 20 mM EDTA and heating to 95 ° C. Products were resolved in denaturing PAGE (20%, 1×TBE), visualized on an FLA-5100 imaging system (Fujifilm) and quantified using ImageQuant TL (GE Healthcare).

12. Ribonucleotide detection in genomic DNA.

For the analysis of possible ribonucleotide insertion into genomic DNA, genomic DNA was isolated from ~ 1 million cells by lysis in ice-cold buffer (Tris-HCl 20 mM pH 7.5, NaCl 75 mM, EDTA 50 mM), then incubated on ice with 200 µg/ml proteinase K (Roche) for 10 min after addition of N-lauroylsarcosine sodium salt (Sigma) at a final concentration of 1%. Nucleic acids were then extracted sequentially with phenol equilibrated with TE, phenol:chloroform:isoamyl alcohol (25: 24: 1) and chloroform. They were then precipitated with isopropanol, washed with 75% ethanol and dissolved in nuclease-free water.

Prior to alkaline gel electrophoresis, 500 ng of total nucleic acids were incubated with 1 pmol of purified recombinant human RNase H2 (Reijns *et al, 2011)* and 0. 25 µg of DNase-free RNase (Roche) for 30 min at 37 °C in 100 µl of reaction buffer (60 mM KCl, 50 mM Tris - HCl pH 8.0, 10 mM MgCl2, 0.01% BSA, 0.01% Triton X-100). Nucleic acids were precipitated with ethanol, dissolved in nuclease-free water and separated in 0.7% agarose in 50 mM NaOH, 1 mM EDTA. After

electrophoresis, the gel was neutralized in 0.7 M Tris-HCl pH 8.0, 1.5 M NaCl and stained with SYBR Gold (Invitrogen). Images were obtained on an FLA-5100 imaging system (Fujifilm) and densitometry plots were generated using an AIDA image analyzer (Raytest).

13. Mutation analysis by PCR and cloning.

In order to test whether RNase H2 deficiency resulted in an increased mutation rate of L1 de novo insertions we used a previously described assay (Bogerd *et al, 2006)*. Parental HeLa, HeLa RNASEH2A-KO1 and KO2 cells were seeded in plates (1×10^5 cells per well in a 6-well plate) and after 18 h were transfected with 1 μg of JM101/L1.3 plasmid or JM101/L1.3-D702A (used as a control) using the conditions described above *(V. MATERIALS AND METHODS, 3. Cell transfection)*. Forty-eight (2 days) and one hundred and twenty hours (5 days) after transfection, genomic DNA (gDNA) was isolated from the transfected cells, digested with SwaI (NEB; there is a single SwaI site within the intron of the mneoI Retrotransposition reporter cassette) and PCRs were performed using primers NEO437 (5' GAGCCCCTGATGCTCTCTTCGTCC) and NEO1808as (5' CATTGAACAAGATGGATGGATTGCACGC) flanking the intron of the *mneoI* reporter cassette. PCRs were performed in a 25 μl reaction volume using KAPA Taq ReadyMIx PCR and 0.4 μM of each primer. DNA-free water (Gibco), untransfected HeLa gDNA and SwaI-digested untransfected HeLa gDNA were included as negative controls. PCR conditions for NEO amplification were as follows: 1 × (95 °C, 3 min); 40 × (15 s, 95 °C; 15 s, 60 °C; 30 s, 72 °C); 1 × (72 °C, 1 min). PCR products were resolved on 1.5% agarose gels and the amplified products were cut, purified and cloned into pGEMT-Easy (Promega). The clones were sequenced using the M13FWD primer.

14. Statistical analysis of the retrotransposition / transposition assays and activity assays.

Unless otherwise stated, all statistics were performed using two-sided unpaired t-test (parametric) or Mann–Whitney test (non-para- metric) comparing replicates as indicated. $P < 0.05$ was considered as statistically significant.

15. Large-scale culture of pluripotent PA-1 (PCs) and differentiated PA-1 (DCs).

In order to culture PA-1 pluripotent (PC) or differentiated (DC) cells on a large scale, the culture periods were different, since differentiation requires the cells to be cultured for 14 days in differentiation medium.

For culturing 15 plates of 500 cm^2 with PA-1 PC cells and obtaining 90% confluency at three days after seeding, 6 T175 flasks (90% confluency) were used. Cells were detached using the procedure described in *2.2. Subculture routine* and the content of each T175 flask was used to seed two and a half 500 cm^2 plates. Subsequently, 70 ml of PA-1 PC culture medium was added to the plates. The cells were cultured at 37°C, 5 % CO2 and atmospheric O2 for three days, being collected following the procedure described in *16. Cell collection and freezing to prepare cell beads*.

In order to culture and differentiate PA-1 DC cells, a total of 8,000,000 pluripotent PA-1 cells were seeded per 500 cm^2 plate with 80 ml of the differentiation culture medium described in *2. Cell line culture*, except for the 10 μM retinoic acid (RA, all trans). The day of seeding, ROCK inhibitor (Stem Cell, Y-27632) was also added to favor the cells, seeded at a very low confluency, to adhere. Subsequently, two days later, 10 uM of retinoic acid (RA, all-trans) were added, starting the differentiation day 0 (D0). From day D0, the culture medium was changed every other day (D2, D5, D7, D9, D12) (**Figure 12**) using the complete differentiation medium as described in *2. Cell line culture*. On day 14 of differentiation, cells were collected following the procedure described in *16. Cell collection and freezing to prepare cell beads*.

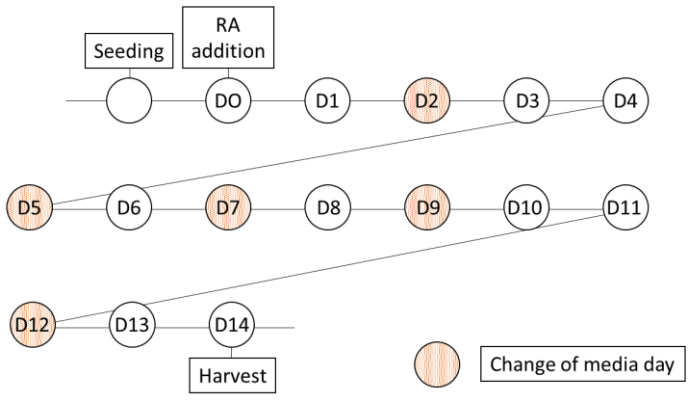


Figure 12. PA-1 differentiation in large scale scheme. *Details are provided in the text.*

16. Cell collection and freezing to prepare cell beads.

At least 15 plates were cultured to obtain pluripotent PA-1 (PC) and at least 20 plates were cultured to obtain differentiated PA-1 (DC) using the culture conditions described in *2. Cell line culture*. Once the cells reached 85-90 % confluency they were harvested as described below according to the protocol of (LaCava *et al, 2016)*. The collection does not have to be performed under sterile conditions.

1. The culture medium from the plates is poured into a large beaker and the culture plate is placed on a tray with ice.
2. Ice-cold phosphate buffered saline (PBS) solution (20 ml) is added to the culture plate and the cells are detached from the plate using a cell scraper; the cells are then transferred to a 50 ml tube previously chilled on ice, and kept on ice.
3. The remaining cells are then collected by pouring an additional 10 ml of ice-cold 1x PBS onto the plate and transferred to the 50 ml tube.
4. The above steps are repeated for each plate; cell suspensions from different plates can be combined to reduce the number of samples and plastic waste. Up to three cell suspension plates can be combined into two 50 ml tubes.
5. Subsequently, the cells are centrifuged for 5 min at 1,000 x g, 4 °C.
6. Carefully pour the supernatant into the waste beaker. Then resuspend each cell pellet in 10 ml of ice-cold 1x PBS. Combine the different pellets with the same volume of 1x PBS until they are all transferred to the same 50 ml tube.
7. Again, centrifuge 5 min at 1,000 x g, 4 ° C.
8. Carefully, pour the supernatant into the beaker for waste. At this point, with all the pellet in a single tube:
    - Remove the plunger from a 20 ml syringe and cap the tip of the syringe.
    - Transfer the cell suspension into the syringe.
    - Place the syringe into a 50 ml tube and centrifuge 5 min at 1,000 x g, 4 ° C.
9. Once the cell pellet is in the syringe, remove the supernatant with a fine-tipped pipette attached to a vacuum trap system until the upper layer of cells starts to be sucked out. This results in a wet cell pellet.
10. Finally, insert the plunger into the syringe and drop the cells dropwise into a large plastic beaker filled with LN2, kept in an LN2 bath, in a Styrofoam box. The cells are frozen in the form of "beads" or spheres of cells.
11. The frozen cells are transferred to a 50 ml tube. The cap of the tube should not be screwed on completely to allow the excess LN2 to evaporate. The

next day the cap should be screwed on completely. The frozen cells can be stored at -80 °C.

17. Cryomilling of frozen cells.

The procedure of cryomilling of frozen cells is described in (LaCava *et al, 2016)*. In order to carry out the cryogenic grinding of cells, different tools and equipment adapted to the RETSCH Planetary Ball Mill PM100 cryogenic grinding system, are necessary.

Briefly, the cryomilling process is carried out in cold, using LN_2 to ensure that the friction and grinding to which the sample is subjected do not cause its deterioration due to heating. It is key to keep the sample at a low temperature to ensure that it remains frozen and that the molecular interactions between the cellular components are preserved.

For this reason, all milling and handling tools are pre-cooled with LN_2. After pre-cooling the entire sample housing system, the crushing jar, two 20 mm balls and the jar lid, the cell "beads" or "spheres" are introduced into the jar, LN_2 is added again to the entire system, equilibrated and the system is started.

Three trituration cycles were carried out using the following program: 400 rpm, 3 min, reverse rotation every minute, with no break intervals. Between each cycle it is necessary to re-cool the crushing jar by adding LN_2 into the system.

Due to small losses of material on the surfaces of the jar and ball during grinding, the percentage of recovered material increases as the mass of the milled cells increases. For this reason, it is recommended to start from 2.5-3 g in order to reduce the loss of material. The amount of lost material is usually of the order of ~ 0.3 g wet cell weight (WCW).

Once cryomilling is complete, the system needs to be carefully opened, the jar placed in LN_2 and allowed to cool. Then, while still in LN_2, carefully remove the lid, remove the metallic beads using tweezers and transfer the powder to a pre-cooled 50 ml tube with a pre-cooled spatula. Once the jar has been opened, a small amount of LN_2 should be added before removing the beads in order to help to recover the powder clumped on the surface. The cell powder should be kept at -80°C or below until use and can be stored in this manner indefinitely without affecting performance.

18. Antibody coupled magnetic beads preparation

The magnetic Dynabeads used for affinity capture (Dynabeads M-270 Epoxy - Invitrogen, Cat. 143.02D) were prepared using 1 mg of antibody per 100 mg of beads. For this purpose, a mixture of 1 mg of antibody in a total volume of 2 ml was prepared in a 1 M solution of $AmSO_4$ and 0.1 M $NaPO_4$, pH 7.4.

Subsequently, the antibody mixture was filtered using a Costar Spin X column (0.45 um). The antibody mixture was then added to beads previously equilibrated in 0.1 M $NaPO_4$ pH 7.4, and incubated at 37 degrees overnight. The next day the beads were washed following the protocol of Michael P. Rout's laboratory (Conjugation of Dynabeads with Rabbit IgG, Rout Lab Protocol). Finally, the beads were resuspended in 2 ml of 50% glycerol, 0.5 mg/ml BSA in 1X PBS, and aliquoted (100 µl aliquots) and stored at -20 ºC.

19. Affinity capture of protein complexes from cell powder

Depending on the experiment, the exact conditions used for affinity capture were slightly modified. The general description of the process is detailed below, with the specific details followed for each experiment indicated in each case.

Samples were kept on ice during the whole procedure. All the affinity captures are started by weighing a certain amount of cell powder, and experiments were carried out with a scale of 25, 50, 100, 200 or 250 mg of material. In each case, volumes and sonication times were applied proportionally to those indicated below, equivalent to those used for a starting material of 50 mg.

1. First, 50 mg of cell powder were weighed into a 1.5- or 2-ml microcentrifuge tube. An analytical balance was used to weigh the empty microcentrifuge tube. This tube was cooled in LN_2, then the balance was tared and the cell powder was added with a spoon or spatula cooled with LN_2.

2. Once weighed, the tube with cell powder was opened and allowed to stand at room temperature for 1 minute (to release pressure and prevent freezing of the extraction solution).

3. Next, 200 µl of extraction buffer (20 mM HEPES, pH 7.4, 1% TritonX-100, 500 mM NaCl) supplemented with protease inhibitors were added and the mixture was shaken with a vortexer.

4. Once the cell powder was resuspended in the extraction solution, an

ultrasonicator QSonica Q700 coupled to a microtip (1/16") was used to give the sample a brief low energy pulse to disperse any aggregate. A sonication of 5 pulses (2 seconds each, 2 A; approximately 15-20 J of total energy was applied to the sample) was performed.

5. Subsequently, the extract was clarified by centrifugation. It was centrifuged at 20,000 x g for 10 min at 4 ° C.

   An aliquot of the cell extract (20 µl of supernatant) and the cell pellet was stored (- 20 ° C).

In order to prepare the magnetic affinity medium, a series of pre-washes to the magnetic beads coupled to antibodies were performed:

Magnetic beads coupled to antibodies are stored in a solution with albumin (BSA) in 1x PBS, being necessary to wash and equilibrate them with extraction buffer before use. A tube rack with a magnet is used to precipitate the magnetic beads. Thus, when the tubes are placed in the tube rack, the magnetic beads accumulate on the side of the tube in a few seconds, allowing the storage solution to be removed.

Depending on the type of magnetic beads used for each experiment, a volume of 2.5, 5, 10 µl of beads was used for 50 mg of cell powder.

To wash the magnetic beads, 500 µl of extraction solution are added and then briefly vortex mixed at medium speed (enough to resuspend). This is followed by a pulse in a mini centrifuge to collect all the content at the bottom of the tube. Subsequently, the tube is placed in the magnet rack and the solution is aspirated. This washing is carried out three times.

6. Once the magnetic beads are prepared, affinity capture is initiated by transferring the clarified cell extract to a 1.5-2.0 ml tube containing pre-washed affinity medium and is briefly shaken to resuspend.

7. The tube is incubated for 1 hr at 4 °C with continuous gentle agitation on a rotating wheel; the beads should remain suspended throughout the incubation.

8. Once the incubation is finished, a pulse is given to the tube in a minicentrifuge to collect all the content at the bottom of the tube. The supernatant is then collected (as FT, "flow through") and the beads are washed three times with 1 ml of cold extraction solution.

The tubes are placed in the magnetic rack, the solution is removed, new

solution is added and the process is repeated.

During the second wash, the beads and wash solution are transferred together to a new microcentrifuge tube by pipetting (to reduce contamination in the elution step).

After the third wash, the tube is pulsed in a minicentrifuge to collect all the content at the bottom of the tube. The tube is returned to the magnetic rack and the residual liquid is removed. This step is important to remove the last µl of solution before the elution to ensure that the eluted samples have uniform volumes and the elution solution is not diluted or contaminated.

9. Elution was carried out with two different solutions, 15 ul Tris pH 8.0, 2% SDS or in LDS Sample buffer (Invitrogen), depending on the intended use of the sample. After adding the solution on the magnetic beads, it was incubated at 70 °C for 5 minutes to elute.

10. The eluate was collected by placing the tube in the magnet rack and pipetting the supernatant into a new tube. Samples were stored frozen at -20 °C for daily use or at -80 °C for prolonged storage.

20. Generation of inducible Tet-On PA-1 cell lines

In order to develop PA-1 cell lines carrying the doxycycline-inducible Tet-On system, cells were transfected with the vectors LD208/PM65 - pTet-On rtTA advance G418 (ClonTech), with selection by geneticin, and the vector LD215/PM66 - pTet-On rtTA advance BSD (modified from the previous one), with selection by blasticidin. The pool of transfected cells was diluted with different factors (1/10, 1/50, 1/100) in 100 cm2 plates to obtain single cells, which were grown for 2 to 3 weeks to form colonies. Subsequently, clones were isolated and those that integrated the pTet-On system were selected by the experiments described in the *VII. RESULTS, 2.1.1. Generation of inducible TetOn Advanced (Clontech) PA-1 cell lines to overexpress LINE-1 proteins.* Section.

21. Sample preparation and liquid Chromatography coupled to Mass spectrometry

After performing affinity capture of L1-ORF1p and its interactors, the resulting eluate was prepared for liquid chromatography coupled to mass spectrometry (LC-MS/MS) by Dr Kelly Molloy.

First, the eluate was reduced by addition of DTT to a final concentration of 25 mM and incubation at 37 °C for 30 min. Subsequently, alkylation of the sample was carried out by addition of iodoacetamide (100 mM final concentration) and incubation in the dark for 30 minutes.

Afterwards, phosphoric acid (1.2 % final concentration) was added to the lysate solubilized in SDS buffer and from this point the S-TrapTM Micro Ultra-High Recovery Protocol was followed. The resulting dehydrated peptides were stored at -80 °C until use.

Dry down peptides in speed vac, store at -80ºC. Resuspend as desired (buffer A or MALDI matrix). The peptides were gradient eluted directly into a Q Exactive (Thermo Fisher) mass spectrometer by Dr Kelly Molloy, and acquired high resolution full MS spectra (similar to Taylor *et al, 2018)*.

22. Mass spectrometry data analysis

The mass spectrometry raw data were analyzed by Dr Mehrnoosh Ogbaie employing the MaxQuant V. 1.6.7.0 proteomic pipeline (Tyanova *et al, 2016)* to quantify the peptides and identify the proteins. Data processing yielded LFQ intensity data (Label Free Quantification, normalized peptide intensity detected) as well as a count of the number of peptides identified for each protein in question (physical evidence).

To identify the detected L1-ORF1 loci, orthogonalization of the data was conducted using the Uniprot FASTA database with the ORF1/ORF2 sequences, as well as the sequences of the ORF1/ORF2 loci from L1Base2 (Penzkofer *et al, 2017)* when using MaxQuant software. Those L1-ORF1 peptides not derived from the ORF1 consensus sequence (Q9UN81) were filtered out and no peptides from ORF2 were identified.

In order to analyze the data and identify those proteins that significantly interact with L1-ORF1p in the case replicates (those performed with antibodies against ORF1) and were not identified in the controls (using antibodies that do not detect proteins expressed in PA-1), two sample t-test and log fold change were

used. However, if data have missing/censored values (where a peptide is not detected or did not hit the minimum threshold) t-test can't be performed. In order to prevent this, data imputation was carried out in order to reduce the missing/censored values of those proteins not identified in all the conditions (case or control). For this purpose, two different imputation methods were used: **Method 1**, more stringent, where missing/censored values in records with at least two existing replicates were imputed with a relative semi-random value (explained in detail in Dou et al, 2020*)*; or **Method 2**, less stringent, where missing/censored values in records with at least one existing replicate were imputed with a relative semi-random value. For those cases where we had only two control replicates (PA-1 DC mαORF1 experiment, where only two controls of mIgG were performed), when both replicates had missing values, these were imputed with small random values.

The analyses by two sample t-test and fold change obtained from the data imputed by both methods were combined: those factors not significant in Method 1 were filtered out but the analyses obtained from Method 2 were considered to rank the factors and compare between the different experiments (PC PA-1 mαORF1, PC PA-1 LαORF1 and DC PA-1 mαORF1). Regarding phosphorylation data, phosphorylation analysis from the imputed data was carried out (from MaxQuant data relative to peptide phosphorylation). Those sites (sites within a protein where phosphorylation was detected) where log2 fold changes > 1 were defined as differentially phosphorylated.

In order to determine whether a host factor is significantly enriched in PC PA-1s or in DC PA-1s, we used a t-test to calculate the p-value between the replicates from each record (p-value), and the Benjamini Hochberg method to calculate the adjusted p-value (p-adjusted) from p-values. Next, using the average log LFQ intensities between the replicates from each record, we calculated log2 fold changes. From these analyses, those proteins whose p-adjusted value were <0.05, whose log2 fold change were >1, with at least one physical evidence detected by MS/MS, and when at least two replicates were available, were considered as significantly different among PC and DC. The data from both PC PA-1 (mαORF1p), DC PA-1 (mαORF1p) and PC PA-1 (LαORF1p) interactomes was compared generating a list of factors mostly present in PA-1 PC.

The volcano and scatter plots were made from the t-test data obtained from the application of Method 2. Significant data resulted from conducting t-test of the imputed data was used to color code the points and to label them. The color cold is: black dots, significant at least in one condition; red dots: ORF1 sequences either consensus or cancer loci; grey dots; not significant proteins. In order to exclude outliers (proteins that were observed in only one condition), linear regression was applied. In addition to linear regression line, sigma statistic was used to draw the line (+/- sigma, dashed blue line) to find the significant proteins that were differentially expressed.

23. LINE-1 Expression analyses by RT-qPCR

LINE-1 expression analyses by RT-qPCR using primers directed to the 5′UTR (N-51) or L1-ORF2p (N-22) sequences were conducted exactly as described in Muñoz-Lopez *et al, 2012*. Briefly, total RNA from pluripotent and differentiated PA-1 cells was extracted using Trizol (Invitrogen), following manufacturer instructions. Total RNAs (2mg) were treated, twice, with 4 Us of RNase-free DNase I (Invitrogen) during 30 min at room temperature, to ensure removal of genomic DNA. After DNase I inactivation, Reverse Transcription was carried out with 1 μg of DNase I-treated total RNAs using a High-Capacity cDNA Reverse Transcription kit (Applied Biosystems), following the instructions provided by the manufacturer. Next, we conducted qPCR in triplicate using cDNAs that were diluted at 1/5 and 1/10 using RNA/DNA-free purified water (Invitrogen). qPCR reactions were carried out using Brilliant SYBR Green QPCRMix (Stratagene) in 20ml volumes, using two previously validated set of primers directed to the 5′end of the L1Hs consensus sequence (N-51 pair, FWD 5′-GAATGATTTTGACGAGCTGAGAGAA, REV 5′- GTCCTCCCGTAGCTCAGAGTAATT) or to L1-ORF2p (N-22 pair, FWD 5′- CAAACACCGCATATTCTCACTCA, REV 5′-CTTCCTGTGTCCATGTGATCTCA) (Muñoz-Lopez *et al, 2012;* Coufal *et al, 2009)*. We used a MX3005P Real-Time PCR machine (Stratagene) and included a melting curve from 50°C to 95°C with reads every 0.2°C, to confirm the identity of amplicons. To normalize expression, we amplified GAPDH as a housekeeping control (primer sequences can be found in Muñoz-Lopez *et al, 2012),* and we used the ΔΔC(t) method to represent results (i.e., the cycle threshold (C(t)) for GAPDH PCR was used to normalize) (Livak & Schmittgen, 2001).

24. DNA-methylation analyses of the LINE-1 promoter

As above, we used a method previously validated by the lab, and we conducted methylation analyses exactly as described (Muñoz-Lopez *et al, 2012;* Coufal *et al, 2009)*. Briefly, genomic DNA from pluripotent and differentiated PA-1 cells was extracted using the Blood&Tissue DNeasy Mini kit (Qiagen), following manufacturer instructions. Next, we bisulfite converted genomic DNAs (2mg each) using the Epitect kit from Qiagen, following the protocol provided by the manufacturer. Converted genomic DNAs were purified using buffers and columns provided in the Epitect kit (Qiagen), and we used 500ng of each converted genomic DNA on duplicated PCRs. In the PCRs (50ml), we amplified a portion of the L1Hs consensus sequence promoter/5'-UTR [a region of 363 bp containing 20 CpGs as described (Coufal *et al, 2009)*], using Taq polymerase (Invitrogen) and the following primers: FWD: 5'-AAGGGGTTAGGGAGTTTTTTT; REV: 5'-TATCTATACCCTACCCCCAAAA (Sigma). Cycling conditions were as follow: 2 min at 95ºC (1x); 30 sec at 94ºC, 30 sec at 54ºC, 60 sec at 72ºC (x35); 5 min at 72ºC (1x). Amplified products were resolved on 1% agarose gels (containing SYBRgreen (Sigma)), excised and purified using the QIAquick gel extraction kit (Qiagen), and cloned on pGEM-T Easy (Promega). We used X-gal/IPTG to screen clones with insert, and we Sanger sequenced 50 clones from each condition (pluripotent and differentiated PA-1). Unique sequences were analyzed using Repeatmasker (http://www.repeatmasker.org/cgi-bin/WEBRepeatMasker), confirming that >95% of clones contained sequences derived from the L1Hs subfamily of elements. Next, we aligned each sequenced amplicon with a L1Hs consensus sequence, to determine the fraction/percentage of methylated/unmethylated CpGs in each of the 20 sites found in the LINE-1 promoter/5'-UTR. Additionally, and to graphically represent the overall methylation level found in pluripotent and differentiated L1Hs promoters, each sequenced amplicon was aligned with an active L1Hs element (L1.3, accession number L19088.1), and we used the amplicons (n=10) with the highest homology to L1.3 to indicate whether CpGs were methylated (black circles) or unmethylated (white circles).

25. Microarray and GO expression analyses

Microarray expression analyses were conducted by MACS Molecular (Germany), using the Agilent Whole Human Genome Oligo Microarray chip (one-color). Briefly, we compared the expression profile of pluripotent and differentiated PA-1 cells in duplicate, and using two distinct methods of PA-1 differentiation that were previously validated (Garcia-Perez *et al, 2010)*. Total RNAs were used in the microarray, and these were extracted from confluent cultures using Trizol (Invitrogen), following manufacturer instructions. After recording intensities, ratios were computed using both "pluripotent" samples as common reference, and ratio datasets were used as basis for subsequent analyses. Ratios were computed using the RosettaResolver™ Software (Rosetta Inpharmatics); this software system has the advantage that, due to the implemented universal error model, not only the ratios are computed, but also associated p-values which are a measure of the reliability of the observed difference between two samples. For the computation of ratios, the two "pluripotent" control samples were merged in silico and used as common reference for each "differentiated" sample. All ratio data were transformed to logarithms to the base 2 (log2 ratio). In addition, for each ratio the corresponding "fold-change" was computed for a more intuitive understanding of the expression changes.

As an initial step in the analysis, the different data sets were compared by a global correlation analysis, in order to get an impression on inter-sample similarity or variability. The correlation coefficients range from -0.592, indicating anticorrelation between the two "pluripotent" samples, to 0.853 in the comparison of two "differentiated" samples (i.e., DM w/o RA samples). Anticorrelation between samples part of a common reference is normal and has no meaning in terms of sample similarity. Notably, the dataset splits into two major sample groups, one containing the "differentiated cell samples ("DM" and "DM w/o RA"), the other one comprising the "pluripotent" cells. Each of the two differentiated cell populations forms a subgroup within the differentiated cell cluster. Based on the results of this analysis, there seem to be some differences in the differentiated cells depending on the culturing conditions they were exposed to. Obviously, the major changes in gene expression were between "pluripotent" and the two "differentiated" states.

For the detection of differentially regulated genes, statistical algorithms such as t-test, ANOVA (analysis of variance), or SAM (significance analysis of microarrays) were the methods of choice. Such algorithms are sensitive to the

magnitude of differences between the sample groups tested and/or to the variability within each of the groups. If the range of expression values covered within one experimental group is too large in comparison to the expression differences to the other experimental group, a gene may not be identified as differentially regulated.

Notably, the variability between each two samples of the replicate pairs does not seem to be too high. However, due to the low number of samples (N = 2), we decided to use a rather relaxed algorithm for the discriminatory genes analysis accepting a potentially higher number of false positive candidates. A t-test was performed (two-tailed, assuming unequal variance, no correction for multiple testing) to determine the significance of expression differences between the "differentiated" samples (both for "DM" and "DM w/o RA" samples), and the "pluripotent" samples. Genes reaching p values of 0.05 or below were called significant (i.e., 5% significance level). Prior to the analysis, no filtering for strongly regulated genes was performed in order to allow also genes with slight but significant changes to be selected. In addition, the two groups of "differentiated" samples were tested for significant expression differences. Finally, genes were selected for the group of "genes of interest" which are differentially regulated in the same orientation in both "differentiation" conditions when compared to the "pluripotent" samples but not between the two "differentiation" conditions (i.e., "DM" and "DM w/o RA") in the direct comparison. This approach returned **808 commonly upregulated** genes and **568 commonly downregulated** genes in the "differentiated" cells compared to the "pluripotent" reference. For clustering, Euclidean distance was chosen as similarity metric and average linkage as algorithm. Excerpts of this heat map showing a selection of up- and downregulated genes are shown in **Figure 48B**. Note that in the heat map, the gene expression changes (as log2 ratios) are color-coded, where Red stands for a relative upregulation of a gene in the sample compared to the control, while Green illustrates downregulation.

To plot Functional Grouping of consistently regulated genes, in order to learn about the molecular processes changed in "differentiated" cells compared with "pluripotent" cells, differentially regulated genes were subjected to an extensive annotation with subsequent grouping into different categories. Up- and downregulated genes were analyzed separately. Genes were annotated with information on their molecular function (based on Gene Ontology; GO) as well as involvement in biological signalling pathways.

Exemplary bar charts of the categories in "biological processes" are shown in **Figure 48D**. Please note that the number of genes in the specific categories does

not allow a direct judgment of a statistically significant enrichment. For example, highly populated categories (found at the top of the bar charts) may arise due to the fact that these categories comprise a higher number of genes in the first place.

VII. RESULTS

1. RNase H2 and LINE-1 retrotransposition

As previously discussed *[III. INTRODUCTION, 5. Aicardi Goutières Syndrome (AGS)]*, it is currently unknown whether RNase H2 would play a role in regulating LINE-1 expression and/or retrotransposition, as other Aicardi Goutières Syndrome (AGS) genes do. RNase H2 has two well characterized catalytic activities: i) the degradation of the RNA strand of DNA:RNA hybrids, similar to those generated during Target Primed Reverse Transcription (TPRT) retrotransposition; and ii) the removal of embedded ribonucleotides miss incorporated in genomic DNAs. To note, and as most vertebrate LINE-1s, human RC-L1s lack a RNase H domain and rely on cellular activities to remove the RNA of DNA:RNA hybrids generated during retrotransposition. In consequence, the first part of this Thesis focuses on studying the regulatory role of RNase H2 on human LINE-1 retrotransposons. The results of this section were published in 2018 (Benitez-Guijarro *et al, 2018)*, a study conducted in collaboration with the lab of Prof Andrew P Jackson (IGC, Scotland, UK), a world known leader in the study of RNase H2 and AGS.

1.1. L1 retrotransposition is compromised in RNase H2 knock out (KO) mutant cell lines.

To test whether RNase H2 regulates LINE-1 expression/retrotransposition, we generated a panel of HeLa RNase H2 knock out (KO) cell lines using CRISPR/Cas9-mediated genome editing, and guide RNAs (gRNAs) targeting the RNASEH2A subunit (following the methodology detailed in *VI. MATERIALS AND METHODS, 8. Generation of KO cell lines and RNase H2 mutants*) (**Figure 13A**). After transfecting with CRISPR/Cas9 plasmids, we generated a panel of single-cell clonal cell lines using FACS-sorter, taking advantage of an EGFP expression cassette present in the backbone. Next, individual clonal lines were selected based on deletions and insertions observed after PCR amplification and sequencing of the RNASEH2A locus targeted by CRISPR/Cas9. In these analyses, we identified several clonal lines containing inactivating mutations in several RNASEH2A alleles per cell line. However, because HeLa cells are karyotypically unstable, to verify loss of RNASEH2A expression we used Western blotting; we specifically explored expression of the three RNASEH2 subunits (A, B and C) by Western blotting in KO cells and controls (**Figure 13B**). Notably, we identified several clones lacking expression of RNASEH2A, which also showed reduced levels of RNASEH2B and C subunits, consistent with instability of RNASEH2 when expression of any of the three subunits is compromised (Reijns *et al*, 2011).

Next, we characterized the KO clonal lines using an established functional RNase H2 assay. As a control to compare with the KO clones, parental cells as well as CRISPR control clones (C) retaining significant RNase H2 expression were used in these and subsequent experiments.

In KO clones, the absence of enzymatic activity able to remove ribonucleotides embedded in double-stranded DNA was confirmed by a FRET-based fluorescent substrate release assay, using cell lysates from KO cell lines and controls (**Figure 13C**). Furthermore, we detected an increase in genomic DNA fragmentation for KO clones after treatment with recombinant RNase H2 followed by electrophoresis in alkaline gel. In fact, we were able to detect the presence of a large number of embedded ribonucleotides in genomic DNAs from KO clones (**Figure 13D**), which is a well-known consequence of RNase H2 deficiency (McElhinny et al, 2010; Hiller *et al*, 2012; Reijns *et al*, 2012).

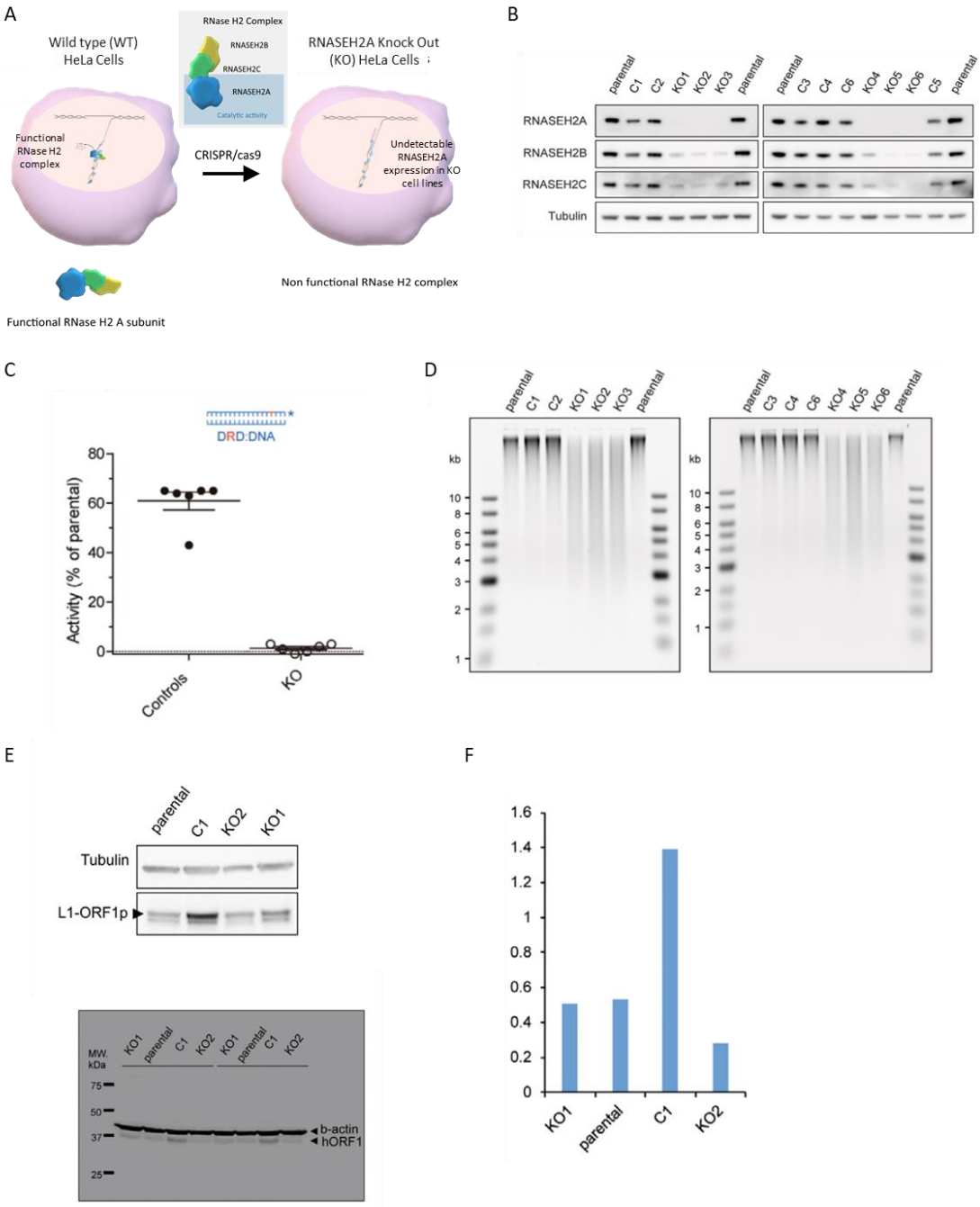


Figure 13. RNase H2 knock out cell lines generation by CRISPR/Cas9. A) *RNase H2 knock out (KO) cell lines were generated by CRISPR/Cas9-mediated genome editing and guide RNAs (gRNAs) targeting the RNASEH2A subunit. The KO resulting cell lines lacked RNASEH2A expression and also showed reduced levels of RNASEH2 B and C subunit expression.* ***B)*** *Western*

blot analyses show absence of RNASEH2A and depletion of RNASEH2B and C in RNASEH2A-KO clones (KO1-6) when compared with control clones (C1- 5) or parental cells. Tubulin was used as a loading control. ***C, D)*** *RNase H2 enzymatic activity in KO clones and CRISPR/Cas9-generated controls. C) The RNase H assay shows the absence of activity against single-embedded ribonucleotides in KO clones, while activity in all control clones was in general unaffected. Activity in parental HeLa cells was set to 100%. Data points indicate the mean of three technical replicates for individual clones. Lines indicate the mean of six biological replicates (C1-6 and KO1-6) ± SEM. D) High levels of ribonucleotides embedded in the genome from KO clones. Genomic DNAs isolated from parental cells, KO clones and control clones were treated with recombinant RNase H2 and resolved by alkaline gel electrophoresis. Smaller fragments indicate a higher number of embedded ribonucleotides in genomic DNAs.* ***E)*** *Western blot analysis of L1-ORF1p expression (in triplicate) in parental HeLa cells, a control clone (C1) and two KO clonal lines (KO1 and KO2), shows no indirect effect of RNase H2 deficiency on L1 expression. Tubulin (E) or β-actin (F) whereas used as loading controls.* ***F)*** *Quantification of Western blot from panel (C) using a LI-COR device following manufacturer's instructions. Adapted from Benitez-Guijarro et al, (2018).*

To start exploring whether RNaseH2 could be involved in regulating human LINE-1s, we initially explored changes on L1 expression. Briefly, we analyzed endogenous L1 expression levels in KO clones and controls, using Western blot analyses and employing a specific antibody against L1-ORF1p. Although there was some variation between clones, we confirmed that L1-ORF1p is expressed at similar levels in RNASEH2A-KO cells and parental cells (**Figure 13E and F**). Thus, these data indicated that RNase H2 doesn't affect L1 expression levels.

Next, we analyzed potential changes on L1 retrotransposition; to do that, both the KO and the control clones (C) were used in L1 retrotransposition assays. The reporter gene used in these assays was the *mblastI* reporter (Morrish *et al, 2002;* Goodier *et al, 2007)*, which activates expression of a blasticidin-resistant gene after retrotransposition (**Figure 14A**, plasmid JJ101/L1.3). It is important to note that this assay is quantitative, and the resulting number of blasticidin-resistant colonies provides a readout of the overall frequency of retrotransposition in cultured cells (Morrish *et al, 2002;* Goodier *et al, 2007)*, as each colony/foci represents a cell with at least a *de novo* retrotransposition event (i.e., a new LINE-1 insertion). Surprisingly, in these assays we observed that LINE-1 retrotransposition was severely reduced in all RNase H2 KO cell lines tested (n = 6 clones), with an average retrotransposition level of 6.0 ± 2.6% (mean SD) in comparison with control cell lines (n = 6, 94 ± 22%; P = 0.0022) or parental cells (adjusted to 100%; **Figure 14B and C**). As an internal negative retrotransposition control, allelic L1 plasmids containing a missense mutation in the RT domain of L1-ORF2p (JJ101/L1.3-D702A) were used, and as expected, they were unable to mobilize in all cell lines tested (**Figure 14C**).

As an additional control, we next tested whether KO and control lines could generate antibiotic resistant foci at the same frequency. Briefly, transfection with pcDNA 6.1 (which constitutively express the blasticidin resistance gene, **Figure 14C**) allows to determine the ability of cell lines carrying an antibiotic resistance gene to form colonies at the antibiotic concentration (blasticidin) used. In fact, the comparison between the number of colonies formed with pcDNA 6.1 and with JJ101/L1.3 allows to determine the rate of LINE-1 retrotransposition, considering the colonies formed by pcDNA 6.1 as the maximum number of colonies that a cell line can form. However, we observed that similar numbers of blasticidin-resistant colonies were generated in all cell lines after transfection with the toxicity/clonability control vector (pcDNA6.1, **Figure 14C**).

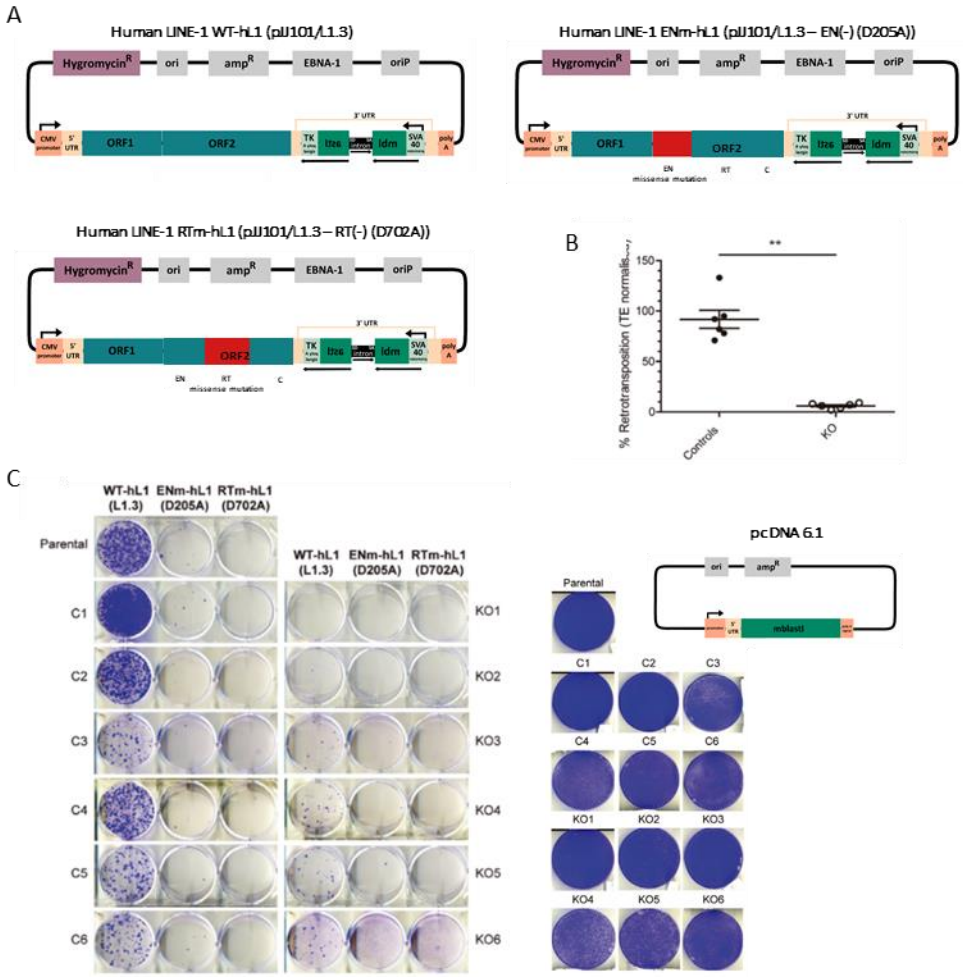


Figure 14. Reduced LINE-1 retrotransposition in HeLa RNase H2 null cells. A) *Schematic of the JJ101/L1.3 (WT-hL1/L1.3) retrotransposon vector and its variants carrying different L1-ORF2p mutations. The relative positions of the EN (endonuclease), RT (reverse transcriptase) and C (cysteine-rich) domains are indicated within L1-ORF2p. The green box labeled BLAST backward represents the mblastI retrotransposon indicator cassette.* ***B)*** *Quantification of human wild type L1 (WT-hL1/L1.3) retrotransposition, normalized to the parental cell line level and normalized for transfection efficiency (TE), set to 100% for comparison. Data points represent the mean of three technical replicates for individual clones. Lines indicate the mean of six biological replicates (C1-6 and KO1-6) ± SEM (representing six independent experiments). Mann-Whitney test; **P < 0.001.* ***C)*** *Representative retrotransposon assay performed on parental cells, control clones (C1-6) and RNASEH2A-KO clones (KO1-6). Cells were transfected with JJ101/L1.3 derived vectors containing an active human LINE-1 (WT-hL1, L1.3 element), an EN mutant LINE-1 (ENm-hL1, L1.3 D205A) or an RT mutant LINE-1 (RTm-hL1, L1.3 D702A). Toxicity controls: a similar number of blasticidin-resistant colonies were generated in all cell lines after transfection with the control vector* pcDNA*6.1 (scheme). Representative results of transfection/selection experiments in parental HeLa cells, control clones (C1-6) and KO clones (KO1-6) are shown. Adapted from Benitez-Guijarro et al, (2018).*

Overall, the above data indicate that L1 retrotransposition is severely compromised in HeLa cells lacking RNase H2 activity. To further explore this apparent need of RNase H2 activity at a mechanistic level, we next explored whether DNA damage accumulated in RNase H2 KO cells could be involved in L1 retrotransposition. To do that, we tested retrotransposition of an EN mutant LINE-1 allelic plasmid (JJ101/L1.3-D205A), and we observed that retrotransposition was also low in parental/control and KO cells, suggesting that DNA lesions in KO cells are not used as integration sites by LINE-1 (WT or EN mutated) (Reijns *et al, 2012)* (**Figure 14C**). Therefore, endonuclease independent LINE-1 retrotransposition does not appear to be increased in KO cells relative to control cells *(*Morrish *et al, 2007, III. INTRODUCTION, 2.2. The LINE-1 retrotransposition mechanism).*

To confirm reduced retrotransposition in RNase H2 KO cells, we next used a retrotransposition assay that activates luciferase expression after integration (*mflucI* cassette, see **Figure 15A**). Briefly, we used a cassette containing the firefly luciferase gene interrupted by an anisense intron that undergoes splicing during retrotransposition, ensuring that luciferase expression can only be activated after a round of retrotransposition event. This assay has two advantages over the antibiotic resistance-based assay: (i) it does not require antibiotic selection or the generation of resistant colonies, and (ii) it allows to measure retrotransposition levels at earlier time points (as early as 96 h after transfection). In this way, long-term culturing of mutant cells is avoided, and, therefore, any possible effects due to differential growth between controls and RNASEH2A-KO cell lines can be ruled

out. Consistent with our initial observations, using the luciferase-based L1 assay we observed a 63% reduction in LINE-1 retrotransposition in RNase H2 KO cell lines (n = 3, 31 ± 6.2%) when compared with controls (n = 3, 84 ± 8.3%; P = 0.0009; **Figure 15B**).

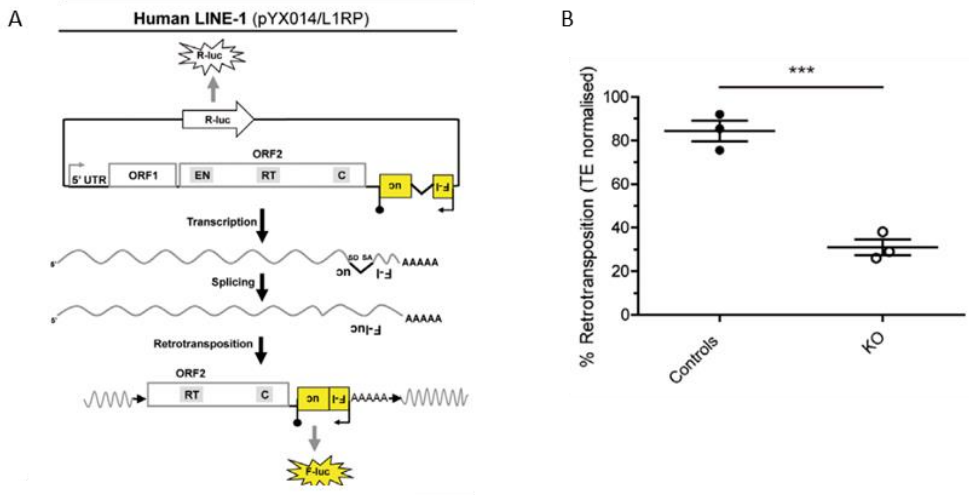


Figure 15. Reduced LINE-1 retrotransposition in HeLa RNase H2 null cells. *Luciferase retrotransposition assay. A) Rationale and schematic of plasmid pYX014. Plasmid pYX014 contains a human RC-L1 tagged with a luciferase retrotransposition indicator cassette (yellow box with a backward F-luc label; its promoter is noted with a black arrow and its polyadenylation signal is depicted with a black lollipop). Using this plasmid, retrotransposition of L1 activates expression of Firefly luciferase. The plasmid backbone contains an expression cassette for Renilla luciferase, which allows to normalize transfection efficiency (large white arrow with R-luc label). Following transfection of the pXY014 plasmid into cells, the L1 mRNA is spliced by canonical cis-splicing and undergoes retrotransposition, leading to the activation of the firefly luciferase reporter and the subsequent translation of the F-luc protein (yellow star with F-luc label). At the same time, the R-luc cassette of the plasmid can be transcribed and translated into Renilla luciferase (R-luc-tagged white star). Within the retrotransposition event shown, the black arrows indicate the presence of TSDs flanking a 5′ truncated L1 insertion. B) Results from retrotransposition assays conducted in HeLa parental cells, three control clones (C) and three RNASEH2A-KO clones (KO). The level of retrotransposition in parental cells was set 100%. Dots represent the mean of three technical replicates for individual clones. The lines indicate the mean of three biological replicates (C2, 4 and 5, and KO2-4) ± SEM (representative of three independent experiments). t-test, ***P < 0.001. From Benitez-Guijarro et al, (2018).*

The data observed in these experiments suggest that, contrary to expectations, RNase H2 would not inhibit LINE-1 expression/retrotransposition, but rather it seems to promote human RC-L1 retrotransposition, at least in cultured HeLa cells.

To determine whether RNase H2 activity is required for LINE-1 mobilization in other cellular environments, we also analyzed retrotransposition in colon carcinoma (HCT116) and osteosarcoma (U2OS) human cells. Thus, we employed the same CRISPR/Cas9 strategy to generate RNASEH2A KO clones and controls, using HCT116 cells lacking p53 expression (HCT116 p53-/-). As above, we performed Western blot, enzymatic activity assays and analysis of ribonucleotide incorporation into the genome to confirm the absence of RNase H2 activity in two clonal KO cell lines (KO, **Figure 16A-C**); several control clones were also analyzed in parallel (Control, **Figure 16A-C**).

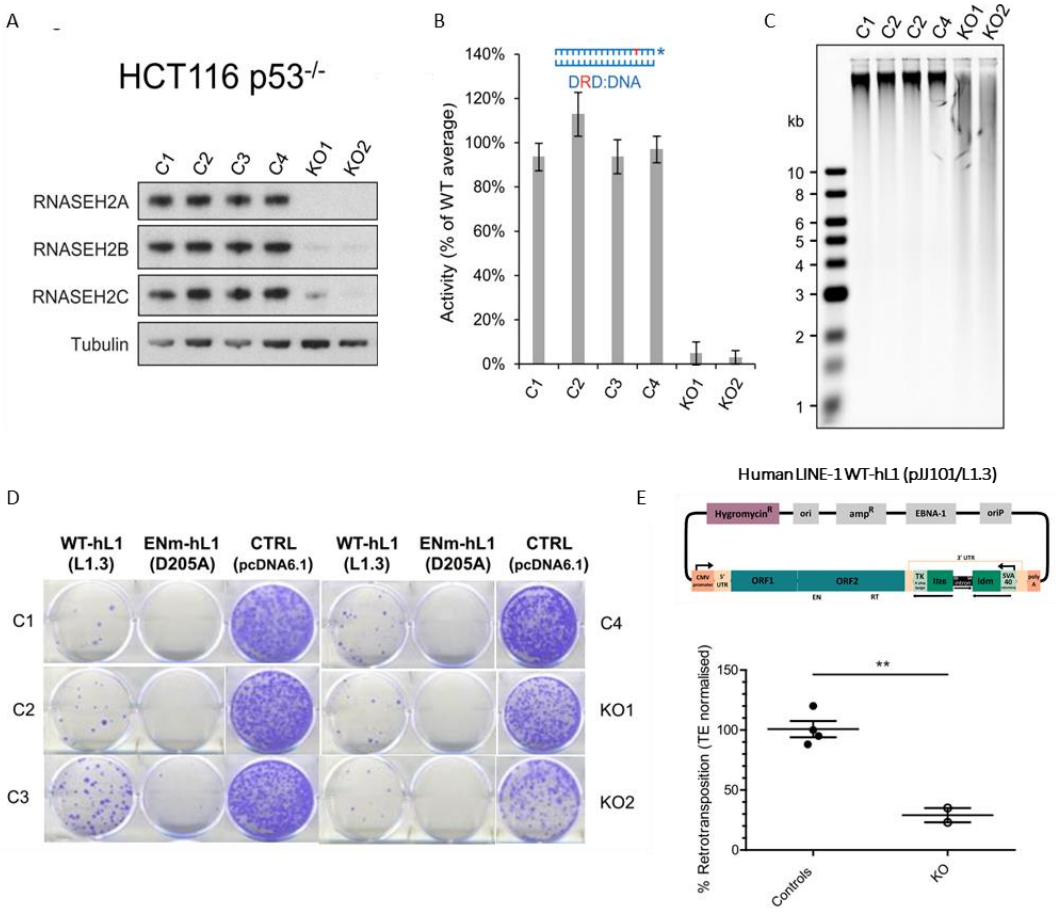


Figure 16. Reduced LINE-1 retrotransposition in RNase H2 null HCT116 p53-/- cells. A) *Western blot analyses show absence of RNASEH2A and reduced expression of RNASEH2B and C in RNASEH2A-KO clones (KO1, KO2), compared to control clones (C1-4). Tubulin was used as a loading control.* ***B)*** *RNase H2 activity against single embedded ribonucleotides in KO clones is absent in comparison to the observed activity in control cells. Data shown correspond to mean ± SD for n= 3 independent experiments.* ***C)*** *As observed with HeLa RNaseH2 KO clonal lines (Figure 14B), high*

levels of genome embedded ribonucleotides were also observed in HCT116 p53-/- RNASEH2-KO clones. Genomic DNAs isolated from parental cells, KO clones and control clones were treated with recombinant RNase H2 and resolved by alkaline gel electrophoresis. Smaller fragments indicate a higher number of embedded ribonucleotides. ***D)*** *Representative retrotransposition and toxicity assays conducted in HCT116 p53-/- control clones (C1-4), and in two RNASEH2A-KO clones (KO1 and KO2). Cells were transfected with active human LINE-1 (WT-hL1, element L1.3), RT-mutant LINE-1 (RTm-hL1, D702A), or a toxicity control vector (CTRL,* pcDNA*6.1).* ***E)*** *Schematic of plasmid JJ101/L1.3 and quantification of L1-WT retrotransposition. Average retrotransposition in control cells has been set to 100% in order to compare with other cell lines. Results show a reduced retrotransposition activity in RNase H2 null cells. Dots represent the mean of 3 technical replicates for individual clones. Lines indicate the mean of n=4 biological replicates for controls (C1-4) and n=2 for KO clones (KO1, 2) ± SEM (representative of 5 independent experiments). t-test, **, p<0.01. Adapted from Benitez-Guijarro et al, (2018).*

Remarkably, as previously observed in HeLa cells, we observed significantly reduced L1 retrotransposition in HCT116 p53-/- RNASEH2A-KO clones using the JJ101/L1.3-based assay (n = 2, 29 ± 8.5%) when compared with control cells (n = 4, 100 ± 14%, P = 0.0028; **Figure 16D and E**). While these data further suggest that RNaseH2 is in fact involved in L1 retrotransposition in a cell line independent manner, we noticed very variable Transfection Efficiency values for HCT116 p53-/- RNASEH2A-KO and control lines, which introduced a bias in our retrotransposition frequency calculations. Thus, we next generated a panel of U2OS RNASEH2A-KO clones and control cell lines (**Figure 17A-C**), using the same CRISPR/Cas9 system, and we then tested L1 retrotransposition (using parental U2OS cells as an additional control). Consistent with the data previously obtained using HeLa and HCT116 p53-/- RNASEH2A KO cell lines, we observed a substantial reduction in LINE-1 retrotransposition (n = 2, 25 ± 9.5%) in U2OS RNASEH2A-KO lines when compared with parental cells (set at 100%) or with a wild-type control clone (n = 2, 94 ± 8.2%; P = 0.016; **Figure 17D**). Notably, toxicity/clonability experiments revealed that all clones were able to produce similar numbers of blasticidin-resistant colonies after transfection with the control vector pcDNA6.1. As expected, no retrotransposition was detected in any U2OS-derived cell line, parental, control or KO, upon transfecting cells with the L1 RT-mutant allelic vector (L1.3-D702A) (**Figure 17D**).

In sum, we concluded that retrotransposition of human L1s is severely compromised in cells lacking RNase H2 activity, and in a cell line independent

manner. Thus, we conclude that RNase H2 activity is naturally involved in human L1 retrotransposition.

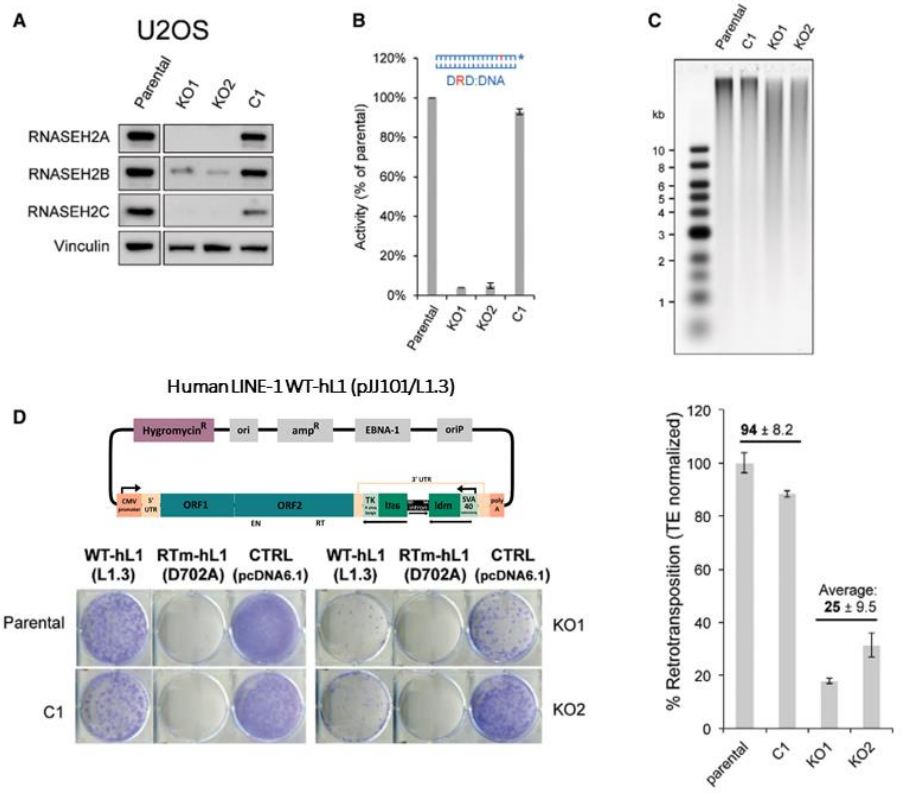


Figure 17. Reduced LINE-1 retrotransposition in RNase H2 null U2OS cells. A*) Western blot analyses show absence of RNASEH2A and reduced RNASEH2B and C in RNASEH2A-KO clones (KO1, KO2), compared to parental cells or a control clone (C1). Vinculin was used as a loading control.* ***B)*** *RNase H2 activity against single embedded ribonucleotides in KO clones is absent in comparison to the observed activity in control cells. Note that activity in parental U2OS cells was set at 100% and shown are mean ± SD for two independent experiments.* ***C)*** *As previously observed in HeLa and HCT116 p53-/- KO cells, high levels of genome embedded ribonucleotides were observed in U2OS RNASEH2-KO clones. Genomic DNAs isolated from parental cells, KO clones and control clones were treated with recombinant RNase H2 and resolved by alkaline gel electrophoresis. Smaller fragments indicate a higher number of embedded ribonucleotides* ***D)*** *Schematic of plasmid JJ101/L1.3 and representative retrotransposition and toxicity assays conducted in parental U2OS cells, a control clone (C1), and two RNASEH2A- KO clones (KO1 and KO2). Cells were transfected with vectors containing an active human LINE-1 (WT-hL1, element L1.3), an RT-mutant (RTm-hL1, L1.3 D702A), or a toxicity control vector (CTRL,* pcDNA*6.1). The level of L1-WT retrotransposition in parental cells was set to 100% in order to compare with KO clones (and controls). Results are shown in the graph, where we plotted mean ± SD for three technical replicates. Numbers indicate the average ± SD of n = 2 controls (parental, C1) and n = 2 (KO1, KO2) (representative of three independent experiments). Adapted from Benitez-Guijarro et al, (2018).*

1.2. RNase H2 facilitates the mobilization of non-LTR retroelements but is not required for LTR retrotransposons and DNA transposons.

During LINE-1 retrotransposition by TPRT, following the endonucleolytic cleavage of genomic DNA and the synthesis of the first cDNA strand by L1-ORF2p, a L1-mRNA:cDNA hybrid attached to genomic DNA is generated (i.e., a Y-branch intermediate). However, the RNA of the hybrid must be removed (or replaced) during subsequent steps, in order to complete second-strand cDNA synthesis of a novel L1 insertion. Since human L1-ORF2p lacks RNase H activity (Mathias *et al*, 1991; Malik *et al*, 1999; Cost *et al*, 2002; Piskareva *et al*, 2003; Piskareva & Schmatchenko, 2006), we reasoned that cellular RNase H2 might be responsible for degrading the L1-mRNA in the L1-RNA:cDNA hybrid during retrotransposition.

To test the above hypothesis, that is whether RNase H2 is required for the mobilization of LINE elements lacking an RNase H domain, we initially investigated the impact of RNase H2 deficiency on the retrotransposition of a zebrafish LINE-2 element (ZfL2-2). Briefly, and as human LINE-1s, zebrafish ZfL2-2 LINE-2 elements lack an RNase H domain but can retrotranspose in human cells (Sugano *et al, 2006;* Garcia-Perez *et al, 2010)*. Thus, we tagged an active Zfl2-2 element with the *mneoI* retrotransposition indicator cassette, which confers resistance to neomycin/G418 upon retrotransposition (Freeman *et al, 1994;* Moran *et al, 1996*; **Figure 18A**), and we measured its activity in parental HeLa, RNASEH2A-KO lines, as well as in control clonal lines. Additionally, and as a control, a human RC-L1 (L1.3) was tagged with the same *mneoI* retrotransposition indicator cassette (JM101/L1.3 vector, **Figure 18A**). Notably, we confirmed reduced retrotransposition of human RC-L1s tagged with *mneoI* **(Figure 18C-D),** similar to when using *mblastI*-based assays in both HeLa and U2OS RNASEH2A-KO clones (**Figure 14 and 17**). Consistent with our hypothesis, ZfL2-2-*mneoI* retrotransposition was also significantly reduced in KO clones (n = 5, 1.6 ± 0.91%) compared with parental cells (set at 100%) and control clones (n = 5, 85 ± 9.5 %; p = 0.0079; **Figure 18A - C**). Remarkably, when the same assay was performed in U2OS RNASEH2A-KO and control clones, we observed a virtually identical result for Zfl2-2 (**Figure 18D**; n = 2 controls, 105 ± 3.5% vs. n = 2 KO, 14 ± 2.4%; P = 0.0012).

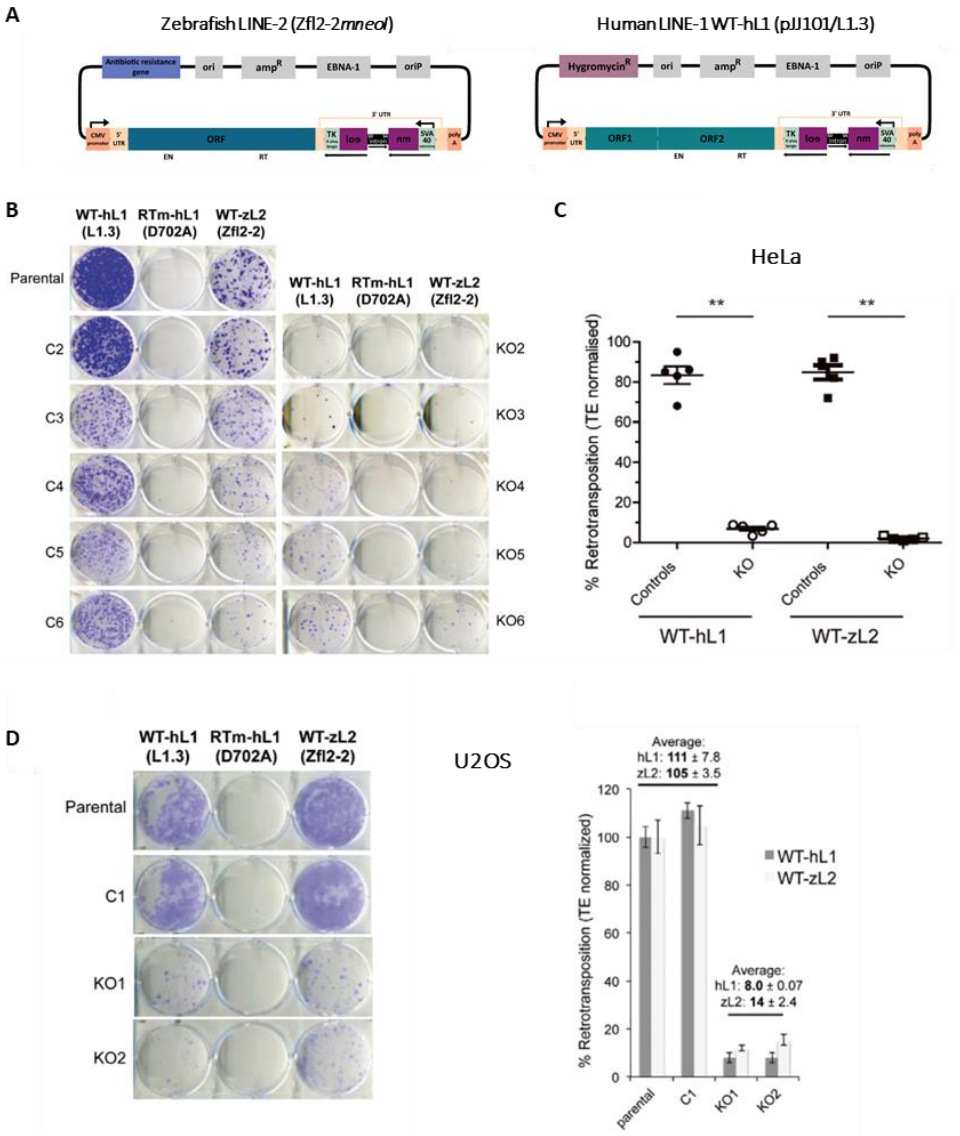


Figure 18. RNase H2 activity is required for LINE activity. A) *Schematic retrotransposition of vectors Zfl2-2mneoI and JM101/L1.3. The purple box with a backward NEO label depicts the* mneoI *retrotransposition indicator cassette.* ***B)*** *Representative retrotransposition assays in parental HeLa cells, control (C) and RNASEH2A-KO (KO) clones. Cells were transfected with vectors containing an active human LINE-1 (WT-hL1, element L1.3), an RT-mutant human LINE-1 (RTm-hL1, L1.3 D702A), or with an active zebrafish LINE-2 (WT-zL2, element Zfl2-2).* ***C)*** *Quantification of WT-hL1 (circles) and WT-zL2 (squares) retrotransposition in HeLa cells, normalized to the retrotransposition level in parental cells (set at 100%). Data points represent the mean of three technical replicates for individual clones. Lines indicate the mean of five biological replicates (C2-6 and KO2-6) ± SEM (representative of three independent experiments). For WT-hL1, in control lines*

*(n = 5) retrotransposition levels averaged 83 ± 2.5%; in null lines (n = 5) retrotransposition levels averaged 7 ± 2.3%. Mann–Whitney test; **p < 0.001.* ***D)*** *Representative retrotransposition assays conducted in parental U2OS cells, a control clone (C1) and two RNASEH2A-KO clones (KO1 and KO2). Cells were transfected with an active human LINE-1 (WT-hL1, element L1.3), an RT-mutant human LINE-1 (RTm-hL1, D702A) or with an active zebrafish LINE-2 (WT-zL2, element Zfl2-2) vector. Right, quantification of WT-hL1 (dark grey bars) and WT-zL2 (light grey bars) retrotransposition in U2OS cells, with retrotransposition in parental cells for both elements set at 100% for comparison. Mean ± SD for n = 3 technical replicates (representative of three independent experiments). Adapted from Benitez-Guijarro et al, (2018).*

In sum, all the data obtained strongly suggest that LINE elements lacking a functional RNase H domain depend on cellular RNase H2 activity for their efficient retrotransposition. Thus, and according to this model, RNase H2 could be consider as an integral part of the LINE retrotransposition machinery. Additionally, we further reasoned that retrotransposons containing a functional RNase H domain would not depend on cellular RNase H2 activity during their retrotransposition. Because LINE-1s containing functional RNase H domains are only found in lower eukaryotes and some plants, we next thought to explore whether mammalian ERVs could be used to solidify our model/hypothesis. Indeed, several active LTR retrotransposons from the mouse genome, such as MusD, are known to contain functional RNase H domains within their *pol* genes (Doolittle *et al, 1989)*; these mobile elements also generate an RNA:cDNA intermediate during their retrotransposition cycle, although in the cytosol, and as LINE-1s, RNA degradation/removal from the intermediate hybrid is necessary to complete LTR retrotransposition. Therefore, we next exploited active mouse MusD LTR retrotransposons, previously identified and cloned by the lab of Thierry Heidmann in France (Ribet *et al, 2004)*, to test their retrotransposition efficiency in RNase H2 KO models and controls. Specifically, we used an active MusD copy tagged with a NEO-based retrotransposition indicator cassette, termed *neoTNF* (construct pCMVMusD-6 *neoTNF*; Fig 19A; Ribet *et al, 2004)*, whose rationale is the same used in the L1 assay. Using this engineered MusD construct, retrotransposition in cultured cells can be quantified using G418 selection (see rationale in **Figure 19**). Consistent with our model/hypothesis, the level of MusD retrotransposition was similar between control and RNASEH2A-KO clones, either using HeLa (**Figure 19A and C**) or U2OS cells (**Figure 19B**). These data suggest that retrotransposons containing a functional RNase H domain do not depend on cellular RNase H2 activity for their efficient mobilization.

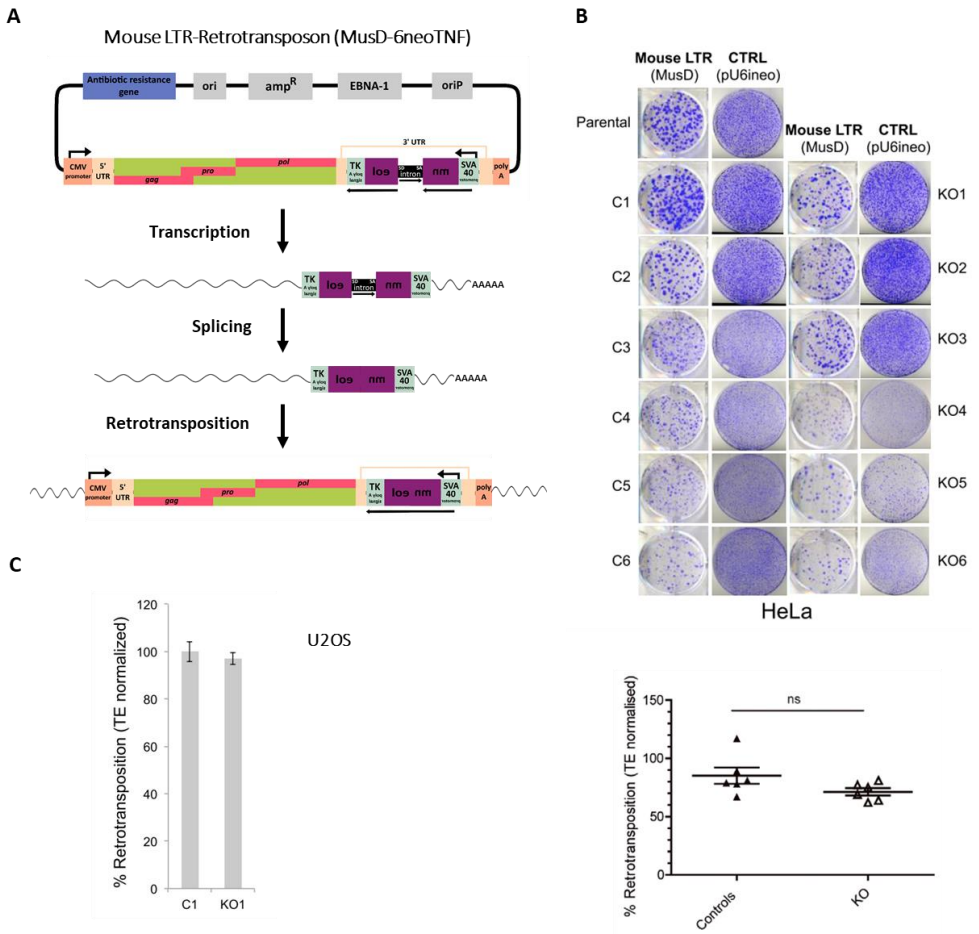


Figure 19. RNase H2 activity is dispensable for LTR-retroelements. A) *Schematic of a MusD LTR-retrotransposon tagged with the neoTNF cassette. The relative position of the gag, pro and pol genes is indicated. The purple box with a backward NEO label depicts the retrotransposition indicator cassette neoTNF. Upon transcription, the intron of the neoTNF cassette is removed by canonical cis and after translation and LTR retrotransposition, expression of NEO is activated from the de novo insertion, conferring neomycin (G418) resistance to cells.* ***B)*** *Above, representative results of LTR-retrotransposition assays in HeLa control (C1-6) and RNASEH2A KO (KO1-6) clones. Cells were transfected with a tagged active mouse MusD element (MusD) or with the toxicity control plasmid (pU6ineo). Below, quantification of MusD retrotransposition, normalized to the level in parental cells (set at 100%). Data points represent the mean of three technical replicates for individual clones. Lines indicate the mean of six biological replicates (C1-6 and KO1-6) ± SEM (representative of three independent experiments). t-test; ns, P > 0.05.* ***C)*** *Quantification of LTR-retrotransposition assays in U2OS control (C1) and RNASEH2A-KO1 clones. Mean ± SD for n=3 technical replicates (representative of 3 independent experiments). Adapted from Benitez-Guijarro* et al, *(2018).*

To further test our model/hypothesis, we next tested whether mobilization of DNA transposons, which move by a cut-and-paste mechanism that does not involve reverse transcription, requires the activity of cellular RNase H2. We employed an active Tc1-like resurrected DNA transposon termed Sleeping Beauty (SB; Ivics *et al*, 1997) that transposes very efficiently in human cells. Using a SB transposition assay based on G418 selection (see rationale in **Figure 20**), no differences in transposition rates were observed when comparing RNASEH2A-KO and control clones, both in HeLa (**Figure 20A and C**) and U2OS cells (**Figure 20A and B**). Therefore, we conclude that neither LTR retrotransposons encoding RNase H activity, nor DNA transposons, depend on cellular RNase H2 activity for their efficient mobilization, and in a cell line independent manner.

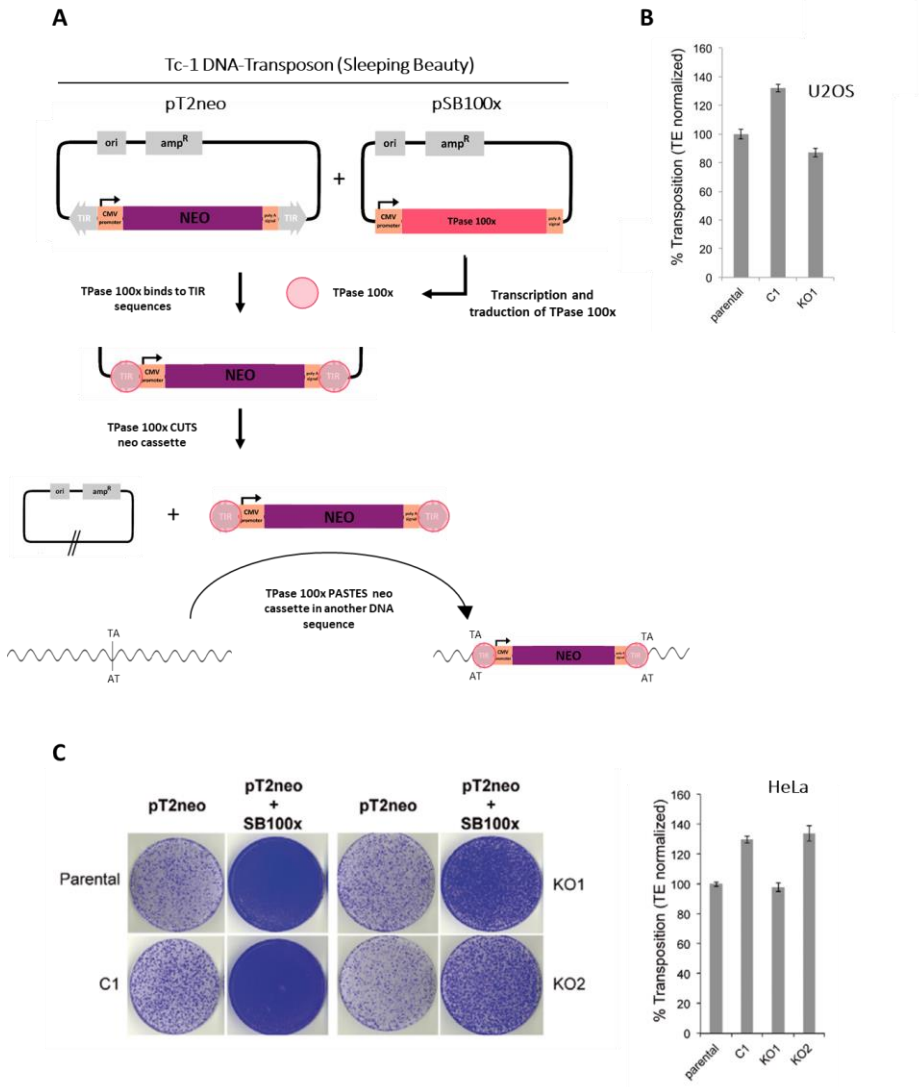


Figure 20. RNase H2 activity is dispensable for DNA-Transposon activity. A) *Schematic*

of plasmids pTneo and pSB100x, used in Sleeping Beauty transposition assay. The purple box with a NEO label depicts the neo expression cassette, flanked by Terminal Inverted Repeats (TIRs). pSB100x encodes an optimized version of the Transposase (TPase 100x) which is able to mobilize the neo expression cassette to a new genomic location by a cut and paste mechanism. ***B)*** *Quantification of SB assays in U2OS cells. Mean ± SD for n=2 technical replicates (representative of 3 independent experiments).* ***C)*** *Left, representative results for DNA-transposition assays (pT2neo + SB100x) or controls (only pT2neo) are shown. Right, quantification of SB transposition assays (pT2neo + SB100x samples), with the level in parental cells set at 100% for comparison. Mean ± SD for n = 3 technical replicates (representative of three independent experiments). Adapted from Benitez-Guijarro et al, (2018).*

1.3. Overexpression of RNase H2 increase L1 retrotransposition.

According to our model/hypothesis, cellular RNase H2 is required for LINE-1 retrotransposition. Thus, we next tested the effects of overexpressing RNase H2 on L1 retrotransposition, as our model would predict that overexpression of RNase H2 might increase the efficiency of L1 retrotransposition. RNase H2 is a heterotrimeric enzyme, and overexpression of the catalytic subunit alone does not significantly increase cellular activity (KR Astell, MAM Reijns, and AP Jackson, unpublished data). Therefore, we co-transfected HeLa cells, or U2OS cells, with three plasmids each expressing one of the RNase H2 subunits (tagged with a V5 epitope tag), along with a human LINE-1 vector where L1 was tagged with the *mblastI* indicator cassette (JJ101/L1.3). As controls, we co-transfected cells with a β-arrestin expression vector [(-ve), which doesn't significantly affect L1 retrotransposition (Bogerd *et al*, 2006)], or with an APOBEC3A overexpression vector [(+ ve), that strongly inhibits LINE-1 retrotransposition (Bogerd, 2006; Richardson *et al*, 2014)]. Cells were also co-transfected in parallel with the toxicity control vector pcDNA6.1, and the resulting number of colonies was used for normalization in the overexpression assays (to control for possible toxic side effects of cDNA overexpression). Notably, upon co-transfecting HeLa or U2OS cells with equal amounts of each overexpression plasmid for RNase H2 subunits (1:1:1 ratio) and the JJ101/L1.3 vector, we detected a significant increase in L1 retrotransposition compared with the β-arrestin control (2.1-fold in HeLa and 1.7-fold in U2OS, **Figure 21**). As expected (Bogerd *et al*, 2006), APOBEC3A overexpression reduced LINE-1 retrotransposition to around 10% of the controls (**Figure 21**). Further consistent with previous experiments, RT-mutant LINE-1 allelic constructs were unable to retrotranspose under any of the conditions tested. Notably, the retrotransposition of EN-mutant L1s, was not affected by RNase H2 overexpression (**Figure 21**), further suggesting that cleavage by RNase

H2 of miss-incorporated ribonucleotides in genomic DNA does not provide an entry point for ENi-retrotransposition.

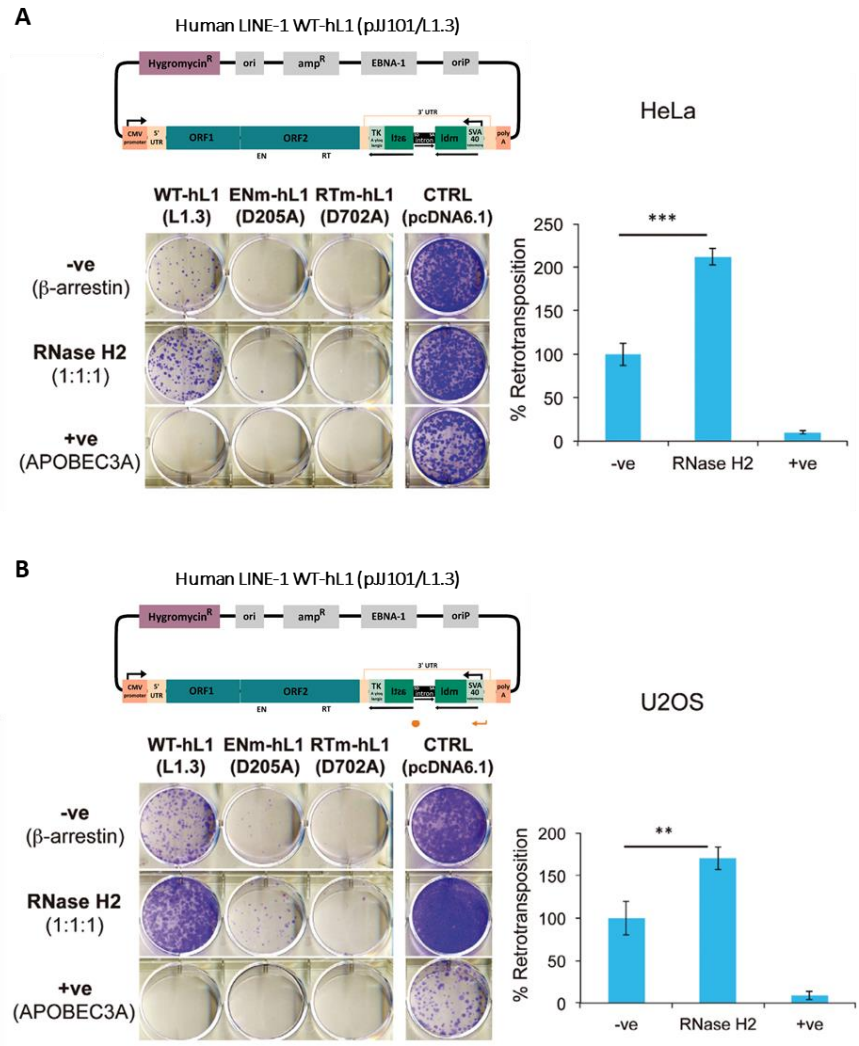


Figure 21. RNase H2 overexpression increases LINE-1 retrotransposition in HeLa and U2OS cells. A, B) *Panels (A) (HeLa cells) and (B) (U2OS cells) show representative results for retrotransposition and toxicity assays, underneath a schematic of retrotransposition vector JJ101/L1.3. Cells were co-transfected with JJ101/L1.3-based vectors [as indicated: WT-hL1 (L1.3), active human LINE-1; ENm-hL1 (D205A), EN-mutant; RTm-hL1 (D702A), RT-mutant] or with the toxicity control vector (CTRL,* pcDNA *6.1), alongside an expression vector for β-arrestin as a negative control (-ve), the three RNase H2 subunits (RNase H2 at a 1:1:1 ratio), or a plasmid expressing APOBEC3A as a positive control (+ve) known to restrict LINE-1 retrotransposition. Right panel, quantification of retrotransposition assays, with the level in cells co-transfected with β-arrestin set at 100% for comparison. Values were normalized for transfection efficiency and toxicity.*

*Mean ± SD for n = 3 technical replicates (representative of four independent experiments). Unpaired two-sided t-test; **p < 0.01; ***p < 0.001. Adapted from Benitez-Guijarro et al, (2018).*

While the above data is consistent with our proposed model/hypothesis, we noticed that the three RNase H2 subunits, which were tagged with a V5 epitope, were not expressed at the same level, as revealed by Western blot analyses (**Figure 22A**). Thus, we next modified the amounts of each plasmid co-transfected, in order to normalize expression of the three RNase H2 subunits. Indeed, we found that a transfection ratio of 14:7:1 (for RNASEH2A, B, and C, respectively) resulted in similar expression levels for each of the subunits (**Figure 22A**). Remarkably, by measuring retrotransposition at the 14:7:1 ratio, we confirmed that overexpression of RNase H2 increased the rate of LINE-1 retrotransposition by about 1.7-fold in HeLa and by ~1.4-fold in U2OS cells (**Figure 22B and C**).

Thus, we conclude that the reduction of LINE-1 retrotransposition in RNase H2 KO cells does not seem to be an indirect effect of its deficiency, and that overexpression data supports a direct role of cellular RNase H2 in facilitating LINE-1 retrotransposition.

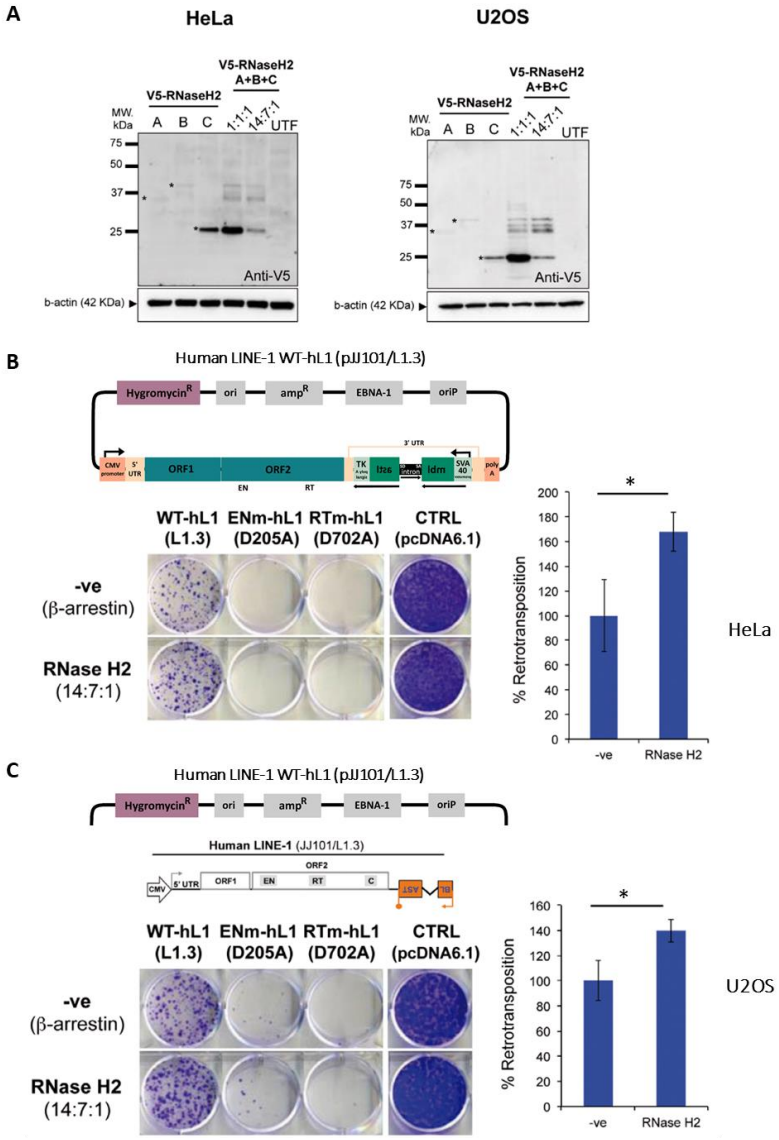


Figure 22. RNase H2 overexpression facilitates L1 retrotransposition. A) *Western blot analyses using an anti-V5 antibody in lysates of HeLa (left) or U2OS (right) cells transfected with the indicated plasmids. Cells were transfected with individual plasmids expressing single RNase H2 subunits or co-transfected with plasmids expressing all three subunits using two different ratios (1:1:1 or 14:7:1). β-Actin was used as loading control. Black asterisks mark the presence of each individual subunit. UTF, untransfected.* ***B, C)*** *Schematic of the retrotransposition vector JJ101/L1.3. Underneath, representative results from retrotransposition and toxicity assays conducted in HeLa (B) or U2OS (C) cells. Cells were co-transfected with vector JJ101/L1.3 or with the toxicity control vector (CTRL,* pcDNA6.1*), alongside an expression vector for β-arrestin (used as a negative control, -ve) or for each of the three RNase H2 subunits at the indicated ratio (RNase H214:7:1). Labels*

indicate if cells were transfected with an active human LINE-1 (WT-hL1, element L1.3), an EN-mutant LINE-1 (ENm-hL1, D205A), an RT-mutant LINE-1 (RTm-hL1, D702A), or with the toxicity control plasmid (CTRL, pcDNA*6.1). Right panels, L1-WT retrotransposition quantification, with the retrotransposition level in cells co-transfected with β-arrestin (-ve) set at 100% for comparison. Values were normalized for transfection efficiency and toxicity. Mean ± SD for n = 3 technical replicates (representative of four independent experiments). Unpaired two-sided t-test; *p < 0.05. Adapted from Benitez-Guijarro et al, (2018).*

1.4. Complementation of RNase H2 KO cell lines with the RNASEH2A subunit, but not with a "separation of function" mutant version, recovers L1 retrotransposition.

As discussed in the Introduction, RNase H2 exhibits two main enzymatic activities: 1) cleavage of ribonucleotides embedded in double-stranded genomic DNA; and 2) hydrolysis of the RNA strand of RNA:DNA heteroduplexes. Therefore, this raises the question of which of these two activities is involved on LINE-1 retrotransposition? Notably, previous research in yeast identified two amino acids (P40 and Y210) that when mutated severely abrogated RNase H2 activity against single ribonucleotides embedded in double-stranded DNAs, while retaining activity against RNA:DNA heteroduplexes (Chon *et al, 2013)*. Thus, we took advantage of this "separation-of-function" (SoF) mutant to identify which of the two enzymatic activities is required during L1 retrotransposition (**Figure 23A**).

First, we tested whether mutation of these two conserved aminoacids in human RNASEH2A would reproduce the phenotype observed in yeast (lack of activity in embedded ribonucleotides). Indeed, we observed that mutating these two conserved amino acids in human RNASEH2A (P40D/Y210A) resulted in yeast-like biochemical characteristics (**Figure 23C-F**, and Chon *et al, 2013)*, as revealed using recombinant purified human RNase H2 protein subunits. The double mutant P40D/Y210A-RNASEH2A subunit, termed human RNase H2 SoF, had virtually no activity against embedded ribonucleotides (**Figure 23C and D**) but retained activity against RNA:DNA heteroduplexes (**Figure 23E and F**). However, at difference with yeast, we noticed that the activity against the RNA of RNA:DNA heteroduplexes was reduced compared with wild-type (WT) RNase H2; unexpectedly, we also found that the human SoF mutant produced longer RNA products compared with wild-type RNase H2 (**Figure 23B**), suggesting that it might exhibit an altered scission pattern on the RNA strand of heteroduplexes. The difference in scission was observed using higher enzyme

concentrations or longer incubation times (**Figure 23B**), suggesting that the altered scission pattern is not due to lower intrinsic activity of the SoF mutant.

As an additional control, we also used a catalytic dead (CD) version of human RNASEH2A, after mutating two previously identified conserved residues (D34A/D169A); Reijns *et al, 2011)*; as previously described (Reijns *et al, 2011)*, we confirmed that recombinant CD RNase H2 lacked the two enzymatic activities of RNase H2 (activity on embedded ribonucleotides and on RNA:DNA heteroduplexes, see **Figure 23**).

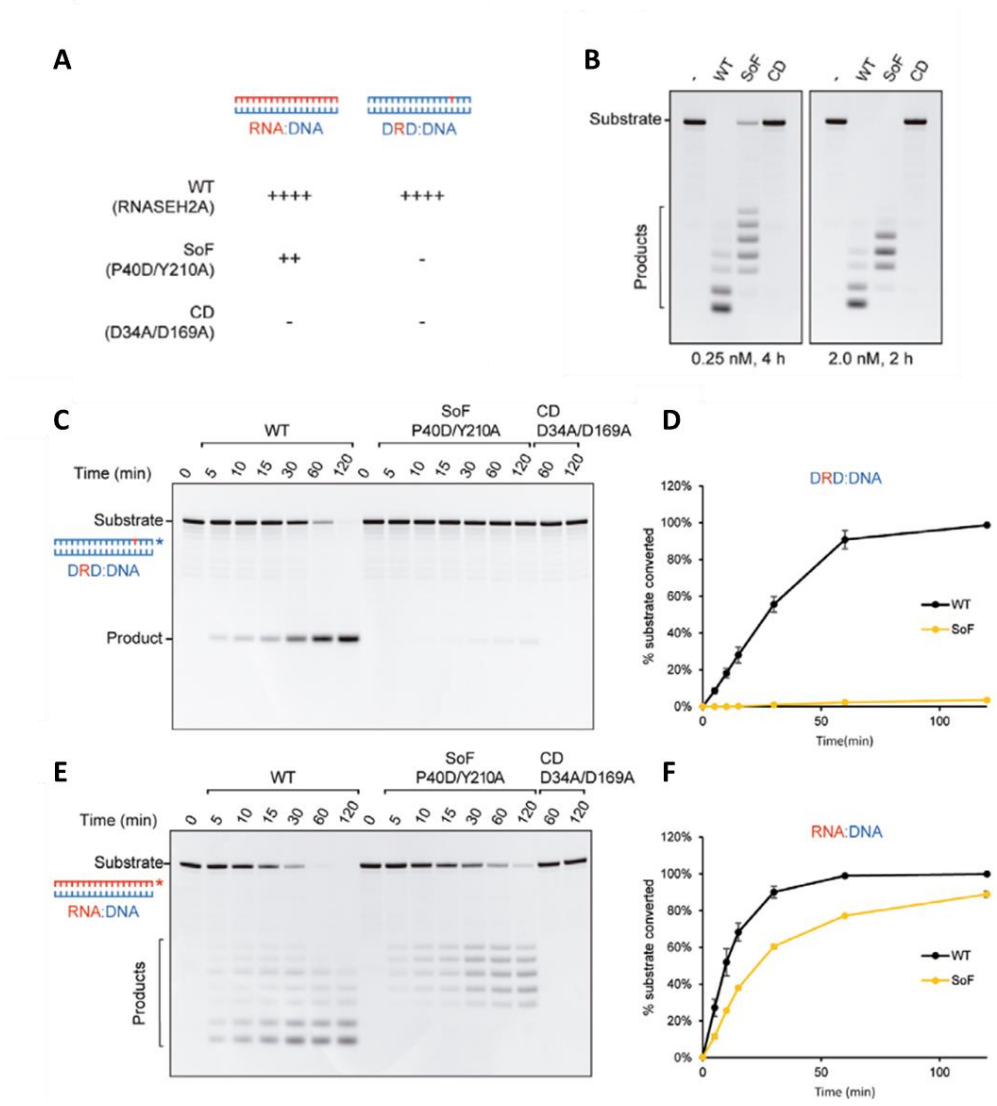


Figure 23. Characterization of a human SoF and CD mutant RNASEH2A. A) *Schematic of substrates used in RNase H activity assays. These assays either use an 18-bp RNA:DNA hybrid (left), or a short dsDNA containing a single-embedded ribonucleotide (DRD:DNA). RNASEH2A-WT can cleave both with high efficiency (++++), whereas RNASEH2A-CD (with D34A and D169A mutations) cannot cleave (-). The separation of function mutant (SoF, with P40D and Y210A mutations) retains some activity against RNA:DNA heteroduplexes (++), but has virtually no activity against single-embedded ribonucleotides (-).* ***B)*** *RNase H2 SoF has reduced activity against RNA:DNA heteroduplexes and does not fully process the hybrid, even at high concentration and/or long incubation times. RNase H activity was measured using the 18-bp RNA:DNA substrate, separating products by denaturing PAGE after cleavage with RNase H2. WT, SoF and CD RNase H2 were used at 0.25 nM for 4 h (left) or 2.0 nM for 2 h (right). Note the different pattern of products generated for SoF and WT.* ***C, D)*** *RNase H activity assays against single-embedded ribonucleotides using recombinant purified proteins (WT-RNase H2 and SoF-RNase H2). Note that only WT-RNase H2 shows activity in this assay. Plotted, mean ± SEM for three independent experiments.* ***E, F)*** *RNase H activity assays against RNA:DNA heteroduplexes using recombinant purified proteins (WT-RNase H2 and SoF-RNase H2). Note that the pattern of products generated by SoF-RNase H2 is different from the wild-type pattern. Plotted, mean ± SEM for three independent experiments. From Benitez-Guijarro et al, (2018).*

After characterizing human RNase H2 SoF and CD mutant proteins, we next exploited cDNA complementation of RNase H2 KO cells to determine which of the two activities of RNase H2 is involved on LINE-1 retrotransposition. To do that, we cloned a WT RNASEH2A cDNA, or any of the two cDNA mutants (SoF or CD), on retroviral vectors, and we next complemented one of the HeLa RNASEH2A KO clonal lines (KO1), generating complemented stable cell lines. As an additional control, we also generated a HeLa KO1 sub cell line expressing the empty vector (EV). Notably, western blot analyses confirmed expression of RNASEH2A and the consequent stabilization of RNASEH2B and C subunits in complemented cells, at levels indistinguishable from those observed in control cells or EV-complemented cells (**Figure 24A**).

We next tested complementation of RNase H2 activity. Assays using genomic DNAs from complemented cells revealed that complementation of clone KO1 with WT-RNASEH2A reduced the level of ribonucleotide incorporation into double-stranded DNA to a level similar to that detected in parental cells and controls (**Figure 24B**). In contrast, no reduction in ribonucleotide incorporation was observed in clonal lines complemented with SoF or CD RNASEH2A, or with EV (**Figure 24B**). Consistently, the same behavior was observed in assays using short synthetic substrates containing a single embedded ribonucleotide and cell lysates from complemented cells (**Figure 24C, E and F**).

We next tested complementation of RNase H2 activity towards RNA:DNA heteroduplexes. Notably, we detected efficient complementation in cell lines

supplemented with wild-type RNASEH2A (+ WT), reaching levels similar to those observed in parental and control cells (C1), whereas cells supplemented with SoF showed <50% activity (**Figure 24D, G and H**), consistent with the activity observed using recombinant RNase H2 SoF (**Figure 23**). As expected, no efficient complementation was observed for cells supplemented with the CD mutant or with EV (**Figure 24D, G and H**).

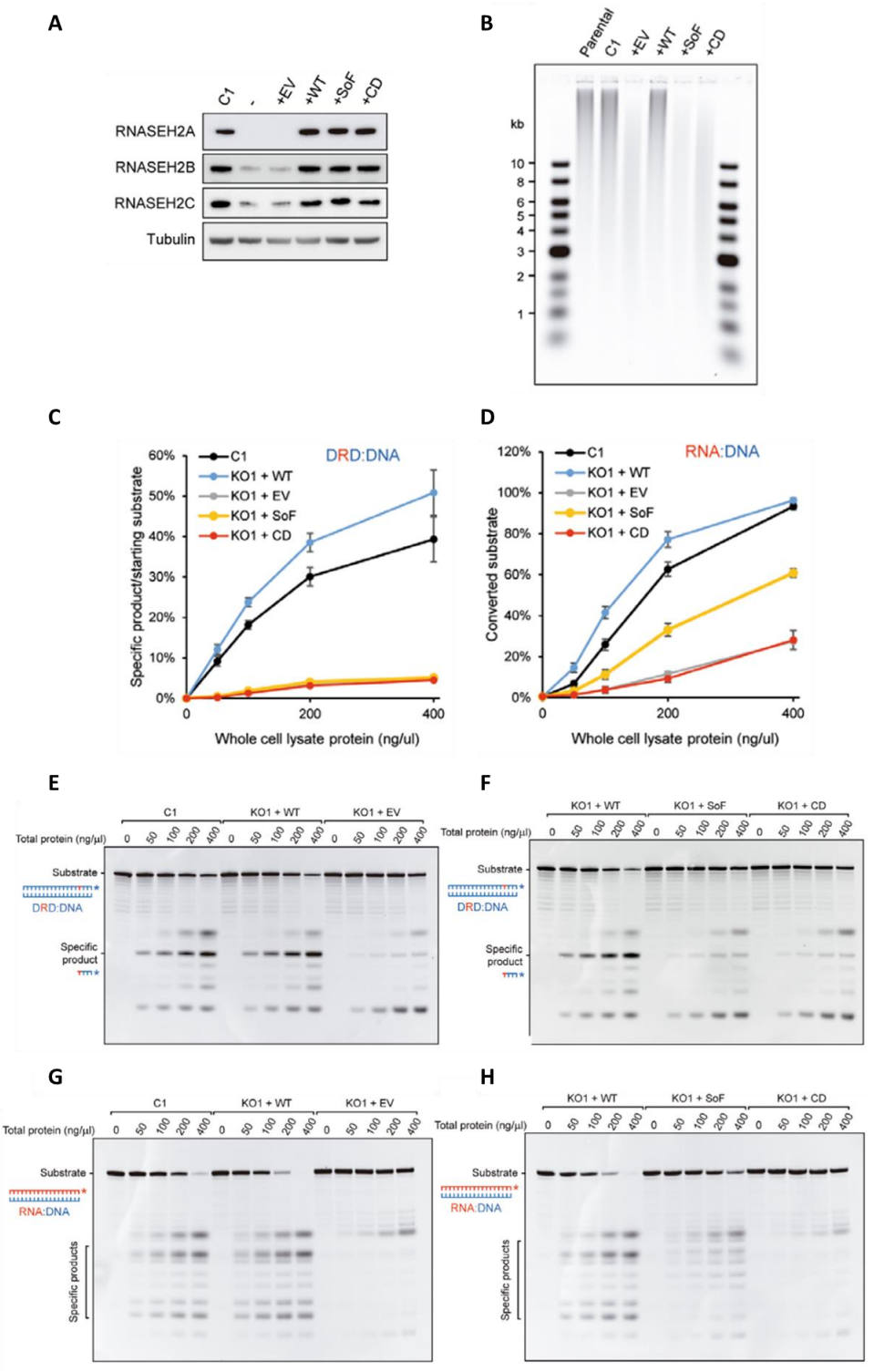

A
C1
+EV
+WT
+SoF
+CD
RNASEH2A
RNASEH2B
RNASEH2C
Tubulin
B
Parental
C1
+EV
+WT
+SoF
+CD
kb
10
8
6
5
4
3
2
1
C
DRD:DNA
C1
KO1 + WT
KO1 + EV
KO1 + SoF
KO1 + CD
Specific product/starting substrate
Whole cell lysate protein (ng/ul)
D
RNA:DNA
Converted substrate
E
F
Total protein (ng/µl)
Substrate
DRD:DNA
Specific product
G
H
RNA:DNA
Specific products

Figure 24. Complementation of RNASEH2A-KO cells with wild-type, but not separation of function RNASEH2A, rescues RNase H activity against RNA:DNA heteroduplexes. A) *Western blot analysis of RNase H2 expression in RNASEH2A-KO HeLa cells complemented with the indicated retroviral vector (EV, empty vector; WT, wild-type RNASEH2A; SoF, RNASEH2A-P40D/Y210A; CD, RNASEH2A-D34A/D169A, see main text for details). Tubulin was used as a loading control.* ***B)*** *Genomic DNAs were isolated from parental cells, a control clone (C1) and the four complemented cell lines (+EV, +WT, +SoF and +CD), RNase H2 treated, and resolved in alkaline gel electrophoresis. As expected, the high levels of genome-embedded ribonucleotides in RNASEH2A-KO cells are rescued only by complementation with wild-type RNASEH2A (+WT), not by SoF RNASEH2A (+SoF), CD RNASEH2A-A (+CD) or the empty vector (+EV). Smaller fragments indicate larger numbers of embedded ribonucleotides.* ***C)*** *DRD:DNA heteroduplex (18 bp; ribonucleotide-containing strand 3′-labelled) was incubated with increasing amounts of whole-cell lysates from the indicated cell line and separated by denaturing PAGE. RNase H activity against single-embedded ribonucleotides in RNASEH2A-KO cells is only rescued by wild-type RNASEH2A (KO1 + WT). The graph shows mean values ± SEM for three independent experiments.* ***D)*** *RNA:DNA heteroduplex (18 bp; ribonucleotide containing strand 3′-labelled) was incubated with increasing amounts of whole-cell lysates from the indicated cell line and separated by denaturing PAGE. RNase H activity against RNA:DNA heteroduplexes in RNASEH2A-KO cells is rescued by wild-type (KO1 + WT), not by CD RNASEH2A (KO1 + CD) or the empty vector (KO1 + EV). Note reduced activity and the difference in cleavage pattern produced by SoF RNASEH2A(KO1 + SoF). Plotted, mean ± SEM for three independent experiments.* ***E-H)*** *Representative gels (used for quantifications in Figure 24F and G) with results from RNase H activity against single-embedded ribonucleotides (D and E) and activity against RNA:DNA heteroduplexes (F and G) assays conducted with lysates from the indicated cell lines. Because RNase H1 is expressed in all cell lines, activity measured against RNA:DNA heteroduplex substrate in RNASEH2A-KO cell lysates is not completely absent. In addition, other nucleases present in the cell lysate act (non-specifically) on the substrate, causing further background activity on both substrates. G). From Benitez-Guijarro et al, (2018).*

To note, RNase H1, which only has activity against RNA:DNA heteroduplexes, is also expressed in human cells, which might explain some of the residual activity against the RNA of RNA:DNA heteroduplexes detected in RNASEH2A-KO cells complemented with empty vector and RNASEH2A-CD. Remarkably, as observed with recombinant purified proteins, the altered RNA scission pattern of RNA:DNA hybrids was also detected in cell lysates from RNASEH2A-SoF complemented cells (**Figure 24H**). These data suggest that, although P40D/Y210A amino acid changes in human RNase H2 act as cleavage-of-function mutations, SoF mutants exhibit RNase H activity against RNA from RNA:DNA heteroduplexes that is compromised both *in vitro* and *in vivo*.

Next, we explored whether RNASEH2A complementation would rescue LINE-1 retrotransposition. To do that, retrotransposition assays were performed

on the complemented cell lines using *mblastI* tagged L1s (i.e., JJ101/L1.3 vector series), and we observed that complementation with RNASEH2A WT efficiently rescued human L1 retrotransposition (**Figure 25A**). To further confirm this, we complemented a second HeLa RNASEH2A-KO clone (KO2) with RNASEH2A WT and observed the same trend (**Figure 25B**), confirming that the reduction of retrotransposition in RNASEH2A-KO cells is due to lack of RNase H2 activity and not to off-target collateral effects of CRISPR/Cas9 (off-targets). In contrast, complementation with RNASEH2A-SoF was unable to rescue L1 retrotransposition, yielding levels similar to those observed in cells complemented with EV or CD (**Figure 25A**). I speculate that the inability to rescue L1 retrotransposition in KO cells complemented with RNASEH2A SoF may be a consequence of the altered biochemical characteristics observed in recombinant RNASEH2A SoF, resulting in an inability to fully/efficiently degrade the RNA from LINE-1 RNA:DNA hybrids. Furthermore, I propose that the altered scission pattern of the SoF mutant may be due to its low affinity for the substrate, an effect that could be more pronounced by getting closer to the 3' end of the substrate.

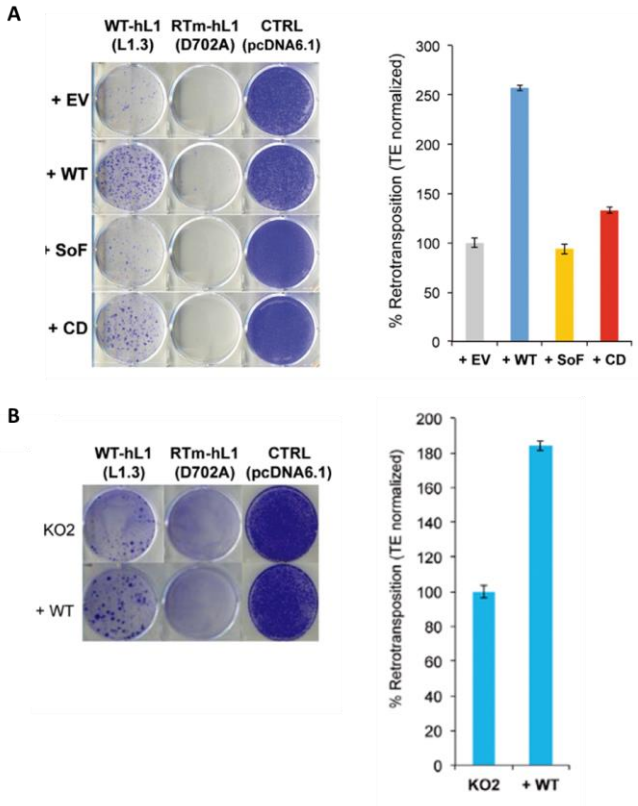


Figure 25. Complementation of RNASEH2A-KO cells with wild-type, but not

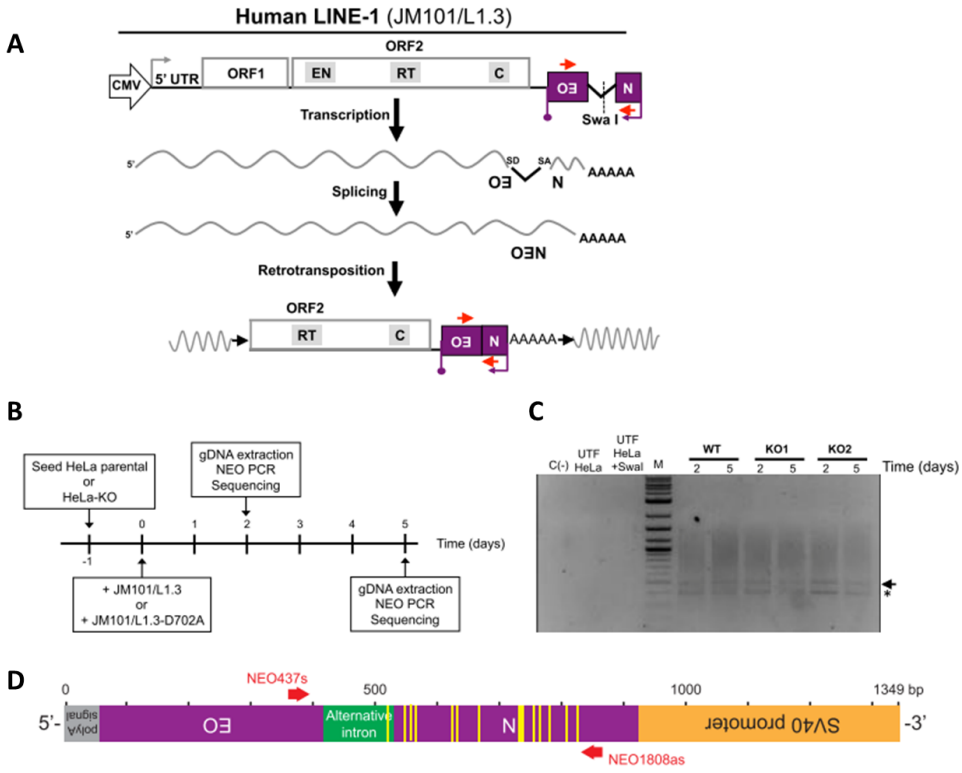
separation of function RNASEH2A, rescues LINE-1 retrotransposition. *A) Only wild-type RNASEH2A rescues the LINE-1 retrotransposition defect found in RNASEH2A-KO cells. Left, representative retrotransposition and toxicity assays conducted in the four complemented lines. Cells were transfected with vectors containing an active human LINE-1 (WT-hL1,L1.3), an RT-mutant LINE-1 (RTm-hL1,D702A) or a toxicity control plasmid (CTRL,* pcDNA *6.1). Right, quantification of L1-WT retrotransposition. For comparison, the retrotransposition level in KO1 cells complemented with the empty vector (EV) was set at 100%. Mean ± SD for n = 3 technical replicates (representative of six independent experiments). B) Wild-type RNASEH2A rescues the LINE-1 retrotransposition defect in RNASEH2A-KO2 cells. Representative retrotransposition and toxicity assays conducted in RNASEH2A-KO2 cells and RNASEH2A-KO2 complemented with wild-type RNASEH2A (+WT). Cells were transfected with an active human LINE-1 (WT-hL1, element L1.3), an RT-mutant LINE-1 (RTm-hL1,D702A), or a toxicity control plasmid (CTRL,* pcDNA*6.1). Right panel, quantification of L1-WT retrotransposition. For comparison, the retrotransposition level in KO2 cells was set at 100%. Mean ± SD for n = 3 technical replicates (representative of three independent experiments). From Benitez-Guijarro et al, (2018).*

1.5. RNase H2 KO cells do not exhibit an increased mutation rate on retrotranscribed L1 DNAs.

While several lines of evidence suggest that RNase H2 activity toward embedded ribonucleotides is not relevant for L1 retrotransposition, it remains a formal possibility. In fact, during human immunodeficiency virus type 1 (HIV-1) reverse transcription, the HIV-1 RT has been shown to miss-incorporate ribonucleotides on cDNAs with a high frequency, especially in macrophages (Kennedy et al, 2012). Previously, in a large-scale screening, it was shown that RNase H2 is important for HIV-1 infection (Genovesio *et al*, 2011), and might be involved in the removal of such incorrectly incorporated ribonucleotides. Although there are major differences between the mechanism of retroviral insertion and retrotransposition of LINE-1s, it is possible that ribonucleotides are incorrectly incorporated during L1 reverse transcription and/or second-strand synthesis, and that their removal by RNase H2-dependent RER might be important to achieve efficient retrotransposition. On the other hand, further research in yeast revealed that absence of RNase H2 lead to the introduction of incorrect ribonucleotides in genomic DNAs, resulting in high rates of 2-5 bp Top1-dependent deletions (McElhinny *et al*, 2010; Kim *et al*, 2011). These mutations were more likely to occur in tandem dinucleotide repeats, particularly of the CA: TG and GA: TC type (Kim *et al*, 2011; Potenski *et al*, 2014). To note, similar dinucleotide repeats are also found in the retrotransposition reporters used in this study, at rates similar to those observed in the reporters used in the yeast studies. Thus, these mutations could in principle inactivate the antibiotic resistance provided by

reporters inserted by retrotransposition, which could explain the reduced integration of L1 observed in RNase H2 KO cells using the engineered L1 assay. Therefore, we decided to analyze whether mutation of dinucleotide repeats could occur during L1 reverse transcription, and how frequently.

To explore editing of L1 cDNAs during retrotransposition, we took advantage of unique sequences within retrotransposition reporters, and using Sanger DNA sequencing, we compared the frequency of mutations in retrotranscribed reporters on parental, RNase H2 KO, and control HeLa clonal lines (see rationale in **Figure 26A and B**). To do that, we transfected cells with plasmid JM101/L1.3, containing a human RC-L1 tagged with the *mneoI* retrotransposition indicator cassette, and allowed cells to grow without G418 selection (**Figure 26A and B**). To note, lack of G418 selection would allow us to detect edited L1 cDNAs without biases. Two and five days after transfection, cells were collected, DNAs extracted, and these were analyzed by conventional PCR using primers flanking the intron of *mneoI*, allowing us to distinguish between products resulting from retrotranscription events (shorter amplification products) from products resulting from the amplification of the transfected vector (**Figure 26 A and C**). This was followed by cloning and sequencing of the amplification products, corresponding to reverse transcribed spliced *mneoI* reporters (i.e., *de novo* insertions of L1 or TPRT complexes). We compared two RNase H2 KO HeLa clones (KO1 and KO2) with parental cells, and the analyses of amplification products revealed a similar pattern of mutations (**Figure 26D and E**), and no increased mutation rate in RNASEH2A-KO cells compared with parental cells. Specifically, only nonsense mutations were identified; no deletions of 2-5 bp were detected in any of the clones analyzed. Therefore, we conclude that the reduction of LINE-1 retrotransposition in RNase H2 KO cells is not caused by hypermutation of L1 *de novo* cDNAs/insertions, which could occur due to the lack of RNase H2 activity capable of cleaving potentially miss-incorporated ribonucleotides during TPRT.



RNase H2	Total analyzed	Mutated	% Mutated
+	67	27	40
-	69	26	38

Figure 26. No increased mutation rate in de novo retrotranscribed LINE-1 sequences in RNase H2 null cells. A*) Schematic of retrotransposition assay using plasmid JM101/L1.3. Red arrows indicate primers used in the PCR assay, flanking the engineered intron present in mneoI. Note the SwaI site in the engineered mneoI intron (dashed vertical line), which, when cut, prevents amplification of intron-containing DNAs (that is, the transfected vector).* ***B)*** *Time line of the mutation detection assay.* ***C)*** *PCR products separated by agarose gel electrophoresis. Products amplified using DNA isolated from WT, KO1 and KO2 cell lines 2 or 5 days after transfection, as indicated; genomic DNAs were digested with SwaI prior to PCR (see A). C (-), negative control without template DNA. M, marker (1-Kb ladder, molecular weight standard). Arrow indicates the expected PCR product; * indicates product resulting from the use of cryptic splice sites in the Neo coding sequence (Gilbert et al, 2002).* ***D)*** *Schematic of the spliced mneoI cassette, where the presence of tandem dinucleotide repeats is indicated using yellow boxes. The green box (labelled Alternative Intron) indicates the relative position of the cryptic intron. Red arrows indicate the relative position of primers used in PCR.* ***E)*** *Mutation rate is not increased in RNase H2 deficient cells (-) when compared to RNase H2 proficient control cells (+). From Benitez-Guijarro et al, (2018).*

1.6. Overexpression of a separation-of-function (SoF) mutant RNASEH2A rescues L1 retrotransposition.

Previously, using retroviral complementation, we tested the effect of overexpressing the SoF RNASEH2A mutant on L1 retrotransposition, and we detected very low complementation (**Figure 25**). However, we also noticed reduced activity against RNA:DNA hybrids on RNASEH2A-SoF complemented cells, and an altered pattern of cleaved RNA products (**Figure 24**). Thus, we explored whether transient complementation with RNASEH2A-SoF, together with RNASEH2B and RNASEH2C, could result in better complementation levels, as we reasoned that transient co-transfection of the three subunits could alleviate the reduced activity of RNASEH2A-SoF. Remarkably, the co-transfection of L1 vectors with RNASEH2A-SoF (P40D/Y210A), RNASEH2B and RNASEH2C, increased L1 retrotransposition significantly when compared to β-arrestin co-transfected controls (**Figure 27A**, P = 0.019). Indeed, under these experimental conditions, complementation of WT-RNASEH2A was very similar to that of SoF-RNASEH2A (**Figure 27A**, right panel).

While the above data might look paradoxical, when we compared the enzymatic kinetics of WT-RNASEH2A and SoF-RNASEH2A against an RNA:DNA substrate (**Figure 27B**), we noticed that the SoF mutant exhibited a much-reduced substrate affinity when compared to WT (K^{SoF}_{M} ~ 16 x K^{WT}_{M}), whereas its maximum substrate conversion rate (catalytic constant) is similar to that of WT RNase H2 (K^{SoF}_{cat} ~ 0.83 x K^{WT}_{cat}). Thus, the low substrate affinity of RNase H2 SoF provides an explanation for our observations. In fact, these results further support our proposed model/hypothesis suggesting that the function of RNase H2 against RNA:DNA hybrids is necessary for efficient LINE-1 retrotransposition.

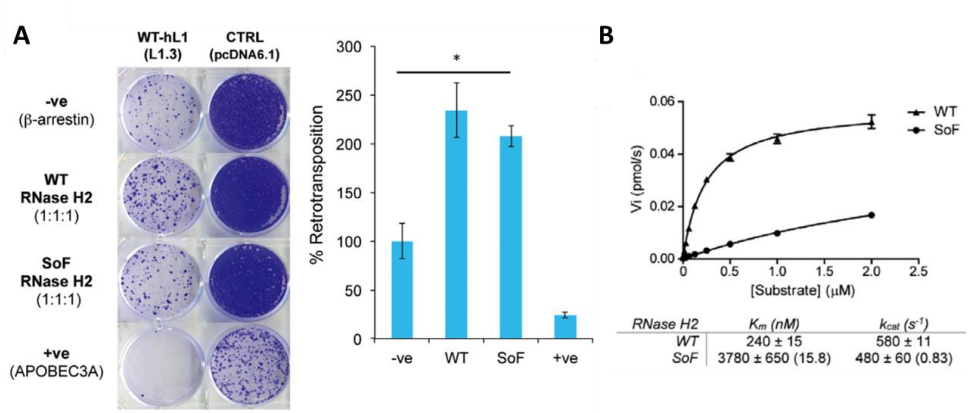


Figure 27. Cellular RNase H activity against RNA:DNA hybrids is important for L1 retrotransposition. A) *Increased L1 retrotransposition upon RNase H2 SoF overexpression. Left, representative retrotransposition and toxicity assays. HeLa cells were co-transfected with vectors containing an active LINE-1 (WT-hL1, L1.3), alongside an expression vector for β-arrestin as a negative control (-ve), the three RNase H2 subunits (RNase H2, at a 1:1:1 ratio; with RNASEH2A-WT or SoF) or a plasmid expressing APOBEC3A, known to restrict LINE-1 retrotransposition, as a positive control (+ve). Right panel, quantification of this retrotransposition assay, with the level in cells co-transfected with β-arrestin set at 100% for comparison. Values were normalized for transfection efficiency and toxicity. Mean ± SD for n = 2 technical replicates (representative of three independent experiments). Unpaired two-sided t-test; *p < 0.05.* ***B)*** *RNase H2 SoF has reduced RNA:DNA heteroduplex substrate affinity. Initial substrate conversion rates (Vi) by 0.1 nM recombinant RNase H2 were measured at different 18-mer RNA:DNA substrate concentrations. Mean ± SEM for n = 3 independent experiments. Km and kcat ± SEM were calculated in GraphPad Prism 5.04, using non-linear regression. Change for SoF compared to WT indicated between brackets. From Benitez-Guijarro et al, (2018).*

1.7. Overexpression of RNase H1 in RNase H2 KO cells partially rescues L1 retrotransposition.

According to our model/hypothesis, we assume that RNase H1 is expressed at normal levels in RNASEH2A-KO cells, similar to those of parental cells; because of the marked reduction on LINE-1 retrotransposition detected in RNASEH2A-KO cells, we further propose that it is unlikely that RNase H1 might play a major role on L1 retrotransposition. However, we noticed variability on L1 retrotransposition among the different RNASEH2A-KO lines tested, and retrotransposition was reduced between 4 to 15-fold depending on the KO line tested when compared to parental cells. Thus, it is formally possible that differences in RNase H1 expression/activity could contribute to the variable level of L1 retrotransposition

detected in RNase H2 KO cells, which also suggest that RNase H1 could also influence L1 retrotransposition.

To test this, we first tried complementing RNASEH2A-KO HeLa cells with RNase H1 cDNAs. Two independent HeLa RNASEH2A-KO clones (KO1 and KO2) were transduced with a retroviral vector expressing the nuclear isoform of human RNase H1. To confirm overexpression of RNase H1, we measured activity against RNA:DNA heteroduplex in lysates from these cells, showing a small but significant increase in enzymatic activity (**Figure 28A**).

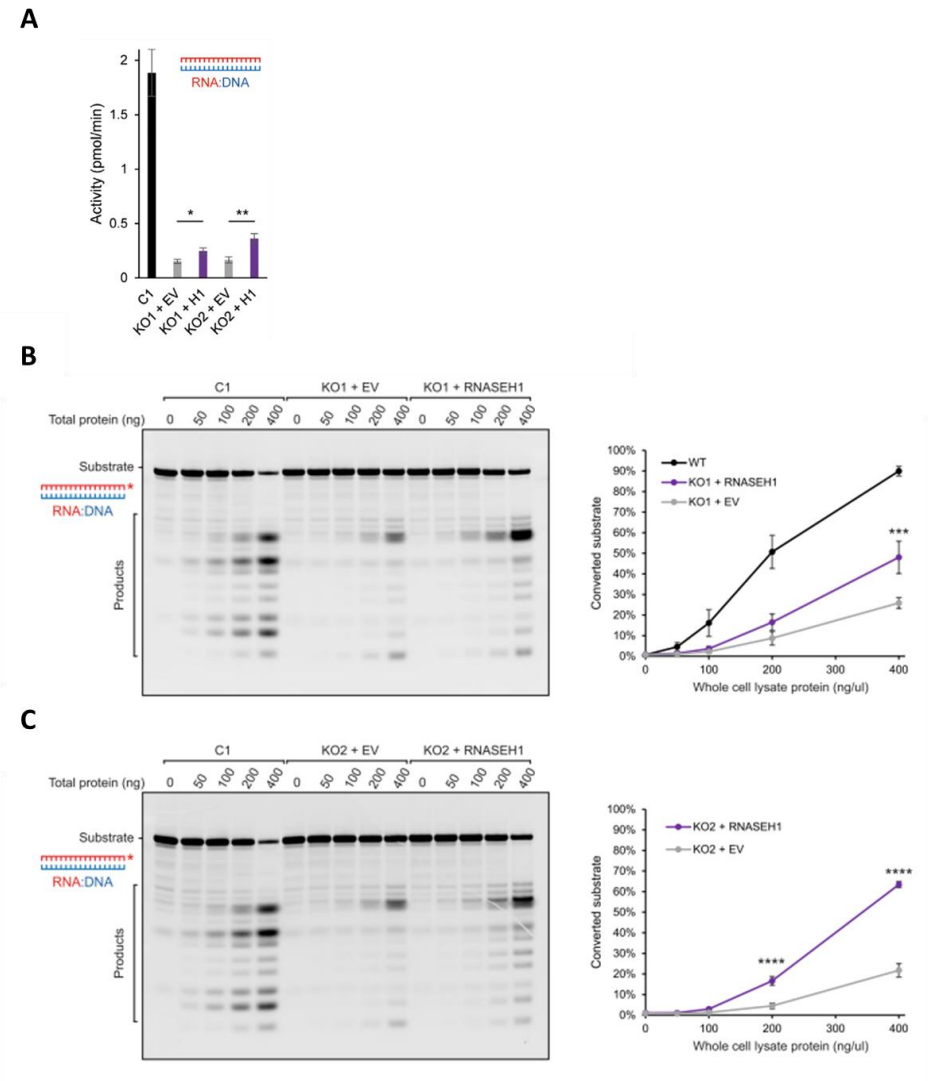


Figure 28. Increased RNase H activity against RNA:DNA heteroduplexes in RNASEH2A-KO HeLa cells complemented with human RNase H1. A) *RNase H activity against RNA:DNA heteroduplexes in RNASEH2A-KO cells (clones KO1 and KO2) is partially rescued by overexpression of human nuclear RNase H1 (KO1 +H1 and KO2 +H1 vs. KO1 + EV and KO2 + EV). An RNA:DNA heteroduplex (18 bp; ribonucleotide containing strand 3′-labelled) was incubated with whole-cell lysates and the speed of cleavage determined using a FRET-based assay. Mean values ± SEM for n = 6 independent experiments.* ***B, C)*** *RNase H activity assays conducted on RNASEH2A-KO cells (A, KO1; B, KO2) complemented with human nuclear RNASEH1 (+RNASEH1) or with empty vector (+EV). RNase H activity was measured using the 18-bp RNA:DNA substrate, separating products by denaturing PAGE after incubation with lysates from the indicated cell lines using increasing amounts of total protein. Left panels, representative gels. Right panels, quantifications showing mean ± SEM of n=4 (KO1) or n=3 (KO2) independent*

*experiments. Mean for n=3 independent experiments. Two-way ANOVA with post-hoc Bonferroni multiple comparison test shows significant increase in activity against RNA:DNA heteroduplexes in KO+RNASEH1 compared to KO+EV cells. ***, p<0.001; ****, p<0.0001. From Benitez-Guijarro et al, (2018).*

Notably, RNASEH2A-KO1 and RNASEH2A-KO2 cells complemented with wild-type RNase H1 showed a 1.6- and 2.2-fold increase in RNA excision activity of RNA:DNA heteroduplexes compared with each corresponding EV (empty vector) control, respectively (**Figure 28A**). Further controls confirmed that the pattern of RNA degradation detected in RNase H1 complemented cells, using either KO1 or KO2, is similar to that detected in parental and control HeLa clonal lines (**Figure 28B and C**).

Once complemented with the nuclear isoform cDNA of human RNaseH1, we next tested the efficiency of human RC-L1 retrotransposition, using *mblastI* tagged L1s (i.e., JJ101/L1.3 vector series), and using WT-RNASEH2A complemented cells as positive controls. Notably, we found that complementation with RNase H1 partially alleviated the L1 retrotransposition defect detected in RNASEH2A-KO clones, but not to the same extent as complementation with WT-RNASEH2A (**Figure 29**). Using KO1, the increase in L1 retrotransposition was 3.4-fold for WT-RNASEH2A complemented cells, versus 2.2-fold for RNase H1 complemented cells; similarly, for KO2, the increases on L1 retrotransposition were of 2.5- and 1.8-fold, for cells complemented with WT-RNASEH2A and RNase H1, respectively (**Figure 29B and C**). Controls revealed that similar numbers of foci were detected upon transfecting the clonability/toxicity pcDNA 6.1 plasmid (**Figure 29B**). Thus, we propose that failure to fully rescue retrotransposition upon RNase H1 complementation could be explained by the relatively low increase in cellular RNase H activity against RNA:DNA heteroduplexes detected in RNase H1 complemented KO-RNaseH2 cells. However, the ability of RNase H1 to partially rescue L1 retrotransposition is consistent with the possibility that both cellular RNase H enzymes might facilitate L1 retrotransposition, in a model in which both nucleases might act by degrading the RNA from L1-RNA:cDNA heteroduplexes formed during TPRT. In sum, we propose that RNase H2 facilitates LINE-1 retrotransposition by removing the L1 RNA from RNA:cDNA hybrids after reverse transcribing the first L1 cDNA, allowing completion of the second L1 cDNA strand, and thus of a *de novo* insertion (**Figure 29D**), with a minor contribution of nuclear RNase H1.

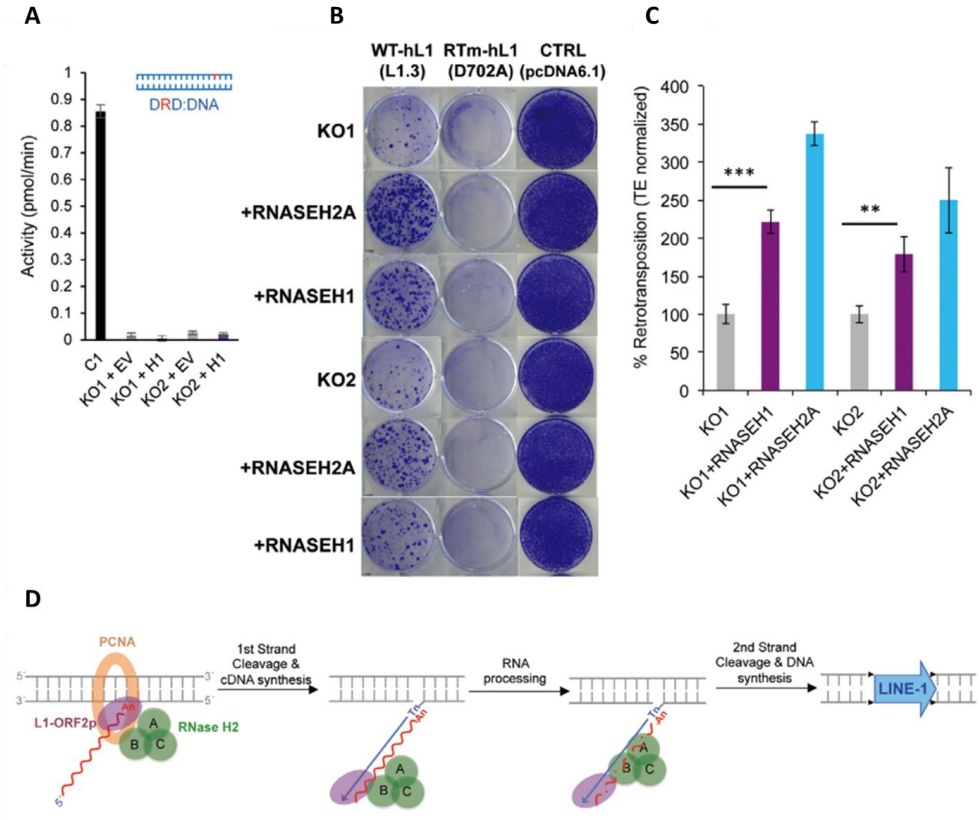


Figure 29. Cellular RNase H activity against RNA:DNA hybrids is important for L1 retrotransposition. A) *RNase H activity against single-embedded ribonucleotides in RNASEH2A-KO cells (clones KO1 and KO2) is not rescued by human RNase H1 (KO1 +H1 and KO2 +H1 vs. KO1 + EV and KO2 + EV). The efficiency of cleaving an 18-bp substrate was determined using a FRET-based assay. Mean values ± SEM for n = 3 independent experiments.* ***B, C)*** *Representative retrotransposition and toxicity assays (B) conducted in two RNASEH2-KO clones (KO1 and KO2) complemented with WT-RNASEH2A (+RNASEH2A) or with human nuclear RNase H1 (+RNASEH1). Cells were transfected with vectors containing an active LINE-1 (WT-hL1, L1.3), an RT-mutant LINE-1 (RTm-hL1, D702A), or a toxicity control plasmid (CTRL,* pcDNA*6.1). Quantification of L1-WT retrotransposition (C). For comparison, the retrotransposition level in KO1 or KO2 cells was set at 100%. Mean ± SD for n = 3 technical replicates (representative of three independent experiments). Unpaired two-sided t-test; **P < 0.01; ***p < 0.001* ***D)*** *Proposed model for the role of RNase H2 during TPRT. Degradation of LINE-1 RNAs in RNA:cDNA hybrids by cellular RNase H2 allow completion of* de novo *LINE-1 insertions. From Benitez-Guijarro et al, (2018).*

1.8. RNase H2 may interact with LINE-1 through PCNA.

According to our proposed model (**Figure 29D**), RNase H2 facilitates LINE-1 retrotransposition by removing the L1-RNA from L1-RNA:cDNA heteroduplexes. However, overexpression of RNase H1 in RNase H2 KO cells failed to fully rescue L1 retrotransposition. Since both RNase H2 and RNase H1 are active against RNA:DNA heteroduplexes, in the previous section we expected to achieve a complete rescue of L1 retrotransposition by overexpression of RNase H1. Such overexpression should have been able to complement the lack of RNase H2 if retrotransposition depends on RNase H activity directed against RNA:DNA hybrids; however, retrotransposition was only partially recovered.

Notably, a functional PCNA-interacting protein (PIP) motif has been recently identified in L1-ORF2p (Taylor *et al*, 2013), between the EN and RT domains. Interestingly, the RNASEH2B subunit also contains a functional PIP motif (Chon *et al*, 2009; Bubeck *et al*, 2011). In this context we wondered whether RNase H2 might be interacting with L1 through PCNA, thereby enabling the PIP interaction motif present in the RNASEH2B subunit to drive RNase H2 to L1 TPRT complexes.

To test this hypothesis, we analyzed whether a L1-PIP mutant (JJ101/L1.3-PIP6, containing two miss-sense mutations in L1-ORF2p (YY414/5AA) that abolish interactions with PCNA, data not shown) could retrotranspose in RNase H2 KO cells (and controls) and at what rate. We reasoned that the L1-PIP mutant would retrotranspose at similar levels in parental and RNase H2 KO HeLa cells, as the interaction with RNase H2 through PCNA is impaired. In other words, because the L1-PIP mutant can´t interact with PCNA, and thus with RNase H2, the status of RNase H2 would not influence its ability to retrotranspose. In these experiments, as an additional internal control of mutant L1 retrotransposition, we tested the frequency of retrotransposition of an EN-mutated L1, using the allelic L1 plasmid JJ101/L1.3-D205A. Consistent with previous reports (Taylor *et al, 2013)*, we found that the PIP-mutant L1 retrotransposed at a very low level in parental HeLa cells (<10% compared with the WT-L1 used (JJ1010/L1.3), **Figure 30**). Controls confirmed reduced retrotransposition of the EN-mutant L1 in parental HeLa cells (<5% compared with the WT-L1 used (JJ1010/L1.3), **Figure 30**). Further controls using overexpression of FLAG-tagged L1-ORF2p in HEK293T cells confirmed interaction of WT-L1 with PCNA, and lack of interaction when the PIP motif of L1 was mutated (i.e., L1-PIP6 mutant, data not shown).

Next, using HeLa RNase H2 KO cells, we confirmed reduced retrotransposition of WT-L1 (<50% when compared with parental HeLa, **Figure 30**). Remarkably, we found that the PIP-mutant L1 retrotranspose at similar levels in parental and RNase H2 KO cells (~4.5% vs ~5% in parental and KO1, respectively; **Figure 30**). Further controls revealed that a similar number of blast-resistant foci were

generated when parental and RNase H2 KO cells were transfected with the toxicity/control plasmid pcDNA6.1 (data not shown). While further research is needed, our data suggest that the PIP motif of L1-ORF2p is relevant to mediate a functional interaction with RNase H2, through PCNA. Indeed, this model (**Figure 29**) also explains why RNase H1 can only partially complement the retrotransposition deficit of RNase H2 KO cells, as the lack of a functional PIP motif in RNase H1 would imply that its interaction with L1-RNPs during TPRT would occur by diffusion.

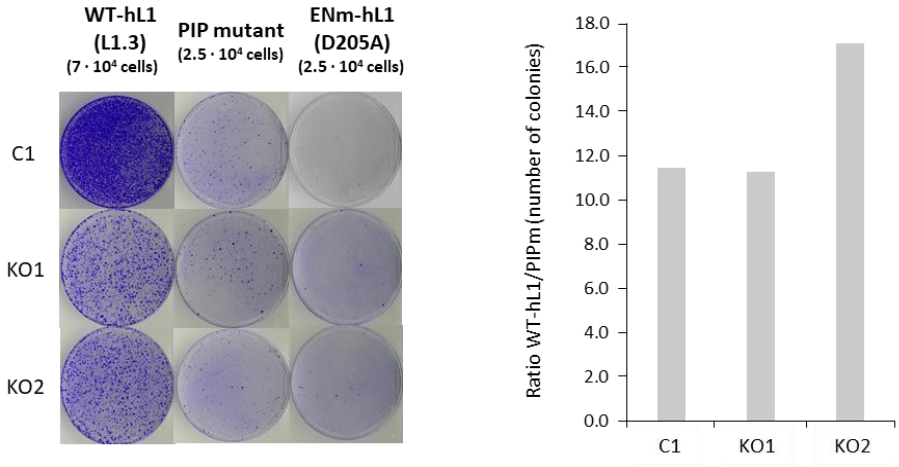


Figure 30. PIP-mutant (JJ101/L1.3-PIP6) retrotransposed at similar levels in parental and RNase H2 KO cells. *Representative retrotransposition results using a WT-L1 (JJ101/L1.3), an EN-mutant L1 (JJ101/L1.3-D205A), or a PIP-mutant L1 (JJ101/L1.3-YY414/5AA) on parental and RNase H2 KO cells. Note that because mobilization of mutant L1s is lower, we plated different number of cells in each assay. Assays were performed in 100 mm plates, using 6 · 10^4 cells with JJ101/L1.3, 3 · 10^5 cells with JJ101/L1.3-PIP6 and 3 · 10^5 cells for JJ101/L1.3-D205A.*

1.9. RNASEH2A mutations characterized in AGS patients also reduce L1 retrotransposition.

Collectively, our data using several L1 retrotransposition assays and different RNASEH2A-KO cell lines strongly suggest that RNase H2 is required for efficient L1 retrotransposition. By extrapolation, we propose that L1 retrotransposition is likely reduced in cells from AGS patients containing mutations in RNase H2, as these mutations are known to result in severely reduced RNase H2 enzymatic activity levels. Two RNASEH2A mutations, G37S and E225G, are frequently found in AGS patients. RNASEH2A-G37S, which causes perinatal lethality (Pokatayev et al, 2016*)*, and RNASEH2A-E225G, are the only miss-sense mutations known to cause changes in the catalytic subunit of RNase H2 when found in homozygosis (Rice *et al*, 2013*)*. On the other hand, common mutations in AGS patients have also been characterized in the RNASE2B (A177T) and RNASEH2C (R69W) subunits.

In order to explore how RNase H2 mutations found in AGS patients would impact L1 retrotransposition, we initially generated recombinant RNase H2 proteins containing mutations characterized in AGS patients. We generated recombinant RNase H2 heterotrimeric complexes containing the following mutations: G37S and E225G in RNASEH2A, A177T in RNASEH2B, and R69W in RNASEH2C (**Figure 31A**). Next, we tested their activity *in vitro*, and consistent with patient characterization, we found that RNASEH2A-G37S and RNASEH2A-E225G lack activity on short RNA:DNA heteroduplexes (**Figure 31A**). In contrast, and as reported (Reijns *et al*, 2011*)*, no major defects on the hydrolytic activity of RNA:DNA heteroduplexes was found for RNASEH2B-A177T, while the defect of RNASEH2C-R69W is small but significant (**Figure 31A**).

To further characterize mutant RNASEH2A proteins, we also analyzed the pattern of RNA cleavage using RNA:DNA heteroduplexes; at difference with SoF-RNASEH2A, we found that the pattern of RNA cleaving for both RNASEH2A mutant proteins, G37S or E225G, is very similar to that of WT-RNASEH2A (**Figure 31A**). Indeed, we confirmed that the two mutant proteins characterized in AGS patients had significantly lower activity on RNA:DNA heteroduplexes, as revealed by *in vitro* assays using higher concentration of recombinant proteins (**Figure 31A**). Thus, according to the proposed model/hypothesis, I speculate that cells containing any of these two AGS mutant RNase H2 proteins would be impaired for L1 retrotransposition. To test this, we next complemented a HeLa RNASEH2A-KO clonal line (KO2) with RNASEH2A cDNAs containing either the G37S or the E225G mutation, or with EV as a control. Cells complemented with WT-RNASEH2A were used as an internal negative control. Western blot analyses revealed restoration of expression of each of the three RNASEH2 subunits when cells were complemented with either AGS-mutant RNASEH2A isoform (G37S or

E225G), at similar levels as those detected with WT-RNASEH2A complemented cells, and in contrast with EV-complemented cells (**Figure 31C**). Consistent with their characterized RNase H2 intrinsic activity, a small but significant increase in cellular RNase H2 activity was detected in cell lysates from cells complemented with AGS-mutant RNASEH2A isoforms (G37S or E225G) when using substrates containing embedded ribonucleotides, but not on RNA:DNA heteroduplexes (**Figure 31D**, compare right and left panel). As expected, no restoration of RNase H2 activity in either assay was found for EV-complemented cells, and cells complemented with WT-RNASEH2A showed fully restored RNase H2 activity levels in both assays, at levels indistinguishable from those detected in parental or control cells (**Figure 31D** and data not shown).

Once characterized, we analyzed changes on L1 retrotransposition using *mblastI* tagged L1s (i.e., JJ101/L1.3 vector series) and cells complemented with AGS-mutant RNASEH2A isoforms (G37S or E225G) or controls (EV and WT-RNASEH2A). Controls revealed lack of L1 retrotransposition in cells transfected with RT-mutant L1s, and a similar number of resistant-foci in cells transfected with the clonability control pcDNA6.1 (**Figure 31E and F**). Interestingly, we observed partial rescue of LINE-1 retrotransposition in cells complemented with both AGS-mutant RNASEH2A isoforms (G37S or E225G) (**Figure 31E and F**), although both AGS-mutant alleles showed significantly reduced retrotransposition when compared with cells complemented with WT-RNASEH2A ($n= 3$, $p = 0.003$ for G37S and $p = 0.018$ for E225G). Therefore, consistent with our model/hypothesis, we propose that AGS patients with RNase H2 mutations would have reduced levels of LINE-1 retrotransposition in their genomes.

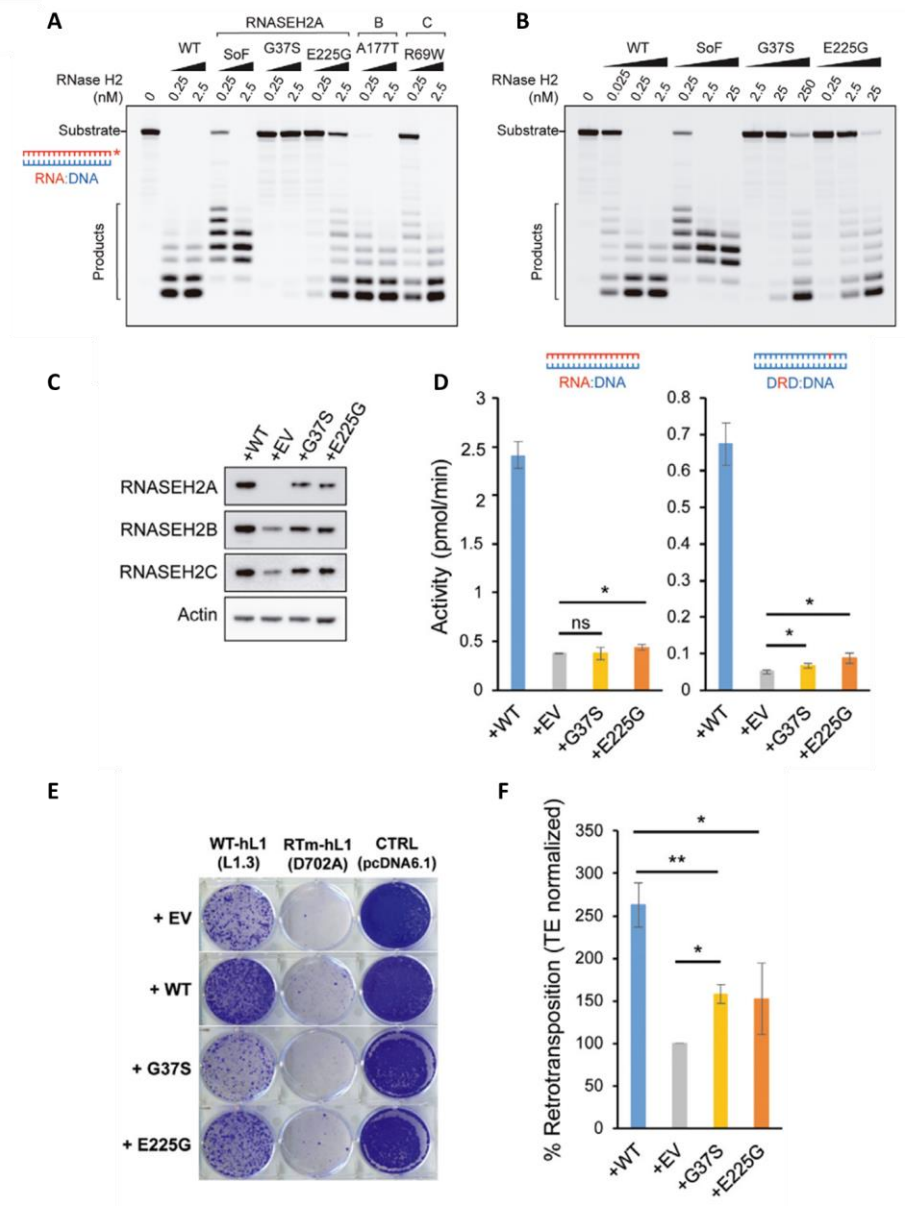


Figure 31. Reduced L1 retrotransposition due to RNase H2 AGS mutations. A, B) *Reduced enzyme activity for recombinant RNase H2 carrying AGS disease mutations. RNase H activity assays (against 18-mer RNA:DNA heteroduplex) using the indicated recombinant purified proteins. Wild type or mutant recombinant RNase H2 (RNASEH2A-P40D/Y210A, RNASEH2A-G37S, RNASEH2A-E225G, RNASEH2B-A177T and RNASEH2C-R69W) were tested using the indicated protein concentration. As previously shown (Reijns et al, 2011), the A177T mutation had limited impact on enzyme activity, whereas the G37S, E225G and R69W mutations all caused a*

substantial reduction in RNase H activity. Note that only RNase H2-SoF (RNASEH2A-P40D/Y210A) generates an altered cleavage pattern on RNA:DNA hybrids. Shown are representative results from three independent experiments. ***C)*** *Western blot analysis of RNase H2 expression in RNASEH2A-KO HeLa cells (KO2) complemented with the indicated retroviral vector (EV, empty vector; WT, wild- type RNASEH2A; RNASEH2A-G37S; RNASEH2A-E225G). Actin was used as a loading control.* ***D****) Complementation of RNASEH2A-KO cells with RNASEH2A with AGS mutations (G37S and E225G) leads to a small but significant increase in RNase H activity. Mean values ± SEM for n = 3 independent experiments. Unpaired two-sided t-test; *p < 0.05; ns, p > 0.05.* ***E, F)*** *Cells expressing AGS mutant RNase H2 fail to support efficient L1 retrotransposition. (E) Representative retrotransposition and toxicity assays conducted in the four complemented lines. Cells were transfected with vectors containing an active human LINE-1 (WT-hL1, L1.3), an RT-mutant LINE-1 (RTm-hL1, D702A) or a toxicity control plasmid (CTRL,* pcDNA *6.1). (F) Quantification of L1-WT retrotransposition in the complemented lines. For comparison, the retrotransposition level in KO cells complemented with empty vector (EV) was set at 100%. Mean ± SD for n = 3 independent experiments (each experiment performed in technical duplicates). Unpaired two-sided t-test; *p < 0.05; **p < 0.01. From Benitez-Guijarro et al, (2018).*

2. Study of the LINE-1 interactome in pluripotent (PCs) and differentiated (DCs) PA-1 cells.

The second part of my Thesis aims to study whether the L1-interactome, that is the collection of host factors interacting with L1-encoded proteins, changes during cellular differentiation, as we now know that some aspects of L1 regulation (i.e., L1-silencing, see Garcia-Perez *et al,* 2010*)* are cell line dependent and/or can change depending on the ontogenic status of cells. Indeed, to further understand how L1-silencing operates in pluripotent cells (PCs), and to further understand why silencing is attenuated in differentiated cells (DCs), we decided to compare the LINE-1 interactome of isogenic PCs and DCs. Our working hypothesis suggest that factors associated with L1-silencing are by definition PC-specific cellular factors, and that some could interact with L1-encoded proteins (i.e., are part of the PC interactome and are absent from the DC interactome). In these experiments, we collaborated with the lab of Dr John LaCava (Rockefeller University, US), as his lab implemented innovative and robust methods to identify specific cellular factors associated with L1-encoded proteins in the context of retrotransposition intermediates (i.e., L1-RNPs). In other words, in this section of my Thesis, I implemented approaches developed by LaCava's group to identify L1 interactors in isogenic human PCs and DCs (Taylor *et al,* 2013, 2018; *III. INTRODUCTION, 6.1. Proteomic study of the LINE-1 interactome*). Thus, the work described in this section of my Thesis was performed with the help and assistance of Dr. John LaCava's team, and up front I would like to indicate that: i) Hua Jiang and Mehrnoosh Oghbaie helped with the purification of L1-RNPs and data analysis, respectively; ii) Kelly Molloy, from Dr. Chait's laboratory (Rockefeller University, US), helped performing the mass spectrometry experiments; iii) John LaCava supervised and guided the proteomic experiments.

In the proteomic experiments, we used human embryonic carcinoma cells (hECs) as a model of human pluripotency (PA-1 cells). As previously described, hECs are derived from human teratomas, which are benign germ cell tumors (Rossant & Papaioannou, 1984; Zeuthen *et al*, 1980; Przyborski *et al*, 2004). Notably, hECs have a transcription profile very similar to hESCs (Sperger *et al,* 2003*)*, and as hESCs, some hEC lines retain their pluripotency but can differentiate into the three main germ layers (ecto, meso and endoderm) (Przyborski *et al,* 2004; Andrews *et al,* 2005*)*. Thus, hECs represent an amenable model to study early human embryogenesis processes, especially for studies requiring a large amount of biological material (LaCava *et al,* 2015*)*. As hESCs, pluripotent PA-1 cells naturally express L1 RNAs and L1-RNPs, consistent with the hypomethylation

status of their genomes. Previous research in our lab has fully characterized PA-1 cells as a pluripotency model, exhibiting the following advantageous features: (i) PA-1 cells are easy to culture using regular tissue culture media, even in large quantities, and can be easily manipulated genetically; (ii) PA-1 cells have a transcription profile very similar to hESCs; (iii) PA-1 cells have a stable diploid genome, with a single balanced translocation; (iv) PA-1 have a stable pluripotent cell phenotype in culture; (v) PA-1 cells have an hypomethylated genome; (vi) PA-1 cells overexpress endogenous L1-RNAs and L1-RNPs; (vii) PA-1 cells accumulate insertions of endogenous L1s; (viii) PA-1 cells epigenetically silence *de novo* L1 insertions (Garcia-Perez *et al*, 2010); and (ix) PA-1 cells can be easily differentiated into DCs *in vitro*, mostly to ectodermal cells (Garcia-Perez *et al*, 2010). In summary, PA-1 cells are an ideal model to study L1-silencing from a proteomic and mechanistic angle.

Previous studies using cultured human transformed cell lines have demonstrated that the affinity capture of L1-ORF1p and/or L1-ORF2p also co-purify L1-RNPs specific interactors (*III. INTRODUCTION, 6.1. Proteomic study of the LINE-1 interactome*). Thus, we used PA-1 cells (in PC and DC states) to purify L1-RNPs (i.e., retrotransposition intermediates) and their interactors. To purify L1-RNPs, we tried to affinity capture both L1-encoded proteins, L1-ORF2p and L1-ORF1p. To achieve this goal, we considered two different strategies: 1) capturing ectopic L1-RNPs from RC-L1s using expression vectors containing epitope tags in L1-ORF1p and L1-ORF2p (using T7 and FLAG, respectively); and 2) capturing naturally expressed endogenous L1-RNPs. Of course, each approach has its own advantages and limitations, and the majority of studies so far have mainly used ectopically expressed active L1s in proteomic studies, likely because these studies used transformed cell lines with relatively low and variable endogenous expression levels of L1-RNPs.

2.1. Capture of ectopic L1-RNPs in PCs and DCs.

Following the successful proteomic approach previously used in transformed cell lines, we initially focused in capturing ectopic L1-RNPs using PC and DC PA-1 cells. To do that, we transfected pluripotent PA-1 cells (PC) with a vector expressing a human RC-L1 that contains a triple-FLAG epitope in the C-terminus of L1-ORF2p and a T7-epitope in the C-terminus of L1-ORF1p (**Figure 32**); in these experiments, we used a well characterized human RC-L1, L1.3 (Sassaman *et al*, 1997; GeneBank L19088.1), which retrotranspose at high frequencies in cultured human cells (i.e., it is a "hot" RC-L1), creating vector pJM101/L1.3-ORF1-T7/ORF2-3xFLAG (**Figure 32**). As a negative control for the capturing of L1-ORF2p, we used

an allelic vector lacking the triple-FLAG epitope in L1-ORF2p, vector pDK101 (**Figure 32**). Epitope tagged RC-L1s were cloned in pCEP4 (Invitrogen), an episomal replicating plasmid that can be selected in human cells, as the backbone contains a selectable marker for hygromycin (**Figure 32**).

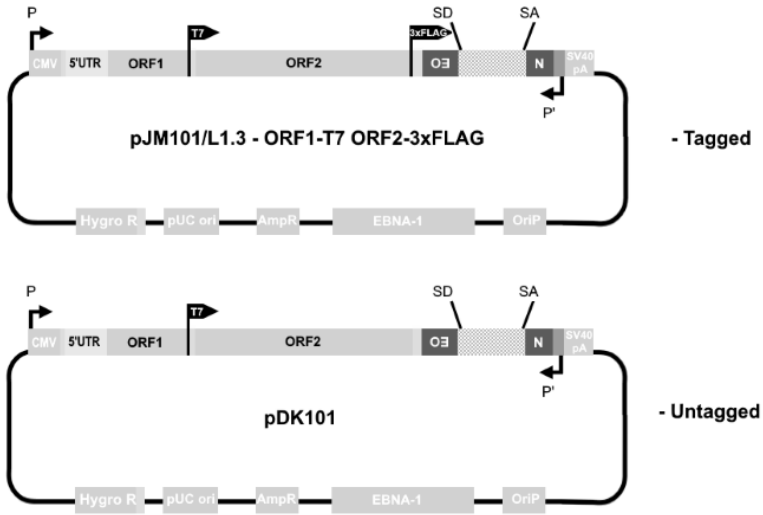


Figure 32. Schematic of vectors pJM101/L1.3 – ORF1-T7 / ORF2-3xFLAG and pDK101. *Both vectors contain a copy of the "hot" human RC-L1 L1.3, where the C-term of L1-ORF1 was tagged with a T7 epitope. pJM101/L1.3 – ORF1-T7 / ORF2-3xFLAG also contains a triple FLAG epitope in the C-term of L1-ORF2p, absent in pDK101. To note, a retrotransposition indicator cassette,* mneoI, *is cloned in the 3´UTR of L1.3. The vector backbone contains a hygromycin resistance cassette as well as OriP and EBNA-1, allowing its selection in cultured cells as a non-integrating episome.*

Using HeLa cells, we demonstrated that the inclusion of epitope tags doesn't interfere with L1 retrotransposition, and both constructs retrotransposed at similar levels as the parental L1 (vector JM101/L1.3) in cultured cells. PC PA-1 cells were transfected with both plasmids and selected with hygromycin during 14 days, in order to obtain a homogeneous population of transfected cells. In parallel, an aliquot of transfected PA-1 cells was cultured under differentiation conditions, to obtain isogenic DCs, using conditions previously known to induce fast and robust cellular differentiation (replace FBS with KnockOut Serum replacement and adding retinoic acid (all-trans) at 1 μM concentration; Garcia-Perez *et al,* 2010). Transfected PCs and DCs were cultured and expanded using 500 cm^2 tissue culture plates, in order to obtain sufficient cellular material to produce cell beads (in the order of several grams of cells, see *VI. MATERIALS AND METHODS*, *16. Cell collection and freezing to prepare cell beads*). After expansion and collection, pelleted

cells were directly frozen in LN_2, and employing cryomilling, BBs were grinded to form cell powder, which was subsequently aliquoted and stored at -80°C (see *VI. MATERIALS AND METHODS*, points *15-17* for a description of the followed process).

Next, we optimized the amount of cell powder/antibody (which was coupled with magnetic beads) to affinity capture L1-RNPs and their interactors. To note, as my main goal was using the affinity capture procedure and liquid chromatography coupled to mass spectrometry (LC-MS/MS) protocol previously optimized by LaCava's team using transformed cell lines (LaCava *et al*, 2015; Taylor *et al*, 2013), I first determined the amount of overexpressed L1-encoded proteins in PCs and DCs. In fact, a prerequisite to use the method established by LaCava's team on transformed cell lines was that a similar yield of L1-ORF1p and/or L1-ORF2p was achieved in PCs and DCs. Previously, LaCava's team used HEK293T$_{LD}$ cells, which incorporate the Tet-On inducible expression system, to stably express 3xFLAG tagged L1-ORF2p from a bicistronic L1 construct, and using either a codon optimized human "hot" RC-L1 (known as ORFeus-Hs (L1$_{RP}$)(An *et al*, 2011); plasmid pLD401) or the native human "hot" RC-L1 (L1$_{RP}$; plasmid pMT302; note that both systems were previously described in Dai *et al*, 2012 and Taylor *et al*, 2013). Thus, I used pLD401 and pMT302 on HEK293T$_{LD}$ cells as a reference sample to compare my PC and DC samples. On preliminary experiments, we affinity captured L1-ORF2p from transfected PCs and DCs, employing an anti-FLAG antibody coupled to magnetic dynabeads, to get a preliminary view of the yield from our transfected PC and DC samples in comparison with pLD401-HEK293T$_{LD}$ samples (**Figure 33A**). Unfortunately, we were unable to detect L1-ORF2p in the lysate, pellet or elution from the affinity capture, either using PC or DC samples. To verify that the loaded material for DC and PC was similar, and to further test the reliability of the immunocapture protocol, the membrane was stripped and probed with an anti-PCNA antibody, as PCNA is a well-known interactor of human L1-ORF2p (Taylor *et al*, 2013). While PCNA was detected in both the pellet and flow through samples using PC and DC, suggesting that we successfully captured 3xFLAG-L1-ORF2p expressed in PA-1 cells, no L1-ORF2p nor L1-ORF1p were detected. This experiment was performed twice, obtaining the same results, suggesting that the expression levels of ectopic L1s in transfected PCs and DCs were very low, clearly below their detection limit. To rule out that the extraction conditions used could affect our capability to capture L1-encoded proteins, as we could lose L1-ORF1p and/or L1-ORF2p during the extraction, we explored expression of L1 proteins in whole cell extracts from PCs and DCs using Western-blotting (**Figure 33B**). To note, we used anti-FLAG antibodies to explore L1-ORF2p expression and a commercial anti-L1-ORF1p antibody to explore L1-ORF1p expression. However, we couldn't detect endogenous/ectopic expressed L1-ORF1p, nor ectopically expressed 3x-FLAG-

tagged L1-ORF2p, either in PA-1 PC or DC samples (**Figure 33B**). Increasing the ratio of cell powder and magnetic dynabeads coupled to FLAG antibody did not dramatically change our capability to detect L1-ORF2p (not shown); however, we observed a faint band corresponding to L1-ORF1p (**Figure 33B**, a combination of ectopic and endogenous L1-ORF1p, as the antibody used could detect both L1-ORF1p types). In sum, and while our data suggest that we could successfully capture ectopically expressed L1-ORF2p from PCs and DCs, as well as some interactors (i.e., PCNA), our overall expression levels are significantly lower than those achieved when using HEK293T cells.

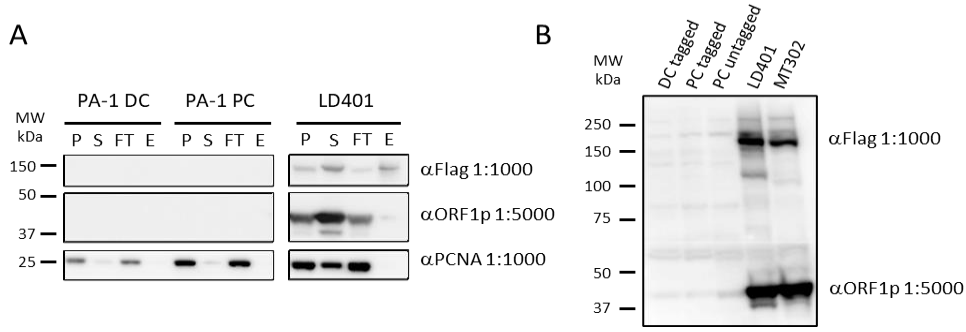


Figure 33. L1-ORF1 and L1-ORF2 detection from PC and DC transfected cells. *Proteins were affinity captured with anti-FLAG dynabeads from total protein extracts.* ***A)*** *Representative results from affinity capture experiments performed with anti-FLAG coupled dynabeads from PC and DC cells. The scale of the extractions was of 200 μl Extraction buffer per every 50 mg of cell powder (1:4 w/v), and using 5 μl of anti-FLAG coupled dynabeads. Western blot was developed with an exposure time of 140 seconds. Both the affinity capture and Western blot were performed as described in VI. MATERIALS AND METHODS. P, pellet; S, supernatant; FT, flow through; E, elution.* ***B)*** *Western blot using total protein extracts from PA-1 DC, PA-1 PC (tagged, pJM101/L1.3 -ORF1-T7/ORF2p-3xFLAG; untagged, pDK101), HEK293T$_{LD}$ LD401 and MT302. Total protein extracts were obtained by extracting 50 mg of cell powder in SDS buffer (2% SDS, 40 mM Tris pH 8) (1:4 w/v), standard sonication (5 times x 2 sec (2 A)) and clarification of the extract. BCA quantification was performed and 25 μg of protein were loaded. Western blot was developed with an exposure time of 30 seconds.*

On a complementary approach, we affinity captured L1-ORF1p instead of 3x-FLAG-L1-ORF2p, using anti-L1-ORF1p coupled dynabeads (mαORF1p, *mouse anti-human L1-ORF1p*); to do that, we used PC and DC PA-1-derived samples, either using naïve untransfected cells or the previously described transfected PA-1 samples (i.e., cells expressing ectopic 3xFLAG-L1-ORF2p and/or T7-L1-ORF1p);

as a positive control, we used MT302-HEK293T$_{LD}$ cells (**Figure 34**).

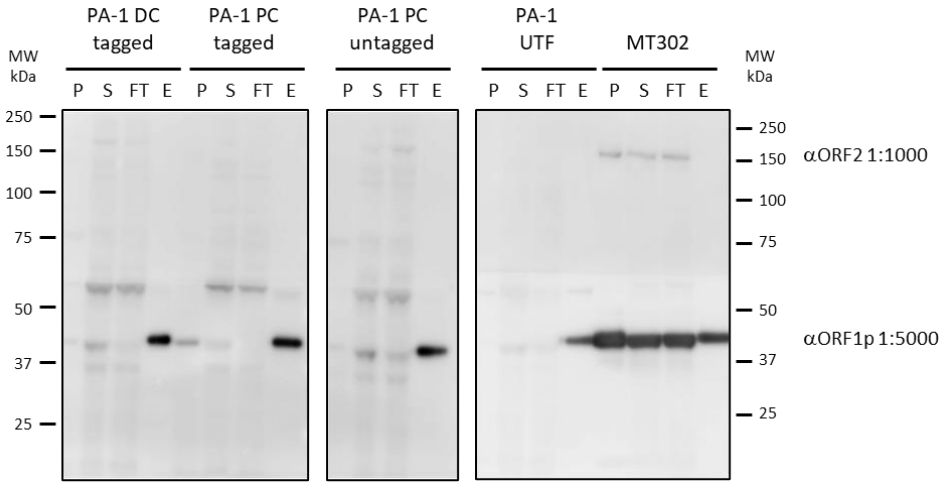


Figure 34. Affinity capture using anti-human_L1ORF1p magnetic dynabeads (mαORF1p). *The scale of the extractions was of 200 µl Extraction buffer per every 50 mg of cell powder (1:4 w/v), and 5 µl of anti-human_L1hORF1p coupled dynabeads. P, pellet; S, supernatant; FT, flow through; E, elution; UTF, untransfected; tagged, cells transfected with pJM101/L1.3-ORF1-T7/ORF2p-3xFLAG; untagged, cells transfected with pDK101; MT302, TetON HEK293T$_{LD}$ cells transfected with MT302 vector (native L1 carrying a 3xFLAG tag in the C-term of L1-ORF2p).*

Notably, when capturing L1-ORF1p, we were able to detect L1-ORF1p in the elutions from transfected PA-1, both as PC or DC, and even on untransfected PA-1 cells (i.e., capture of endogenously expressed L1-ORF1p). However, the L1-ORF1p signal observed in transfected cells likely corresponds to the combination of endogenous and ectopic (vector-derived) expressed L1-ORF1p. Indeed, comparing the L1-ORF1p signal from the elution of untransfected PA-1 with that obtained using transfected material, we observed that only half of the L1-ORF1p signal obtained for the elutions of PA-1 DC tagged, PA-1 PC tagged and PA-1 PC untagged was derived from ectopic expressed L1-ORF1p. In sum, and while transfected cells might contain slightly more than twice L1-ORF1p compared to untransfected PA-1 cells, our expression levels are much lower than those achieved using MT302-HEK293T$_{LD}$ cells (**Figure 34**).

In conclusion, using transfection and long-term selection of episomal plasmids to overexpress proteins of interest in PC and DC PA-1 cells (i.e., tagged L1-ORF1p or L1-ORF2p), our expression analyses indicate that expression levels achieved are not high enough to implement previous protocols optimized by LaCava's group using transformed human cells (that is, using HEK293T).

2.1.1. Generation of inducible TetOn Advanced (Clontech) PA-1 cell lines to overexpress LINE-1 proteins.

To increase ectopic LINE-1 expression levels in PCs/DCs, we next decided to generate a Tet-On system using PA-1 cells, allowing to overexpress proteins of interest after doxycycline induction; exploiting the Tet-On method in PA-1 cells would create a system similar to those previously used with success using HEK293T cells by the LaCava lab (described in Dai *et al*, 2012; Taylor *et al*, 2013). Briefly, the pTet-On Advanced system is used to generate stable Tet-On Advanced PA-1 cell lines that contain a doxycycline-inducible gene expression system. In doing that, transfection with a vector containing the protein complex of interest under the control of a Tet-sensitive element (TRE-Tight or TRE2) into the generated PA-1 Tet-On cell lines, would allow rtTA-Advanced to bind TRE and activate the transcription/translation of the protein complex of interest, using doxycycline to induce and in a dose-dependent manner (**Figure 35**). In principle, and if the system behaves as in HEK293T cells, it would allow to overexpress L1 encoded proteins in PA-1 cells in sufficient amount and in an inducible manner, which ultimately would allow us to implement previous proteomic protocols optimized by LaCava's group to study the L1-interactome.

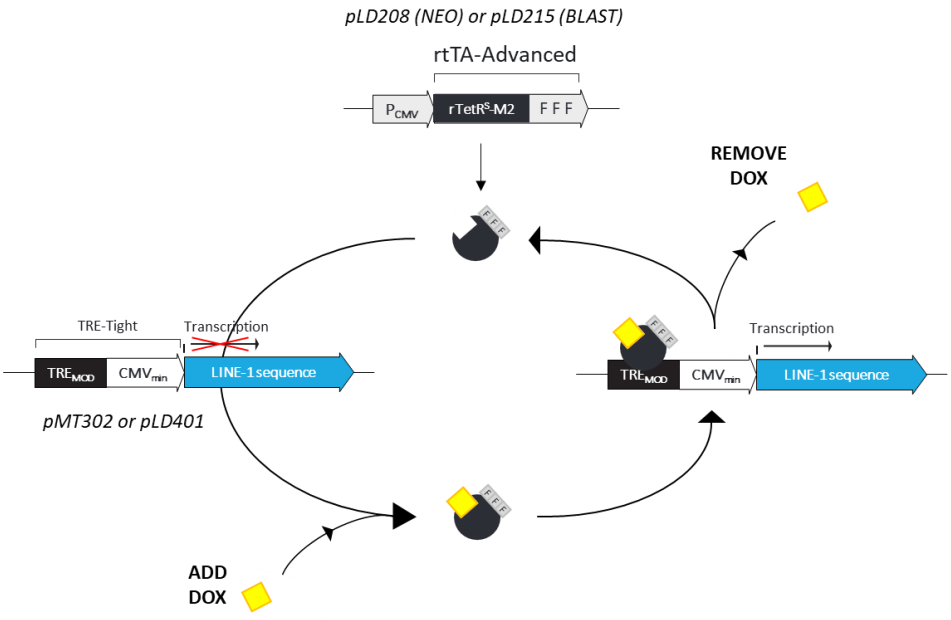


Figure 35. Schematic of gene regulation in the Tet-On Advanced System. *We generated two groups of PA-1 Tet-On cell lines, after transfecting with plasmid pLD208 (carrying rtTA-Advanced and a NEO resistance gene) or with plasmid pLD215 (carrying the rtTA-Advanced and a BLAST resistance gene). Once the rtTA-Advanced regulator is integrated in the genome of cells, the*

rtTA-Advanced protein is constitutively expressed. Next, when rtTA-Advanced cell lines are transfected with a vector encoding a gene of interest under the transcriptional control of a TRE-Tight element [in our case a full-length human L1 with epitope-tagged ORFs, and either natural (pLD401) or codon optimized (pMT302)] protein overexpression can be induced with doxycycline (DOX). Dox binds the rtTA-Advanced regulator, which allows activating gene expression of genes located downstream of TRE-Tight promoters. Once DOX is removed from the culture media, rtTA no longer can bind promoters, stopping expression of the gene of interest (i.e., L1 encoded proteins).

Using pluripotent PA-1 cells, we generated two independent groups of Tet-On cell lines, using vectors conferring neo or blast-resistance to cells [plasmids LD208 (rtTA-Advanced and neo resistance gene) and LD215 (rtTA-Advanced and blast resistance gene)]. To generate these cell lines, I collaborated with Dr. Paolo Mita (NYU Langone, NY, EEUU), as he had experience with this same system, and following the procedure described in *VI. MATERIALS AND METHODS.* Upon plasmid transfection and antibiotic selection, we generated several clonal cell lines which were subsequently characterized in detail, to demonstrate that they could support L1 retrotransposition/silencing, and that they could be used to overexpress proteins of interest upon transfection and Dox addition.

As described in the introduction, our interest in using embryonic carcinoma cell lines like PA-1 (as well as others described in Garcia-Perez *et al,* 2010*)* is double, as in one hand these are pluripotent cells mimicking a cellular niche where endogenous L1s retrotranspose in the human population (early embryogenesis), but also because pluripotent cells are capable of silencing new L1 insertions. While we know some aspects of the L1-silencing response operating in PCs, we don't know which host factors mediated L1-silencing nor how the mechanism of silencing work at a mechanistic level. Because I aim to test whether proteomic approaches could be used to dissect the mechanism of L1-silencing, an important prerequisite of the PA-1 Tet-On cell lines generated is that L1-silencing occur as in naïve cells. Thus, we first tested L1-silencing in the clonal PA-1 Tet-On cell lines generated, following the assay originally described in Garcia-Perez *et al,* 2010 (**Figure 36**). To do that, naïve PA-1 cells or the PA-1 Tet-On cell lines were transfected with a plasmid containing a "hot" RC-L1 (LRE3; Brouha *et al*, 2003) tagged with an *EGFPmI* retrotransposition indicator cassette (Ostertag, 2000), which express the fluorescent marker EGFP only after a round of L1 retrotransposition (plasmid 99-UB-LRE3, **Figure 36**); 8 days after transfection, an aliquot of transfected cells is treated with 500nM Trichostatin A (TSA, a histone deacetylase inhibitor), and 18h after the treatment the percentage of cells expressing EGFP is determined using cytometry (i.e., FACS). Because in pluripotent cells *de novo* L1 insertions are strongly silenced by histone modifications shortly after their insertion, very little EGFP is produced from

retrotransposition events. However, TSA can rapidly revert silencing of new L1 insertions, increasing the amount of EGFP expressed from retrotransposition events, and allowing us to measure the efficiency of L1-silencing by simply comparing the percentage of EGFP expression in untreated and TSA-treated populations. Notably, as *bona fide* pluripotent cells, we observed that L1-silencing is robust in the panel of PA-1 Tet-On cell lines generated (6 lines tested, see **Figure 36**), and that the overall efficiency of L1 retrotransposition is similar to that detected in naïve PA-1 cells (compare overall EGFP-expression percentages in TSA-treated cells, **Figure 36**). Indeed, and while we detect variability in the strength of L1-silencing, L1-silencing was as efficient as in naïve cells in all but one clonal line (Clon 46, **Figure 36**). As additional controls, we also used HeLa cells to explore L1-silencing. However, as differentiated-like cultured cells, no L1-silencing was observed in HeLa cells, and TSA had a very minor effect on the percentage of EGFP-expressing cells detected (**Figure 36**). We also demonstrated that an allelic L1 containing two missense mutations in the RNA binding domain of L1-ORF1p (RR261/62AA, construct 99-UB-JM111 (Moran *et al*, 1996; Coufal *et al*, 2009) failed to retrotranspose in all cell lines tested (not shown). In sum, most of the PA-1 Tet-On cell lines generated retained their capacity to silence new L1 insertions (as described in Garcia-Perez *et al*, 2010*)*.

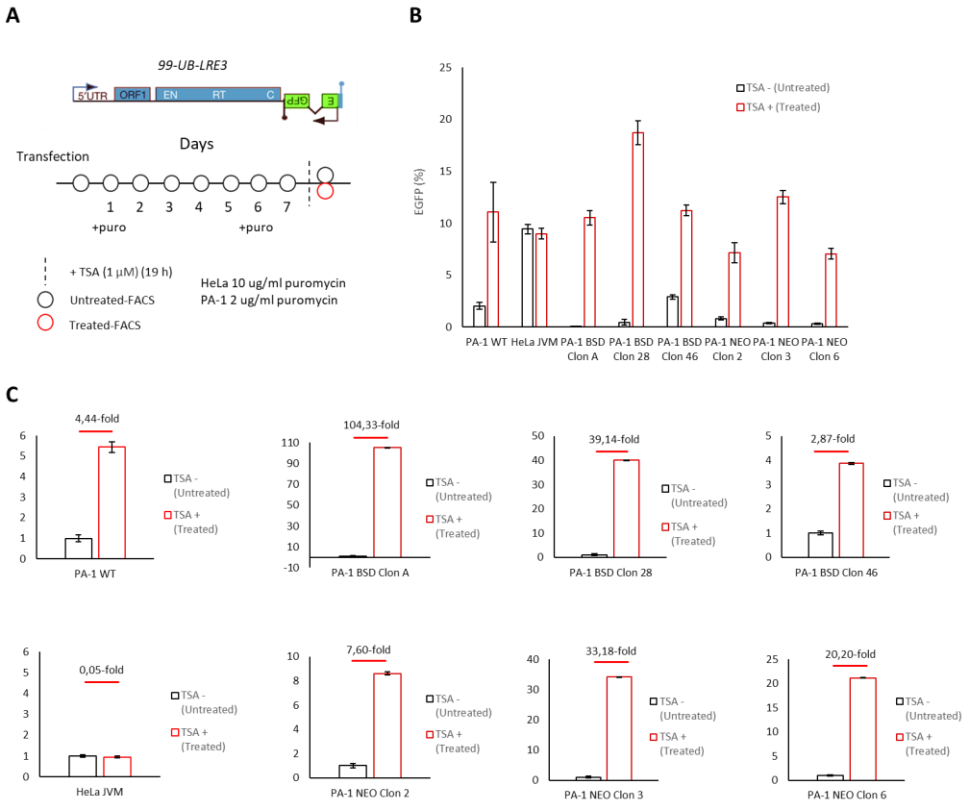


Figure 36. Testing L1 silencing in the panel of PA-1 Tet-On cell lines generated. A) *Schematic of plasmid 99-UB-LRE3 and outline of the L1-silencing assay (see main text for details).* ***B)*** *The graph depicts the percentage of EGFP-expressing cells detected in retrotransposition assays using naïve PA-1 cells, the new 6 PA-1 Tet-On cell lines, and HeLa cells as a negative control; black bars, percentage of EGFP-expressing cells in vehicle-treated cultures; red bars, percentage of EGFP-expressing cells in TSA-treated cultures (during 18 h); the graph shows data from triplicate assays (technical), and the SD of the assay is also indicated.* ***C)*** *Each graph depicts L1-silencing results in the indicated cell line, using the nomenclature described in panel B; the strength of L1-silencong is also indicated (i.e., fold change differences among vehicle and TSA-treated cells). BSD, clones transfected with the pLD215 vector (conferring resistance to blast); NEO, clones transfected with the pLD208 vector (conferring resistance to neo/G418).*

Once validated for L1-silencing, we next tested how well protein induction would work in these new cell lines. To do that, we specifically explored whether the new PA-1 Tet-On clonal lines would overexpress L1-ORF1p upon transfecting with a TRE-Tight driven L1 expression vector and upon doxycycline addition, with the ultimate goal of selecting the clonal line with the highest overexpression of L1-ORF1p. Briefly, PA-1 Tet-On clonal lines were transfected, in duplicate, with

plasmid pMT302 (contains a native human "hot" RC-L1 with a 3xFLAG-epitope in the C-terminus of L1-ORF2p under the transcriptional control of a TRE-Tight promoter), and 24h later transfected cells were selected with puromycin during 4 days (pMT302 contains a puromycin resistance cassette in the backbone). After selection, doxycycline was added to one of the replicates during 24-48 hours (**two concentrations were used, 1 and 2 μg/ml**), allowing overexpression of L1-encoded proteins. Western blot was used to measure L1-ORF1p overexpression in doxycycline (+) or untreated controls (-). Remarkably, we observed increased expression of L1-ORF1p upon Dox addition in most PA-1 Tet-On clonal lines tested (**Figure 37A**). While we observed variability in the strength of L1-ORF1p induction and expression using PA-1 lines, we also noticed that the levels of expression achieved are much lower than when using HEK293T$_{LD}$ cells, which were used as a positive control in parallel (**Figure 37A**). We suspect that the presence of large T antigen from the SV40 virus in HEK293T cells, which helps to maintain a high plasmid copy number (Lin *et al,* 2014*)*, might be responsible for the large difference with PA-1 cells. However, PA-1 cells do not express this viral antigen, which also induces undesirable changes in cells, such as loss of genomic stability. Alternatively, it is possible that not enough rtTA-Advanced regulator might be expressed in our panel of PA-1 Tet-On clonal lines. Thus, we next explored expression levels of the rtTA-Advanced regulator in our panel of PA-1 Tet-On clonal lines, using Western blot. Pellets from PA-1 Tet-On clonal cell lines were analyzed using a modified Western blot, termed Wes (Protein Simple), together with an anti-rtTA antibody. We noticed variable expression levels of the rtTA-Advanced regulator, independent of whether clonal lines were generated using pLD215 (confers resistance to blast) or pLD208 (confers resistance to NEO) (**Figure 37B**), but we noticed that expression levels of the rtTA-Advanced regulator were much lower than those detected in HEK293T$_{LD}$ cells, used in parallel as a control (**Figure 37B**). To note, and as expected, we observed a very good correlation between expression of the rtTA-Advanced regulator and expression of L1-ORF1p upon Dox addition (see Clon 3, **Figure 37A and B**), suggesting that the Tet-On system operates well pluripotent human PA-1 cells (i.e., expression levels of the rtTA-Advanced regulator are proportional to the level of induction that can be obtained upon Dox addition). While we don't know why the rtTA-Advanced regulator cannot be expressed at a high level in PA-1 cells, previous reports suggested that rtTA is toxic for most cell lines (Morimoto & Kopan, 2009), which is an inherent limitation of the Tet-On system (Benabdellah *et al,* 2011*)*. In conclusion, and while the Tet-On system operates correctly in pluripotent PA-1 cells, the overall overexpression level of L1-ORF1p is not enough to implement previous proteomic protocols optimized by LaCava's group.

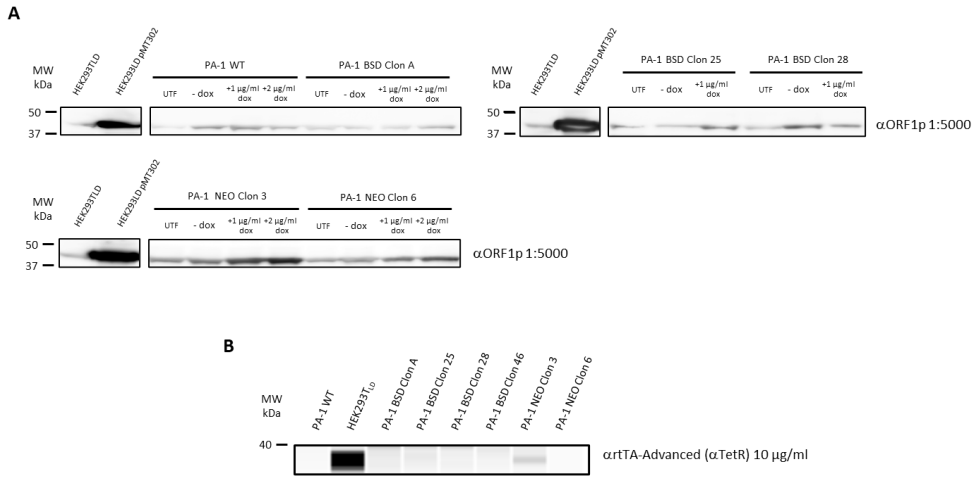


Figure 37. PA-1 Tet-On cell lines induce expression of L1-ORF1p upon doxycycline addition. A) *Representative results from a doxycycline induction assay (details described in the text). Clonal cell lines were cultured in 6-well plates and L1-protein expression induced with doxycycline during 24 h [using 1 µg/ml for HEK293T$_{LD}$ and 1 or 2 µg/ml for PA-1 Tet-On clones). L1-ORF1p expression was subsequently analyzed using Western blot.* ***B)*** *Wes (Protein Simple) using PA-1 Tet-On clonal cell extracts. PA-1 WT and HEK293T$_{LD}$ were used as controls. Expression of the rtTA-Advanced regulator is much lower than when using HEK293T$_{LD}$.*

2.2. Capture of endogenous L1-RNPs using PC and DC PA-1 cells.

Because expression of ectopic L1-encoded proteins in hECs proved challenging, we next explored and alternative experimental design, where we took advantage of the natural overexpression of endogenous L1-RNPs typical of hECs. As previously reported, several hECs (such as PA-1, NTera2D1, 833KE and 2102Ep cells) are characterized for naturally overexpressing endogenous L1 RNAs and L1-RNPs (Hohjoh & Singer, 1997; Kulpa & Moran, 2005; Martin, 1991; Garcia-Perez *et al*, 2010*)*. Remarkably, we demonstrated that L1-RNPs are also present at a high level during early differentiation of PA-1 cells (not shown but see **Figure 34** as an example), making the comparison between PC and DC interactomes a valid alternative to establish the mechanism of L1-silencing. Indeed, we found that L1-ORF1p expression is similar in PCs and in DCs, at least 7-14 days after the initiation of cellular differentiation using RA (not shown). Furthermore, we have demonstrated that a fraction of endogenously expressed L1-RNPs in PA-1 cells are derived from RC-L1s, as revealed by the characterization of *de novo* insertions using Next Generation DNA Sequencing (NGS) methods (unpublished data from the lab).

In the previous section (*2.1. Capture of ectopic L1-RNPs in PCs and DCs*), we optimized a protocol to affinity capture L1-RNPs using anti-human_L1-ORF1p antibody coupled magnetic dynabeads (mαORF1p), and we demonstrated that endogenous L1-RNPs can be captured from naïve pluripotent PA-1 cells (**Figure 34**). However, the yield of endogenous L1-ORF1p from PA-1 PC cells was still several times lower than that obtained using HEK293T$_{LD}$ cells expressing pMT302. Thus, we further optimized L1-ORF1p affinity capture, in order to obtain a yield high enough to perform the planned LC-MS/MS proteomic analyses. To do that, we applied the same experimental procedure described above, starting with the large culturing of PC and DC PA-1 cells, to obtain enough cells to produce BBs, which subsequently were cryomilled to obtain the starting material for the affinity capture experiments.

2.2.1. Optimization of endogenous L1-ORF1p affinity capture.

We started by testing several ratios of PA-1 PC or DC cell powder/anti-human-L1-ORF1p antibody (mαORF1p) coupled magnetic dynabeads, in order to find the ratio which gives a similar L1-ORF1p yield to that obtained using HEK293T$_{LD}$ cells expressing pMT302. Upon traditional Western blot and Wes (Protein Simple) system analyses, we found that PC PA-1 approximately express 10-times less L1-ORF1p compared to HEK293T$_{LD}$-pMT302 (**Figure 38**). Captured complexes were also analyzed running a 4-12% Bis-Tris Sypro stained gel, allowing to visualize affinity captured proteins and their interactors. While we readily detected captured L1-ORF1p in the Sypro-stained gel, we also observed an intense band over ~50 kDa in PC and DC PA-1, but not in HEK203T$_{LD}$ cells expressing pMT302 (**Figure 38**).

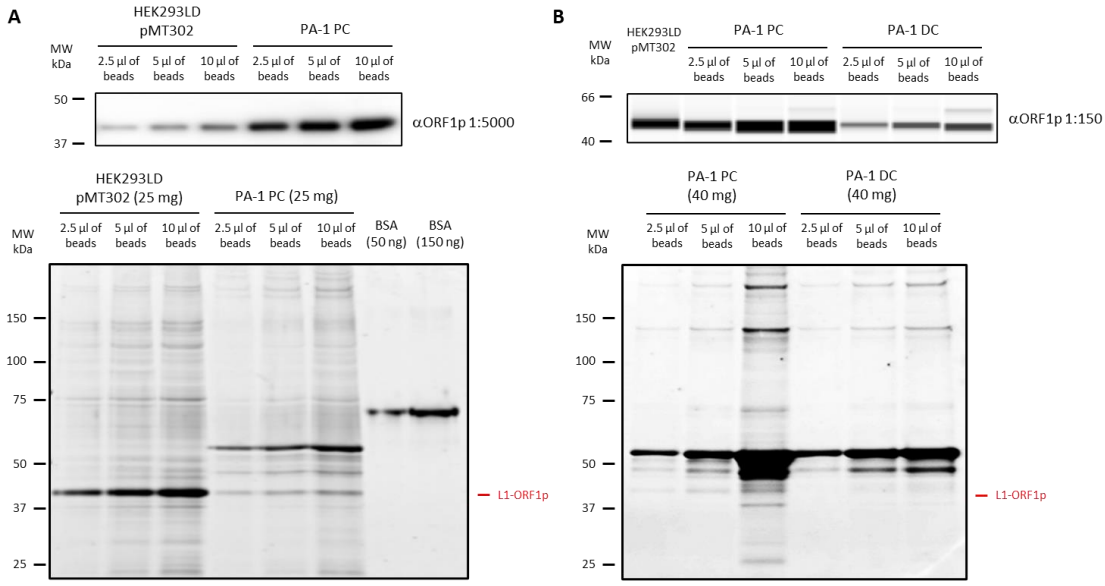


Figure 38. Optimization of capturing using mαORF1p beads/PA-1 cell powder. A) *A screening employing different volumes of mαORF1p beads and 50 mg of PA-1 PC/DC cell powder. Each elution was collected in 20 µl of 2% SDS, 40 mM Tris pH 8 buffer. HEK293TLD pMT302 was processed in parallel as a control (50 mg of cell powder and 5 µl of mαORF1p beads). The equivalent to 25 mg of cell powder (both for PA-1 and HEK293TLD) was loaded in the Western-blot. The material (considered in mg or cell powder) loaded in Sypro-stained gel is indicated in the figure. BSA, bovine serum albumin was loaded as a control.* ***B)*** *Same screening than in A) but elution was performed in 15 µl of buffer. Only 3 µl of elution (equivalent to 10 mg of cell powder from PA-1 PC or DC) were loaded for Wes (Protein simple). HEK293TLD pMT302 was processed in parallel as a control (50 mg of cell powder and 5 µl of mαORF1p beads) and the elution volume equivalent to 1.5 µg of cell powder was loaded in the Wes run. The material (considered in mg or cell powder) loaded in the Sypro-stained gel is indicated in the figure. Note: For Wes we have verified that L1-ORF1p runs higher than expected (data not shown).*

In order to confirm whether the ~50 kDa band was specific for PA-1 cells, either as a contaminant unspecific protein or as a L1-ORF1p specific interactor, we next employed magnetic dynabeads coupled to a different anti-human_L1-ORF1p antibody, using HEK293TLD as a control (**Figure 39**). In the first experiments (**Figure 38**), the anti-human_L1-ORF1p antibody coupled magnetic dynabeads used was prepared using a commercially available mouse monoclonal anti-human_L1-ORF1p antibody (mαORF1p) (clone 4H1 - Sigma-Aldrich, Cat. MABC1152). Next, as a second capture antibody, we used an in-house generated llama-derived anti-human_L1-ORF1p nanobody (LαORF1p, *llama anti-ORF1p*, from Dr. M. Rout laboratory, The Rockefeller University, US). Llama nanobodies

are fragments of standard antibodies consisting only of a single monomeric variable domain from the heavy chain of the standard antibodies (V_HH), having a lower molecular weight (12-15 kDa) and being found in camelids (**Figure 39**)(Harmsen & De Haard, 2007). Thus, we compared the efficiency of affinity capture using dynabeads coupled to mαORF1p or to LαORF1p, using HEK293T$_{LD}$ untransfected or pMT302 transfected cells as a source of L1-ORF1p. Remarkably, we observed a similar pattern of bands using either antibody, and the unknown protein of ~50 kDa was present when using any of these two antibodies, although was detected only when using untransfected HEK293T$_{LD}$ cells (**Figure 39**). However, the ~50 kDa signal was less intense in the elutions from nanobodies (LαORF1p). We speculate that when L1-ORF1p expression is not as high as in HEK293T$_{LD}$ pMT302 overexpressing cells, an unidentified protein of ~50 kDa is captured by anti-ORF1p antibodies (**Figure 39**). Thus, we decided to use both types of antibodies to study the LINE-1 interactome in PA-1 cells.

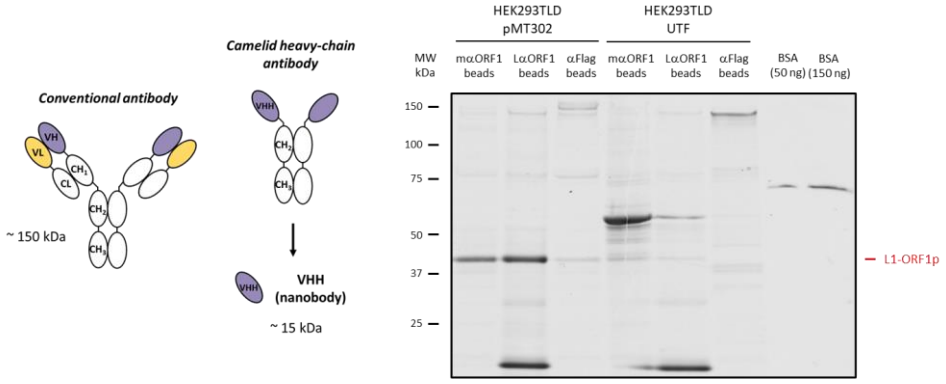


Figure 39. Comparison of anti-ORF1p magnetic dynabeads. *Left, schematic of a conventional antibody (left side) and of a camelid nanobody (right side). Right, representative affinity capture results using mαORF1 or LαORF1 beads. The scale was 50 mg of cell powder and 10 ul of beads, for HEK293T$_{LD}$ expressing pMT302 or untransfected (UTF). The elution volume equivalent to 25 mg was loaded on a 4-12% Bis-Tris gel. As expected, L1-ORF1p is efficiently captured when using HEK293T$_{LD}$ pMT302; however, we also detected captured L1-ORF1p in UTF cells, although at a much lower level. The ~50 kDa band was predominantly observed when using UTF HEK293T$_{LD}$, using either mαORF1 or LαORF1 beads.*

Next, we continued the optimization of the capturing, testing different ratios of cell powder (mg)/ anti-human_L1-ORF1p magnetic dynabeads; to do that, we used both coupled antibodies, that is the mouse monoclonal (mαORF1p) and the nanobody (LαORF1p), and we used PA-1 PC and HEK293T$_{LD}$-pMT302 cells, in order to identify the combination providing the higher yield of L1-ORF1p (**Figure 40**). Notably, we determined that the ratio which gives the highest yield of L1-ORF1p is 10 μl of coupled magnetic dynabeads per 50 mg of PA-1 PC powder, for

both mouse monoclonal (mαORF1p) and the nanobody (LαORF1p) (**Figure 40**). The volume of extraction buffer (20 mM HEPES, pH 7.4, 1% Tritonx-100, 500 mM NaCl) used per mg of cell powder was stablished as ¼ (w/v) (the standard volume routinely used by LaCava's team); similarly, the sonication conditions were determined in relation to the amount of starting material, choosing an energy of 8.5 J for every 25 mg of cell powder used. To establish the sonication energy, several sonication programs were tested, and the resulting cell pellets after clarification of the extract were analyzed, selecting the condition yielding the clearest extract and with the smallest pellet. To note, the temperature of extracts during the sonication test was monitored, ensuring that it didn't increase.

Next, we determined the amount of starting material (mgs of cell powder) to carry out L1-ORF1p affinity capture and LC-MS/MS with a high enough yield to obtain a signal intensity similar to that obtained when using pMT302-HEK293T$_{LD}$ overexpressing cells. The results obtained were key to conduct the final capture experiments, despite using only triplicates of each sample and no negative controls to detect false positives. In conclusion, the affinity capture of L1-ORF1p and its interactors in PCs was performed using 100 mg of PA-1 PC cell powder (and 100 mg of HEK293TLD overexpressing pMT302 in order to compare the intensities), 20 µl of mouse monoclonal anti-human_L1-ORF1p antibody coupled magnetic dynabeads (mαORF1p), and the elution was carried out in 18 µl NuPAGE LDS Sample Buffer (4X)(ThermoFisher), which were finally loaded in Gel-plugs prepared as described Taylor *et al,* (2018), followed by LC-MS/MS analyses. Mass spectrometry analyses were carried out by experienced researchers at Rockefeller University (Dr. Kelly Molloy from Dr. Chait's Lab, a mass spectrometry specialized lab).

To establish the scale of the affinity captures required to analyze endogenous L1-ORF1p and is interactors in PC and DC PA-1 cells by mass spectrometry, we used the resulting intensity of the top 50 hits obtained, which were overlapped with the list of factors previously detected by LaCava's team using pMT302-HEK293T$_{LD}$ overexpressing cells (as described in Taylor *et al,* 2018*).* By scaling up immunoprecipitations with 100 mg of PA-1 PC powder and comparing the results with those obtained using 100 mg of powder from pMT302-HEK293T$_{LD}$ cells, the analyses of detected protein intensities determined that affinity captures must be performed at a scale of 400 mg per replicate, in order to get a high enough intensity. Additionally, for subsequent experiments we used the S-Trap™ Micro Ultra-High Recovery Protocol to prepare mass spectrometry samples (protein sample preparation method for mass spectrometry that provides high peptide recovery and shorter processing time), rather than using gel plugs. Indeed, this was the only

difference with previous proteomic analyses using pMT302-HEK293T$_{LD}$ overexpressing cells.

As a negative control when using the mouse monoclonal anti-human_L1-ORF1p antibody coupled magnetic dynabeads (mαORF1p), we used dynabeads coupled with mouse IgG (mIgG), as both antibodies share the same constant region. When using llama-derived anti-human_L1-ORF1p nanobody coupled magnetic dynabeads (LαORF1p), we used another in-house generated nanobody, directed against the EGFP protein (LαG94-10), as both share the structure (but not the target) of the nanobody. We prepared and tested several batches of each type of magnetic dynabeads (mαORF1p, mIgG, LαORF1p and LαG94-10 beads), as we required as much as 400 µl of each dynabead type. To note, as when using mouse monoclonal anti-human_L1-ORF1p antibody coupled magnetic dynabeads (mαORF1p), we performed a screening to identify the best ratio of LαORF1 beads and cell powder, using the same protocol employed with mαORF1p beads (**Figure 40**).

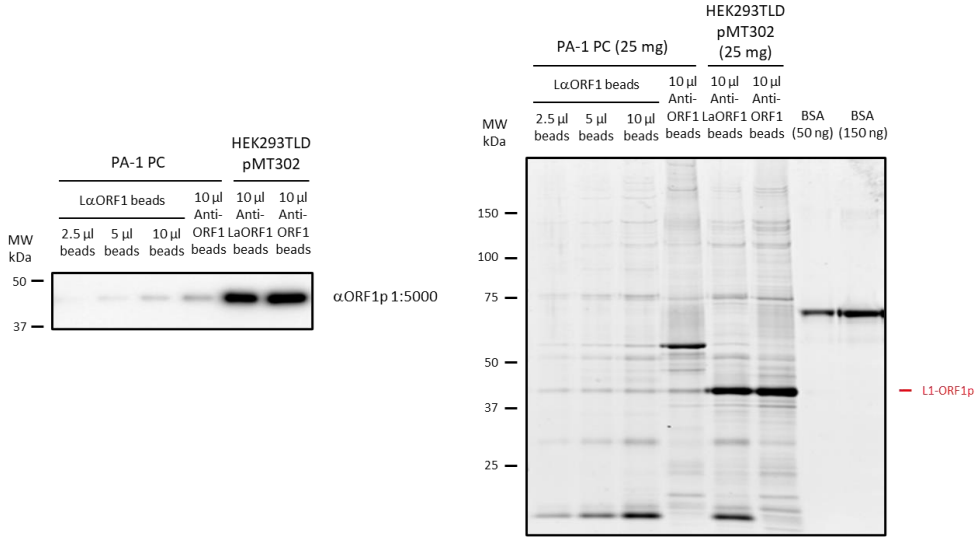


Figure 40. Optimization of LαORF1p beads/PA-1 cell powder ratio. *A screening employing different volumes of LαORF1p beads for 50 mg of PA-1 PC/DC and HEK293T$_{LD}$ pMT302 cell powder was carried out. Each elution was performed in 20 µl of 2% SDS, 40 mM Tris pH 8 buffer. The equivalent to 25 mg of cell powder (both for PA-1 and HEK293T$_{LD}$) was loaded on the Western-blot. The material (considered in mg of cell powder) loaded in the Sypro-stained gel is indicated in the figure. BSA, bovine serum albumin, loaded as a control.*

Notably, in the screening using llama-derived anti-human_L1-ORF1p nanobody coupled magnetic dynabeads (LαORF1p), we reproducibly detected the same ~50 kDa band. Thus, we decided to identify the nature of the ~50 kDa protein/s by LC-MS/MS; to do that, we repeated the affinity capture using the antibodies described in the figure, loading the equivalent to 100 mg of cell powder por PA-1 PC. Notably, we found that the ~50 kDa band corresponded to vimentin (shown in **Figure 41**). However, further assays are required to determine whether this is an unspecific contaminant protein or whether vimentin can indeed interact with L1-ORF1p in PA-1 cells (and untransfected HEK293T cells).

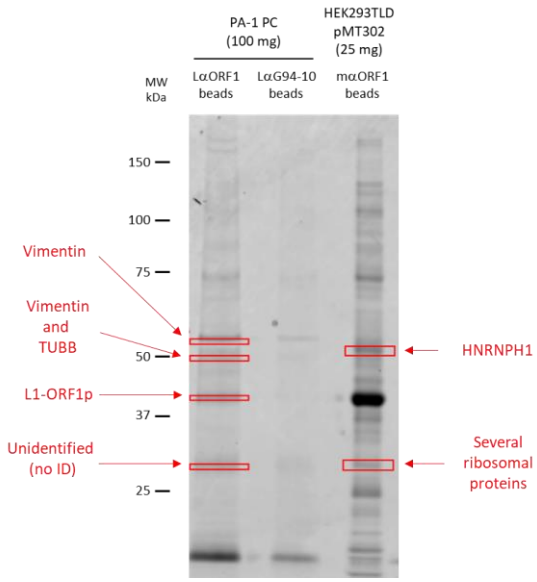


Figure 41. Identification of recurrent bands in L1-ORF1p affinity captures. *Two affinity captures using 20 µl of LαORF1 and LαG94-10 on PC PA-1 (100 mg of cell powder each), and one affinity capture using mαORF1 on HEK293TLD-pMT302 (25 mg, 5 µl of beads) were performed. Highlighted bands were cut and processed for identification by LC-MS/MS.*

2.3. Capturing L1-RNPs and their interactors on PC and DC PA-1 cells.

After optimizing the affinity capture conditions to increase the yield of L1-ORF1p to levels required for LC-MS/MS analyses, we next established the interactome of endogenously expressed L1-RNPs using PC and DC PA-1 (i.e., capturing L1-ORF1p and its interactors to study the composition of the L1 interactome in PCs and DCs). Specifically, we used mouse monoclonal anti-human_L1-ORF1p antibody coupled magnetic dynabeads (mαORF1p) to capture L1-ORF1p and interactors from PC and DC PA-1 cells, and we used llama-derived anti-human_L1-ORF1p nanobody coupled dynabeads (LαORF1p) to capture L1-ORF1p and interactors from PC PA-1 cells as an internal control (see below).

We designed experiments considering previously described optimization variables (sonication, extraction buffer, material weight/antibody coupled dynabeads ratio) (**Figure 42**); additionally, to increase protein yield, we used 500 mg of cell powder in each replicate. Because the volume when using 500 mg of cell powder was too large to process, each replicate was split into two separate tubes, each with 250 mg, although elutions were collected on a single tube (i.e., the elutions from 2x 250mg captures were combined). As described above, we used the S-Trap™ Micro Ultra-High Recovery Protocol to prepare samples, to increase peptide recovery (**Figure 42**).

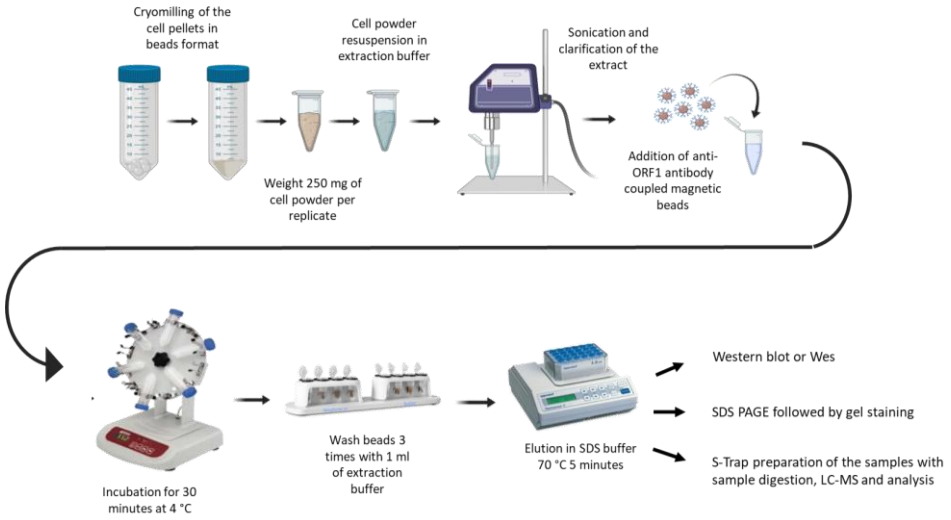


Figure 42. Schematic of the optimized affinity capture protocol of L1-ORF1p and interactors using mouse monoclonal antibody (mαORF1p) or llama derived nanobody magnetic beads (LαORF1p). *The protocol is described in detail in section VI. MATERIALS AND METHODS, 19.*

2.3.1. Capture of L1-ORF1p and interactors from PC PA-1 using affinity capture and mouse monoclonal anti-human_L1-ORF1p antibody coupled magnetic dynabeads (mαORF1p).

The interactome of L1-RNPs in pluripotent PA-1 cells (PCs) was established using the optimized affinity capture protocol and dynabeads coupled to mouse anti-human_L1-ORF1p antibody (mαORF1p). In parallel, as a negative control, we used the same optimized affinity capture protocol as well as mouse IgG coupled magnetic dynabeads (mIgG). Using this rationale, we could identify specific interactors of L1-ORF1p in PC PA-1, by simply comparing co-immunoprecipitated proteins using the anti-human_L1-ORF1p antibody with those captured with mIgG. To obtain consistent and robust data, we performed 4 biological replicates of both Case (mαORF1p) and Control (mIgG), each using 500 mg of PC PA-1 cell powder (4 grams of PC PA-1 in total). Elution was carried out directly at room temperature for 10 min in a buffer containing 2.5% SDS, 4M urea and 50 mM glycine. A small fraction of the elution from each biological replicate, corresponding to ~75 mg, was used to perform a Western blot (25 mg), and to run a 4-12% Bis-Tris, Sypro stained gel (200 mg obtained from combining 50 mg from each biological replicate were load), in order to characterize the elutions. The remaining elution (equivalent to 425 mg) was prepared by S-Trap™ Micro Ultra-High Recovery Protocol for LC-MS/MS *(VI. MATERIALS AND METHODS, 21. Sample preparation and liquid chromatography coupled to Mass spectrometry)*. As a control of protein yield, in parallel we captured L1-ORF1p from HEK293T$_{LD}$ overexpressing pMT302 cells, which we used in Western blot and in Sypro-stained gel analyses. In fact, we confirmed that PC PA-1 has ~10-times less L1-ORF1p than pMT302-HEK293T$_{LD}$ overexpressing cells (**Figure 43**).

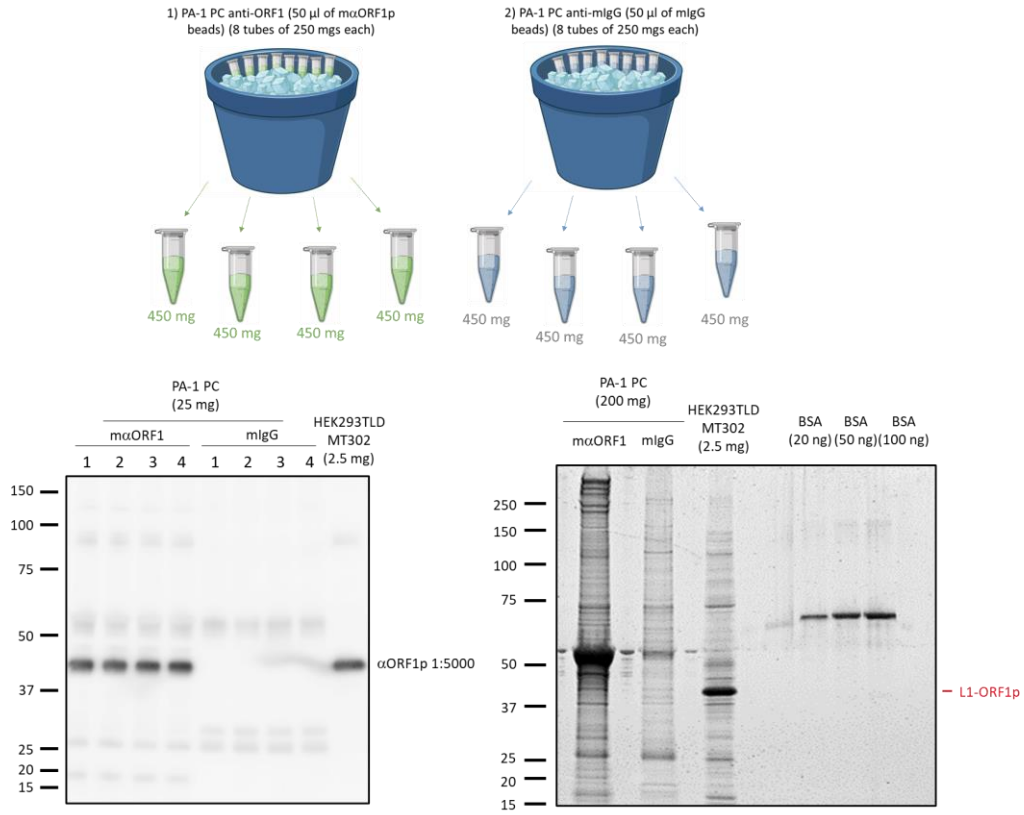


Figure 43. Capturing L1-RNPs from PC PA-1 cells using mouse anti-human_L1-ORF1p beads (mαORF1p) and mIgG as a negative control. *Shown are representative results from the four biological replicates, analyzed by Western blot (left side) and by 4-12% Bis-Tris, Sypro stained (left side), and using elutions (the relative amount of material loaded is indicated in the figure). The protocol used is described in full detail in section VI. MATERIALS AND METHODS.*

While analyzing the Sypro-stained gel, we noticed that the L1-ORF1p (~40 kDa) band was faint; however, Western blot analyses revealed that the intensity of L1-ORF1p was similar to that detected when using HEK293T$_{LD}$ overexpressing pMT302 cells. We speculate that the elution buffer/conditions used likely failed to elute all captured L1-ORF1p from the dynabeads; however, it seems that L1-ORF1p interactors were successfully eluted, as many proteins with a wide range of molecular weights were detected in the Sypro-stained gel (**Figure 43**).

After preparing samples using the S-Trap™ Micro Ultra-High Recovery Protocol, LC-MS/MS mass spectrometry was used to identify the nature of L1-interactors in PC PA-1, in collaboration with Drs Kelly Molloy (expert in LC-MS/MS) and Mehrnoosh Oghbaie (Bioinformatician expert in LaCava's team) *(VI. MATERIALS AND METHODS, 22. Mass spectrometry data analysis)*. To visualize data, we used volcano plots (**Figure 44**), allowing us to distinguish L1-ORF1p interactors in isogenic PC and DC PA-1 cells (see below). To note, mouse IgG was used to remove non-specific interactors in the final list of identified L1 interactors. Interestingly, we identified well-known L1-ORF1p specific interactors, which were described in previous studies using transformed cell lines and ectopically expressed L1s. In fact, a total of 10 specific significant interactors on PC PA-1 (corresponding to 7.5 % of specific interactors) were previously found in prior studies (**Table 4**; Goodier *et al,* 2013; Taylor *et al,* 2013; Moldovan & Moran, 2015; Taylor *et al,* 2018*)*. Additionally, we found 2 factors that were previously described as endogenous L1-ORF1p interactors in hESCs in Vuong *et al,* 2019. Thus, we found a relative low number of significant *bona fide* L1-ORF1p interactors in PC PA-1 cells (n=12, including well-known L1 regulators such as MOV10, PABPC1, HSP90AB1, TUBB, TUBB4B, ZCCHC3, LARP1, IGF2BP3, KIF5B, and YBX1) providing consistency and reliability to our results. Remarkably, most of the detected L1-interactors in PCs were novel [121 specific interactors on PC PA-1 (corresponding to 90 % of specific interactors)], and only 28 were described as not significant L1-interactors in Taylor *et al,* 2013. Among the large number of novel factors, several ZNF (Zinc finger proteins) appear to interact with L1-RNPs in PCs (see below). Thus, these data suggest that the L1-interactome is not a constant entity. It is tempting to speculate that L1 regulation might be equally variable among cell lines.

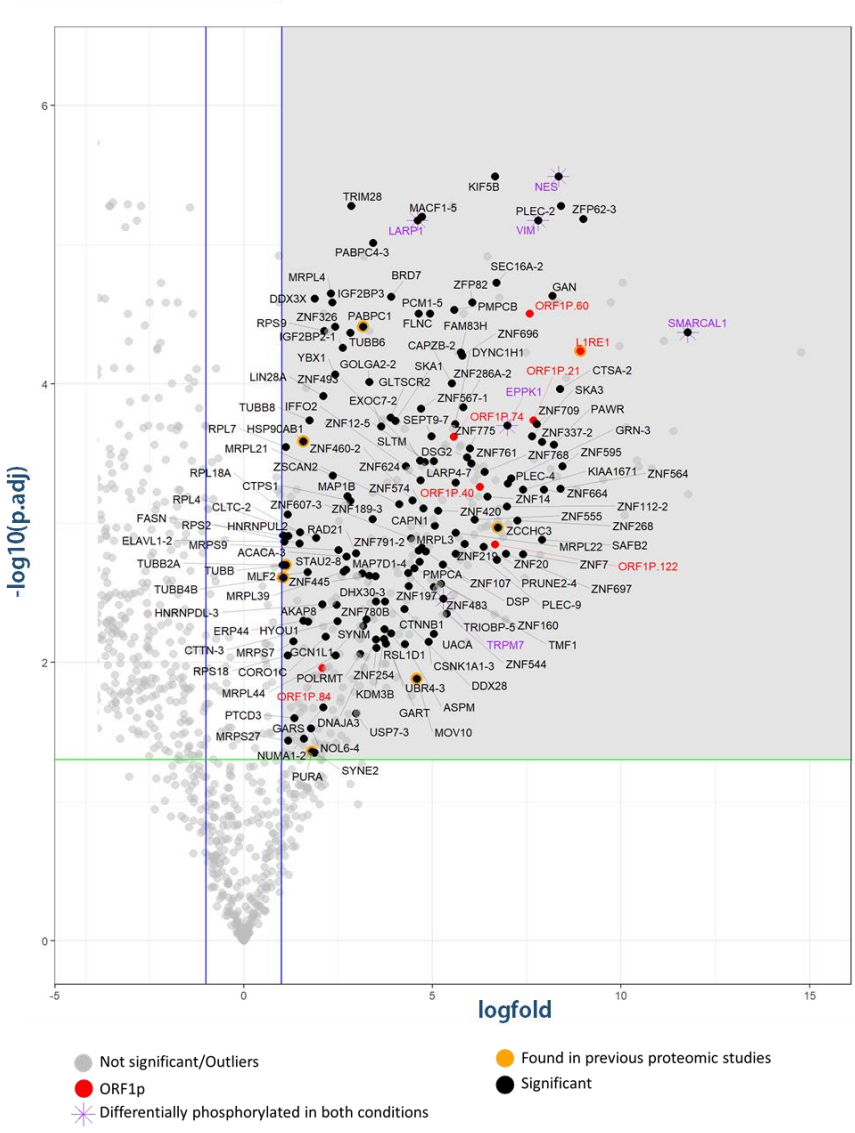


Figure 44. L1-ORF1p specific interactors found in PC PA-1 cells, detected with the mouse monoclonal anti-human_L1-ORF1p antibody (mαORF1p). *A Volcano plot indicates PC specific PA-1 interactors (see Method 2, VI. MATERIALS AND METHODS, 22.. Mass spectrometry data analysis to learn how the plot was generated). In the volcano plot, the name of the main interacting proteins of L1-ORF1p in pluripotent PA-1 cells (PC) is indicated. Highlighted in the upper right quadrant are significant L1-ORF1p interacting factors [-log^{10} of p-adjusted value < 0.05 (p-adjusted value from Benjamini Hochberg method applied to t-test p-value); and logfold > 1 (log2fold change of LFQ intensity between cases and controls)]. Highlighted in purple are factors*

with different phosphorylation; L1-ORF1p is highlighted in red; highlighted in yellow are known L1 interactors (i.e., factors found in previous studies (Taylor et al, 2013 and 2018).

Table 4. Factors identified in PC PA-1 using mαORF1p beads in common with previous studies.

	Gene names	(Goodier *et al.*, 2013)	(Taylor *et al.*, 2013)	(Moldovan and Moran, 2015)	(Taylor *et al.*, 2018)	(Vuong *et al.*, 2019)
1	MOV10	Yes*	Yes*	Yes*	Yes*	-
2	PABPC1	Yes*	Yes*	-	Yes*	-
3	HSP90AB1	-	Yes*	-	Yes*	-
4	TUBB	-	Yes	-	Yes*	-
5	TUBB4B	-	Yes	-	Yes*	-
6	ZCCHC3	-	Yes*	-	Yes*	-
7	LARP1	Yes*	Yes*	Yes*	-	-
8	IGF2BP3	-	Yes	Yes*	-	-
9	KIF5B	-	-	Yes*	-	-
10	YBX1	Yes*	Yes	-	-	-
11	ZNF268	-	-	-	-	Yes
12	ZNF624	-	-	-	-	Yes
	Factors found	4	5	4	6	0

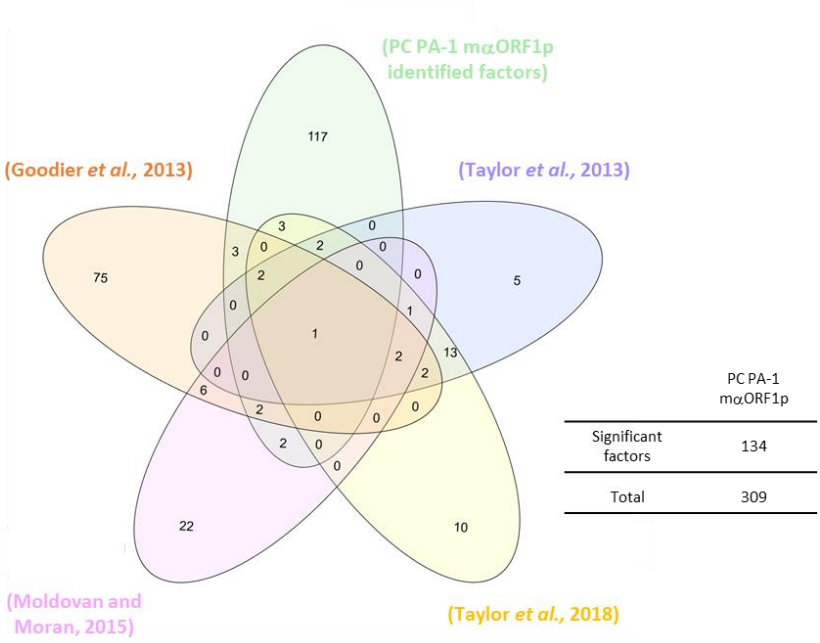


Table 4. Factors identified in PC PA-1 using mαORF1p beads that have been found in previous studies. *Asterisks indicate that the factor was significant in that condition. "Yes" indicates that the protein was identified in the study, but did not reach the significance threshold established by authors.*

2.3.2. Capture of L1-ORF1p and interactors from PC PA-1 using affinity capture and llama-derived anti-human_L1-ORF1p antibody coupled magnetic dynabeads (LαORF1p).

As described above, the analysis of the L1-ORF1p interactome on pluripotent PA-1 cells using a mouse monoclonal antibody (mαORF1p) uncovered a significant number of novel L1-interactors. To corroborate these data, and to further explore whether the composition of the L1-ORF1p interactome is influenced by the cellular background, we next analyzed the L1-ORF1p interactome using the llama-derived anti-human_L1-ORF1p antibody coupled magnetic dynabeads (LαORF1p). To note, these two antibodies are very different in size (nanobody ~13-15 kDa vs mouse monoclonal antibody ~150 kDa), a factor that could influence the efficiency and kinetic of capturing L1-RNPs, which could at the same time influence the composition of the L1-ORF1p interactome in cultured cells.

While the closest to an "apple vs apple comparison" would be the ideal situation, we decided to change some aspects of the elution process when using the llama-derived anti-human_L1-ORF1p antibody. As discussed above, we noticed difficulties in eluting all captured L1-ORF1p from beads when using the mouse monoclonal antibody. Thus, we optimized the elution of captured L1-ORF1p from beads, by testing different buffers and elution conditions. To do that, we used pMT302- HEK293T$_{LD}$ derived material at a constant ratio of 10 µl of buffer for every 25 mg of pMT302-HEK293T$_{LD}$ cell powder, analyzing the following buffers and conditions: 1) NuPAGE LDS Sample Buffer (4X)(ThermoFisher) at 70 °C for 5 minutes with mixing; 2) 2% SDS/40mM Tris, pH8, at 70 °C for 5 minutes with mixing; 3) 2% SDS/40mM Tris, pH8, at room temperature (RT) for 10 minutes, with mixing 4) 5% SDS/8M urea/100 mM glycine) at room temperature for 10 minutes; and 5) 2.5% SDS / 4M urea / 50 mM glycine at room temperature (RT) for 10 minutes, which were the conditions used above [i.e., conditions used to characterize the L1-ORF1p interactome of PC PA-1 cells using anti-human_L1-ORF1p mouse monoclonal coupled magnetic dynabeads (mαORF1p)]. The results of these analyses indicated that in general all conditions tested resulted in similar elution efficiencies (**Figure 45**); we also noticed that NuPAGE LDS Buffer was not compatible with S-Trap™, and this condition was not further considered. However, among the tested conditions, the highest efficiency was achieved when using 2% SDS 40 mM Tris pH 8 at 70 ºC for 5 min, and we decided to use these elution conditions from now on (**Figure 45**).

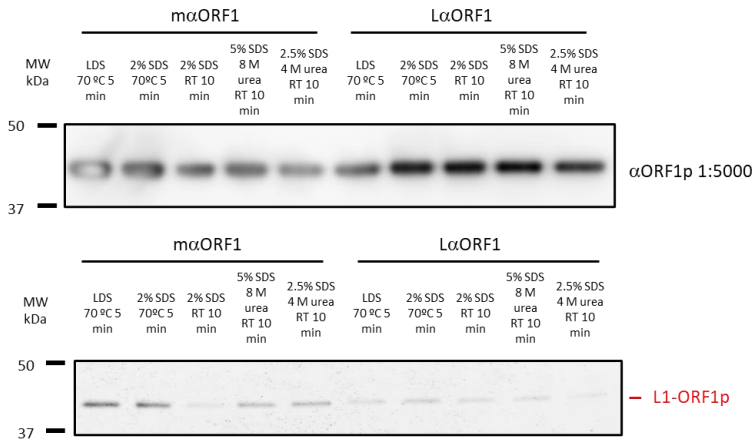


Figure 45. Elution conditions tested using pMT302-HEK293T$_{LD}$ material. *Shown are results from Western blot (top) and 4-12% Bis-Tris Sypro stained gel (bottom) analyses of L1-ORF1p elutions. LDS, NuPAGE LDS Sample Buffer (4X) (ThermoFisher). See main text for further details.*

After the optimization of the elution process, we next captured the L1-ORF1p interactome using the llama-derived anti-human_L1-ORF1p nanobody coupled magnetic dynabeads (LαORF1p) and PC PA-1 cells. We used the same procedure and conditions described above when using the mouse monoclonal antibody for capturing, except for using new elution conditions (use of 2% SDS 40 mM Tris pH 8 at 70 ºC for 5 min instead of 2.5% SDS / 4M urea / 50 mM glycine at room temperature (RT) for 10 minutes). As above, we performed 4 independent replicates for each nanobody (LαORF1p, and the negative control). As a negative control, we used the LαG94-10 nanobody coupled magnetic dynabeads (against EGFP, not expressed by PA-1 cells), and the amount of material used was of 4 grams of PC PA-1 cells (500 mg per capture) and 400 µl of each type of magnetic bead (50 µl per capture). To analyze the capture of the L1-ORF1p interactome using the nanobody in PC PA-1, we used an elution volume equivalent to 75 mg from each replicate, which were subsequently analyzed by Western blot and by 4-12% Bis-Tris, Sypro stained gel electrophoresis (**Figure 46**).

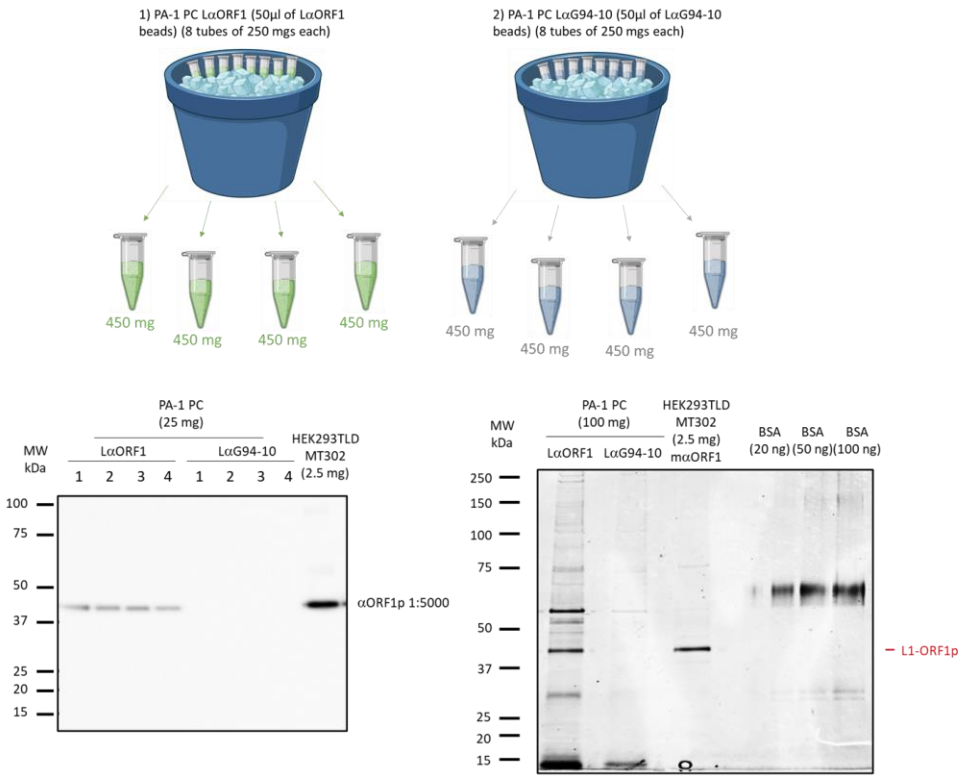


Figure 46. Capturing L1-RNPs from PC PA-1 cells using the llama-derived anti-human_L1-ORF1p nanobody beads (LαORF1p). *Schematic of the experiment. Below are shown results from Western blot (left side) and 4-12% Bis-Tris Sypro stained gel electrophoresis (right side) analyses of the elutions (amount of material loaded is indicated in the figure). Nanobody beads from LαG94-10 (anti-EGFP) were used in parallel as a negative control. Protocol described in detail in VI. MATERIALS AND METHODS.*

Compared with the results obtained using the mouse monoclonal and the former elution conditions, we observed a clear increase in the elution efficiency of L1-ORF1p from beads, although relatively lower levels of L1-interactors were observed when using nanobody beads (compare **Figures 43 and 46**). However, as above, we next processed the L1-ORF1p elutions (volume equivalent to ~425 mg per replicate) using the S-Trap™ Micro Ultra-High Recovery Protocol; after processing, samples were analyzed by LC-MS/MS (with the help of Dr Kelly Molloy), and the data generated was processed and analyzed as above (with the help of by Dr Mehrnoosh Oghbaie), to ultimately identify L1-interactors in PC PA-1 cells. Remarkably, after removing non-specific interactors using data from LαG94-10 control captures, we identified a significant number of L1-ORF1p specific interactors, which are shown in the volcano plot included in **Figure 47**. As

when using the mouse monoclonal antibody, we noticed that some L1-ORF1p interactors have been described previously in proteomic studies using ectopically expressed L1s and transformed cell lines [8 out 65 factors (12.3% of the factors) were previously found in previous studies: MOV10, PABPC1, ZCCHC3, TUBB4B, FAM120A, MATR3, DDX5, and MARS see **Table 5**; (Goodier *et al*, 2013; Taylor *et al*, 2013; Moldovan a& Moran, 2015; Taylor *et al*, 2018*)*]. However, as when using the mouse monoclonal, we identified a high number of novel L1-interactors when using the nanobody for capturing L1-ORF1p, never found in previous studies [57 out 65 (87.7% of the factors) correspond to novel L1-interactors factors absent in Goodier *et al*, 2013 and Taylor *et al*, 2013, where 16 were found but were not significant; Moldovan and Moran, 2015; Taylor *et al*, 2018*]*.

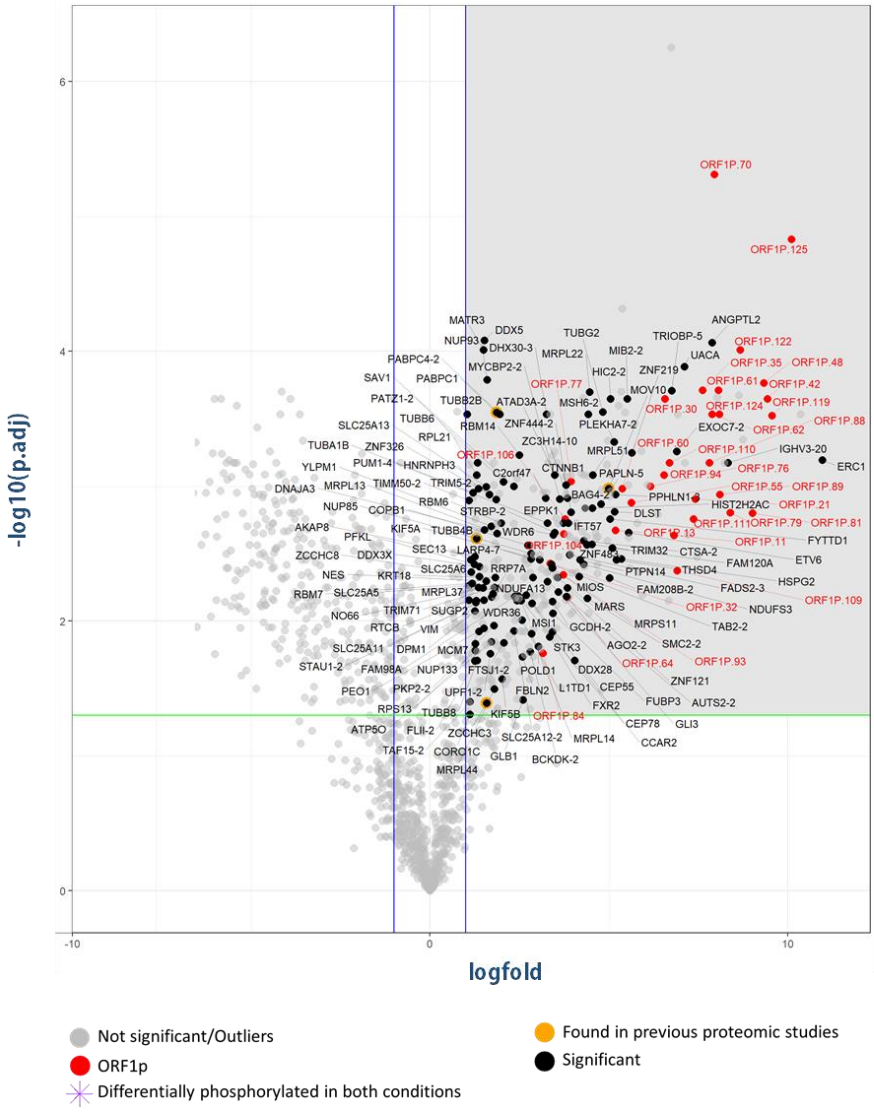


Figure 47. L1-ORF1p specific interactors found in PC PA-1 cells, detected with the llama anti-human_L1-ORF1p nanobody (LαORF1p). *The volcano plot indicates PC specific PA-1 interactors (see Method 2 (VI. MATERIALS AND METHODS, 22. Mass spectrometry data analysis). In the plot, the name of the main L1-ORF1p interacting proteins in pluripotent PA-1 cells (PC) is indicated. Highlighted in the upper right quadrant are significant L1-ORF1p interacting factors [-log^{10} of p-adjusted value < 0.05 (p-adjusted value from Benjamini Hochberg method applied to t-test p-value); and logfold > 1 (log2fold change of LFQ intensity between cases and controls)]. Highlighted in purple are factors with diferent phosphorylation, L1-ORF1p is highlighted in red, and highlighted in yellow are known L1-interactors (i.e., factors found in previous studies (Taylor et al, 2013, 2018).*

Table 5. Factors identified in PC PA-1using LαORF1p beads in common with previous studies.

	Gene names	(Goodier *et al.*, 2013)	(Taylor *et al.*, 2013)	(Moldovan and Moran, 2015)	(Taylor *et al.*, 2018)	(Vuong *et al.*, 2019)
1	MOV10	Yes*	Yes*	Yes*	Yes*	-
2	PABPC1	Yes*	Yes*	-	Yes*	-
3	ZCCHC3	-	Yes*	-	Yes*	-
4	TUBB4B	-	Yes	-	Yes*	-
5	FAM120A	Yes*	Yes	Yes*	-	-
6	MATR3	Yes*	Yes	Yes*	-	-
7	DDX5	Yes*	Yes	-	-	-
8	MARS	Yes*	Yes	-	-	-
	Factors found	6	3	3	4	0

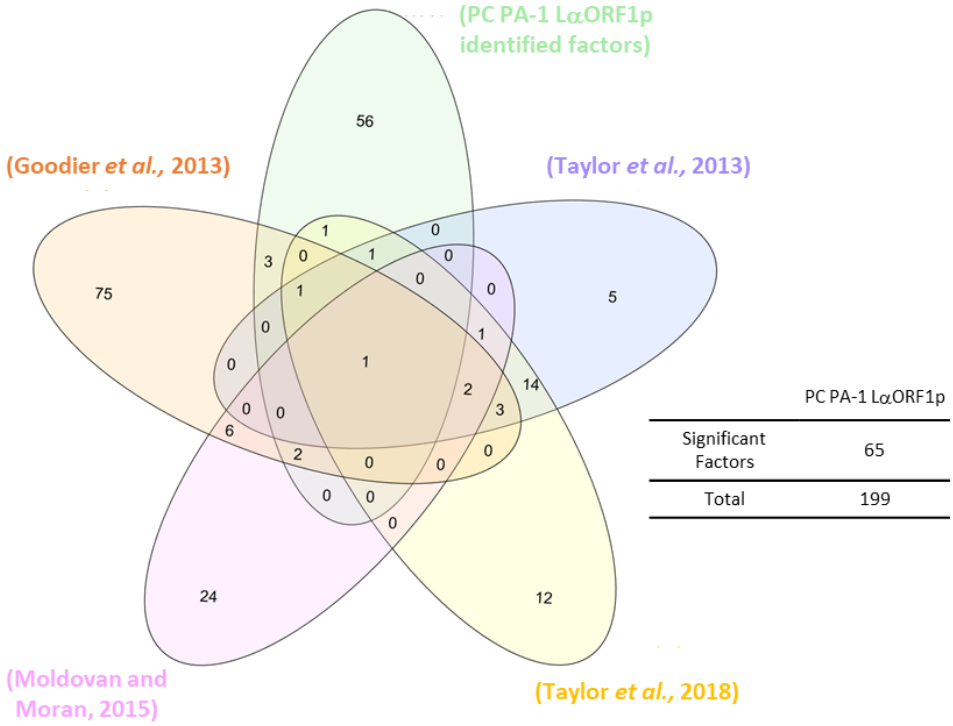


Table 5. Factors identified in PC PA-1 using LαORF1p beads that have been founf in previous studies. *Asterisks indicate that the factor was significant in that condition. "Yes" indicates that the protein was identified in the study, but did not reach the significance threshold established by authors.*

Although using different antibodies and elution conditions, we next analyzed the overlap between the PC PA-1 interactomes inferred using the mouse monoclonal or the nanobody against human_L1-ORF1p. Notably, these analyses revealed that 21 out 65 (32%) factors identified were found in both L1-ORF1p interactomes (nanobody and mouse monoclonal), including well known L1-interactors such as MOV10 (Goodier *et al*, 2013; Taylor *et al*, 2013; Moldovan and Moran, 2015; Taylor *et al*, 2018), PABPC1 (Goodier *et al*, 2013, ; Taylor *et al*, 2013; Taylor *et al*, 2018), ZCCHC3 (Taylor *et al*, 2013; Taylor *et al*, 2018) and TUBB4B (Taylor *et al*, 2018). Thus, these data indicate that our proteomic analyses are reliable and robust. Surprisingly, we also found that a significant number of L1-ORF1p interactors were unique to the nanobody or to the mouse monoclonal dataset [(44 of 65 (67.7%) factors were unique to the nanobody dataset; 113 of 134 (84.3%) were unique to the mouse monoclonal antibody dataset]. While inherent characteristics of mouse monoclonal IgG antibodies and llama nanobodies might explain the marked differences found between both L1-interactomes, their size is perhaps the main differential characteristic. The differences observed among the PC PA-1 interactomes will be further discussed in section *VIII. DISCUSSION*.

2.3.3. Capture of L1-ORF1p and interactors from DC PA-1 using affinity capture and mouse monoclonal anti-human_L1-ORF1p antibody coupled magnetic dynabeads (mαORF1p).

As described above, we characterized the interactome of L1-ORF1p on pluripotent human cells, using two different antibodies. As previously discussed, a main goal of this Thesis is to analyze whether proteomic approaches could be used to dissect the mechanism of L1-silencing, which operates only in pluripotent cells. Following this rationale, we next characterized the interactome of L1-ORF1p but on differentiated isogenic human cells (DCs). To identify PC-specific interactors, we would ultimately compare the L1-interactome of PC and DCs, and of previous studies; thus, to avoid comparative biases, we decide to use the mouse monoclonal anti-human_L1-ORF1p antibody to determine the L1-interactome of DC PA-1s, simply because all previous proteomic L1 studies have exploited conventional antibodies rather than nanobodies.

While L1 expression is attenuated/silenced in differentiated cells, we should notice that we exploited Retinoic Acid (RA, all trans) (Gudas & Wagner, 2011) to induce strong, robust and quick differentiation of PA-1 cells, as previously described (Garcia-Perez *et al*, 2010). As described, our lab previously demonstrated that L1-silencing is rapidly attenuated when PA-1 cells are differentiated, independently of the method used to differentiate hECs. Indeed, we established

that a reproducible and robust method to differentiate PA-1 cells is to culture cells during 14 days without Fetal Bovine Serum (FBS) but in the presence of 20% KnockOut™ Serum Replacement (ThermoFisher) and 1μM RA (all-trans, Sigma), termed DM for Differentiation Media; these were the differentiation conditions used in this Thesis (see *VI. MATERIALS AND METHODS*). Consistently, we also determined that the expression of several pluripotent markers (OCT4, SOX2, etc) is severely downregulated as early as 7 days after initiation of differentiation using DM (data not shown but see Garcia-Perez *et al*, 2010).

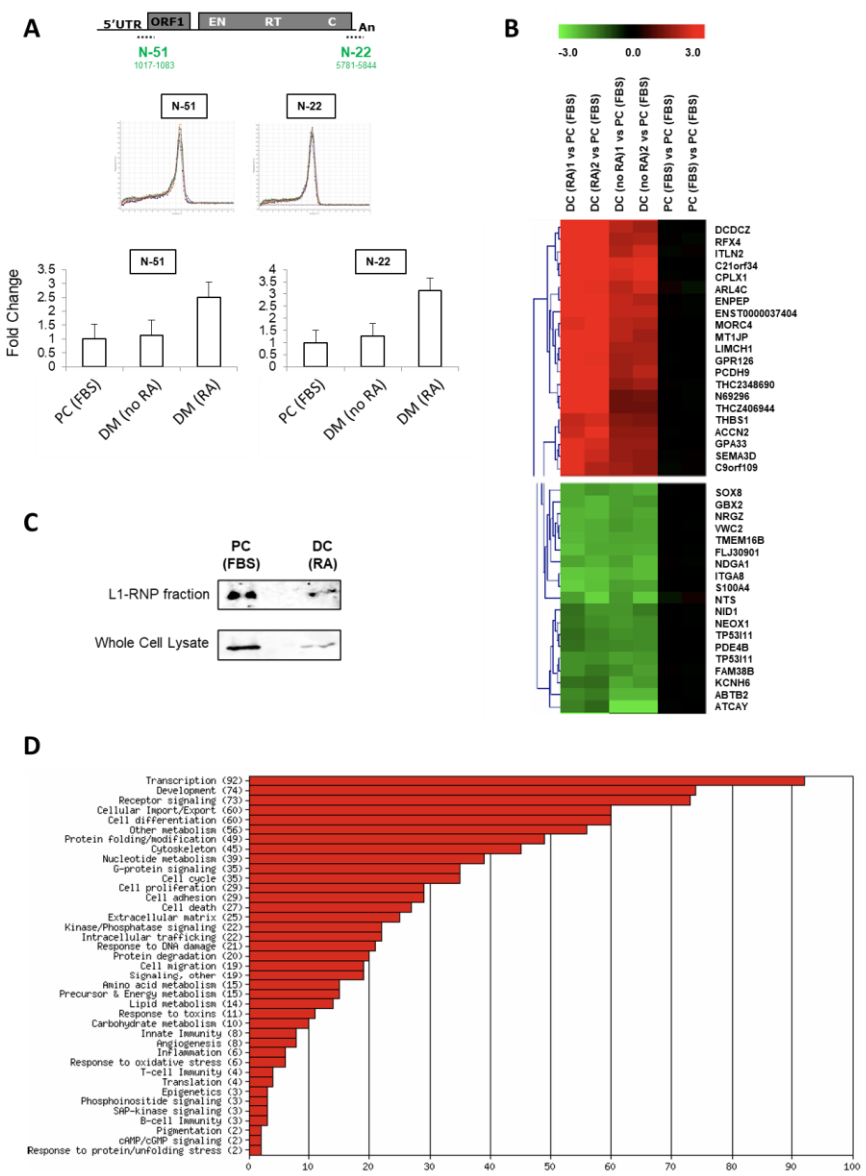


Figure 48. PA-1 differentiation analyses. *PA-1 cells were differentiated during 14 days, using the conditions described in Garcia-Perez et al, 2010. Two differentiation conditions were employed: DC (RA), containing 10 % of Knockout Serum Replacement (KOSR) and 1 μM of Retinoic Acid (RA, all trans) (and no fetal bovine serum (FBS)); and DC (no RA), containing 20 % of KOSR, no Retinoic Acid (RA) and no FBS. PA-1 cells cultured with standard PA-1 media (see MATERIALS AND METHODS) were cultured with 10% FBS as a pluripotent control (PC (FBS)). A) In the upper section of the figure, a schematic of the LINE-1 sequence indicating the location of the regions amplified to identify the L1-RNA are shown. Two regions located in the 5'UTR and L1-ORF2p of the L1 sequence were amplified, termed primer sets N-51 and N-22, respectively. The lower graphs*

show the expression levels detected for PC (FBS), DC (non-RA) and DC (RA). B) Results from microarray expression analyses. The heat map shows the genes whose expression change the most when comparing PCs and DCs, either up or downregulated. Each line is a gene, and the columns from left to right show the following comparisons: columns 1 and 2, DM (RA) vs PC (FBS) (duplicates); columns 3 and 4, DM (no RA) vs PC (FBS); and columns 5 and 6, PC (FBS) vs PC (FBS) as a control, with no significant changes detected. C) L1-RNPs were isolated by sucrose cushion (Kulpa and Moran, 2006) and L1-ORF1p expression was analyzed by Western blot using β-actin as a loading control. D) GO analysis shows the top downregulated pathways upon differentiation (pathways downregulated in DC conditions).

Consistently, using whole genome transcriptomic analyses, we confirmed that PA-1 cells are fully differentiated 14 days after initiation of differentiation using DM (mostly towards ectodermal linages, see **Figure 48B**); similarly, Gene Ontology (GO) analyses revealed that relevant pathways such as *Development* and *Cell Differentiation*, were significantly upregulated in DC PA-1 (**Figure 48D**). Remarkably, we found that expression of endogenous L1 RNAs (**Figure 48A**) and L1-RNPs (**Figure 48C**) is similar between PC and DC. Consistent with canonical embryonic differentiation, we found that the methylation status of the promoter of endogenous L1Hs elements increase as PA-1 cells differentiate using DM (**Figure 49**); however, the increase in DNA methylation observed was modest, and even 14 days after the initiation of cellular differentiation using DM, we could detect fully unmethylated L1Hs promoters (**Figure 49**). While these data might look paradoxical, it´s worth noting that we analyzed very early differentiation times, and we speculate that as cells continue to differentiate, it is likely that the extent of L1-promoter methylation would increase to levels that would result in reduced RNA expression levels. However, the endogenous expression of L1-RNPs in DCs allowed us to characterize the L1-ORF1p interactome of isogenic DCs, using the same proteomic approach and the optimized protocols established using PC PA-1s.

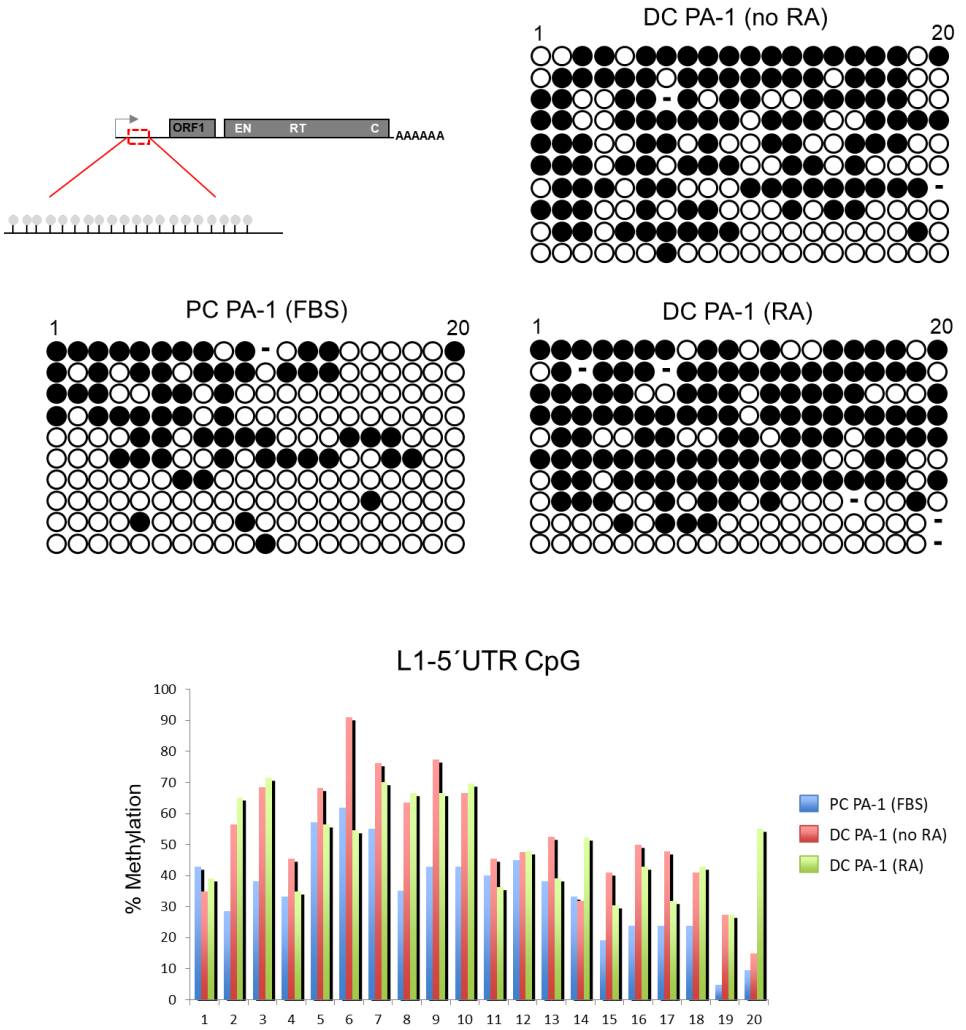


Figure 49. PA-1 differentiation analyses. *We determined the methylation levels of the CpG islands located in the 5'UTR promoter of L1. Note how methylation increases with differentiation. Upper panel, schematic of the L1 sequence indicating the location of the 20 CpG motifs found in the L1 promoter. The methylation status of 10 sequences showing the highest homology with human RC-L1s are plotted in the different conditions tested. Black and white circles denote methylated and unmethylated CpG residues in the L1 promoter, respectively. Bellow, the graph shows the % of methylation in each of the 20 CpG motifs found in the L1 promoter and in each condition (PC PA-1 (FBS), DC PA-1 (non-RA), DC PA-1 (RA)).*

Thus, we next conducted proteomic analyses of DC PA-1 cells, using anti-human_L1-ORF1p antibody coupled magnetic dynabeads (mαORF1p) and the protocol used when exploring the L1-interactome of PC PA-1. However, we first determined the scale of PA-1 DC (amount of mg per replicate) necessary to obtain a L1-ORF1p yield equivalent to that obtained in PC PA-1. These analyses revealed that the same amount of material used for PA-1 PCs gave a similar yield when using DC PA-1s, consistent with PCs and DCs expressing similar levels of L1 RNAs and L1-RNPs (**Figure 48A and C**). To replicate the experiment under the same experimental conditions used with PCs, 4 grams of DC PA-1s (500 mg per capture) were required; however, likely because differentiated PA-1 cells are larger in size than pluripotent cells, we only had 3 grams of DC PA-1 available to conduct all the proteomic analyses planned. In fact, while culturing PC and DC PA-1 in 500 cm^2 tissue culture plates, we noticed that to obtain the same amount of DC PA-1s, we used three times the materials used when culturing PC PA-1s. To obtain 1g of PC PA-1s, we routinely used five 500 cm^2 plates; in contrast, to obtain 1g of DC PA-1s, we required around ten 500 cm^2 plates. Thus, to characterize the interactome of L1-ORF1p on DC PA-1s, we performed 4 biological replicates for the Case (mαORF1p) and only 2 for the Control (mIgG), each using 500 mg of PA-1 DC cell powder (**Figure 50**). To analyze the L1-interactome of DC PA-1s, we used the same experimental conditions used with PC PA-1s, except for elution [to elute, we used the conditions most effective to release L1-ORF1p from beads, see **Figure 45**; that is 2% SDS 40 mM Tris pH 8 at 70 ºC for 5 min (**Figure 50**)]. While using slightly different elution conditions could create some caveats to compare the L1-interactome of PCs and DCs, it is worth noting that all the elution conditions tested were similar in their capability to extract L1-ORF1p (and interactors) from beads (**Figure 45**).

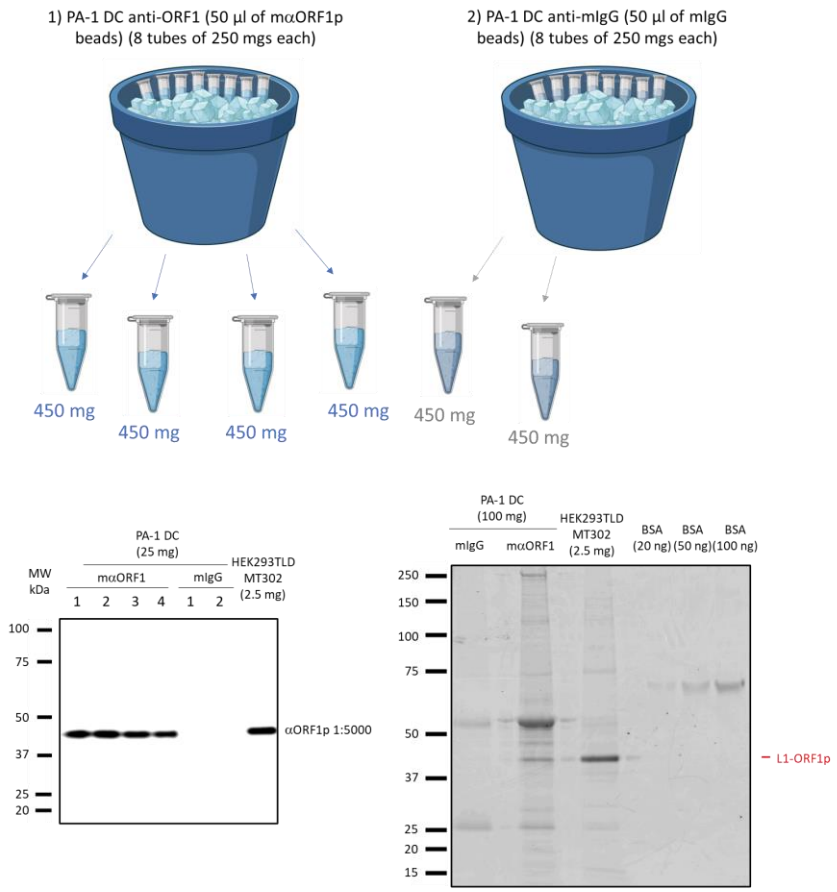


Figure 50. Capturing L1-RNPs from DC PA-1 DC using mouse anti-human_L1-ORF1p beads (mαORF1p). *Schematic of the experiment. Below are shown results from Western blot (left side) and 4-12% Bis-Tris Sypro stained gel electrophoresis (right side) analyses of the elutions (amount of material loaded is indicated in the figure). Mouse IgG coupled beads were used in parallel as a negative control. Protocol described in detail in VI. MATERIALS AND METHODS.*

Upon completion of the capture protocol, we next used the equivalent of 50 mg of elution from each replicate in order to analyze the capture of L1-ORF1p/interactors on DC PA-1s by Western blot (25 mg from each replicate) and by Sypro-stained gel electrophoresis (25 mg of each replicate, combined in 100 mg). However, on this occasion we only used the equivalent of 100 mg of cell powder (instead of 200 mg) to load the Sypro-stained gel. Notably, we successfully released L1-ORF1p and interactors from beads on DC PA-1 captures, as revealed by Sypro-stained gel electrophoresis (**Figure 50**, right panel). Thus, the remaining elution, ~450 mg, were processed by S-Trap and subjected to LC-MS/MS, using the conditions described above. Upon bioinformatic analyses, we establish a list of L1-ORF1p specific interactors, using IgG as a negative control. Remarkably, while we

noticed significant differences in the composition of the L1-interactome when compared to PCs (either comparing to the mouse monoclonal dataset or comparing to the nanobody dataset), we also identified several proteins that were detected in previous proteomic L1 studies (Taylor *et al*, 2013; 2018) (highlighted in yellow in **Figure 51**). In fact, we noticed that 10 of 81 (12.3%) DC L1-interactors were found in previous proteomic analyses using ectopically expressed L1s and transformed cell lines (Goodier *et al*, 2013; Taylor *et al*, 2013; Moldovan & Moran, 2015; Taylor *et al*, 2018). Similarly, we also found that 18 of 81 (22.2%) and 9 of 81 (11.1%) L1-interactors were also found in the L1-interactome of isogenic PC PA-1s, analyzed with the mouse monoclonal and the llama nanobody, respectively (**Table 6**). A detailed comparison of DC- and PC-PA-1 L1-interactomes will be described in the following pages. However, it is worth noting that our analyses using isogenic models revealed that the L1-interactome is highly dynamic during cellular differentiation, and we speculate that L1 regulation would be different when PCs and DCs are compared. Indeed, L1-silencing is a perfect example of how the ontogenic status of cells (i.e., their pluripotency) influence how cells react to new L1 retrotransposition events.

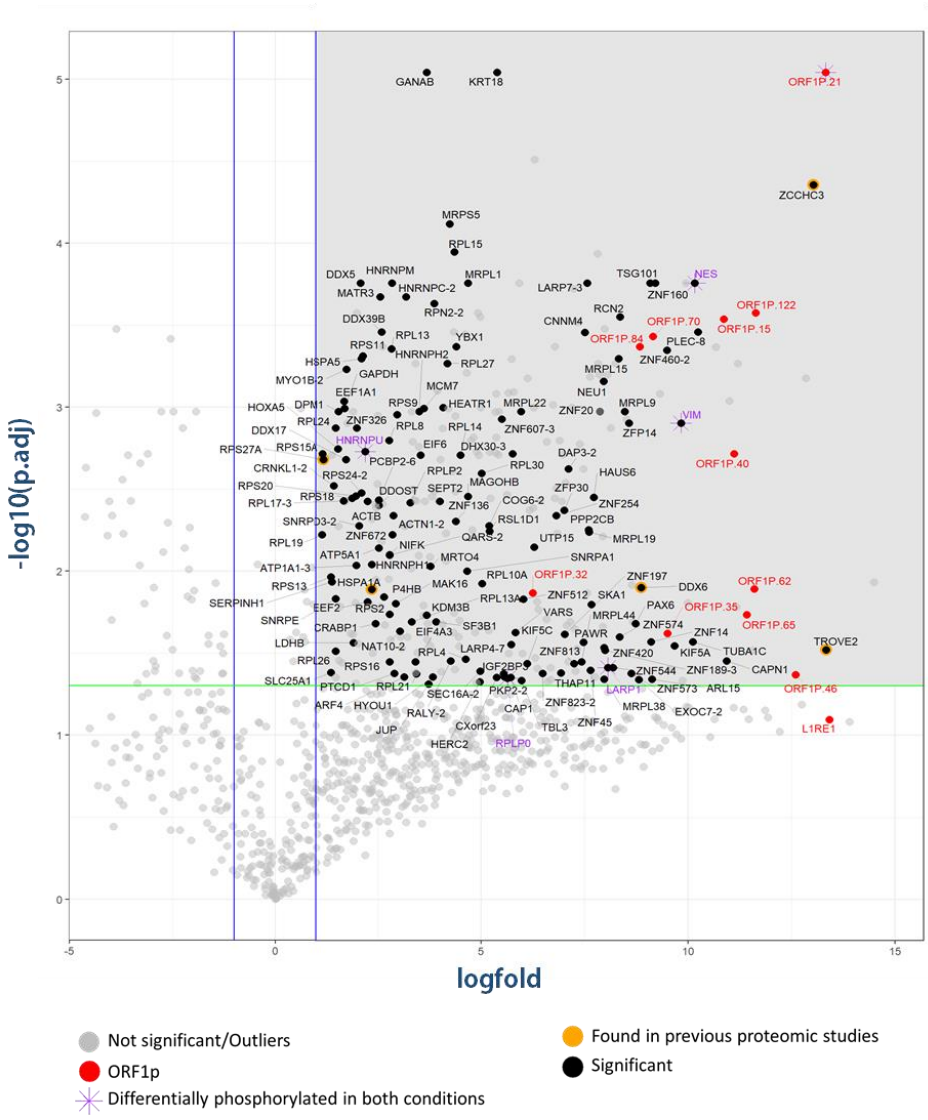


Figure 51. L1-ORF1p specific interactors found in DC PA-1 cells, detected with the mouse monoclonal anti-human_L1-ORF1p antibody (mαORF1p). *The volcano plot indicates DC specific PA-1 interactors (see Method 2 (VI. MATERIALS AND METHODS, 22. Mass spectrometry data analysis). In the plot, the name of the main L1-ORF1p interacting proteins in differentiated PA-1 cells (DC) is indicated. Highlighted in the upper right quadrant are significant DC L1-ORF1p interacting factors [-log[10] of p-adjusted value < 0.05 (p-adjusted value from Benjamini Hochberg method applied to t-test p-value); and logfold > 1 (log2fold change of LFQ intensity between cases and controls)]. Highlighted in purple are factors with different phosphorylation, L1-ORF1p is highlighted in red, and highlighted in yellow are known L1-interactors (i.e., factors found in previous studies, Taylor et al, 2013; 2018).*

Table 6. Factors identified in DC PA-1 using mαORF1p beads in common with previous studies.

	Gene names	(Goodier *et al.*, 2013)	(Taylor *et al.*, 2013)	(Moldovan and Moran, 2015)	(Taylor *et al.*, 2018)	(Vuong *et al.*, 2019)
1	PURA	Yes*	Yes*	Yes*	Yes*	Yes
2	TROVE2	Yes*	Yes*	Yes*	Yes*	-
3	RPS27A	-	Yes*	-	Yes*	-
4	ZCCHC3	-	Yes*	-	Yes*	-
5	HSPA1A	-	-	-	Yes*	-
6	MATR3	Yes*	-	Yes*	-	-
7	DDX17	Yes*	Yes	-	-	-
8	DDX5	Yes*	Yes	-	-	-
9	HNRNPU	Yes*	Yes	-	-	-
10	YBX1	Yes*	Yes	-	-	-
	Factors found	7	4	3	5	0

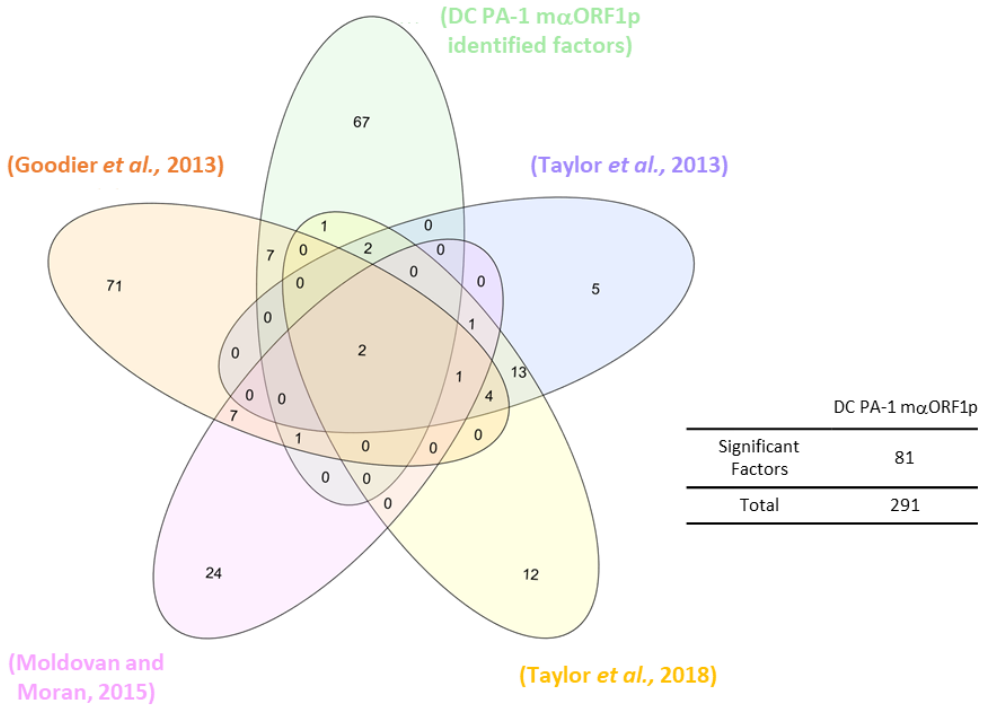


Table 6. Factors identified in DC PA-1 using mαORF1p beads that have been found in previous studies. *Asterisks indicate that the factor was significant in that condition. "Yes" indicates that the protein was identified in the study, but did not reach the significance threshold established by authors.*

2.4. Comparison of L1-ORF1p interactomes from isogenic PCs and DCs.

As described above, beside studying how dynamic the L1-interactome is (or not), and how this might influence L1 regulation, a main goal of my Thesis was to explore whether advanced proteomic approaches could be used to dissect the mechanism of L1-silencing. Thus, we next analyzed in detail the composition of the L1-ORF1p interactome from isogenic PCs and DCs, focusing on host factors that are present in PC but absent in DCs, as we previously demonstrated that L1-silencing is attenuated in DCs.

However, we started these analyses by first comparing the interactors identified in PCs; specifically, we compared L1-ORF1p interactors identified using mαORF1p and LαORF1p to capture L1-RNPs. To compare these groups, we generated a scatter plot that combined both interactomes (**Figure 52**). As briefly described above, we noticed that a decent number of interactors were successfully co-immunoprecipitated by both antibodies [21 out of 134 for mαORF1p (15.7%) and 21 out of 65 for LαORF1p (32.3%) (**Table 7**). However, we also found a relatively large number of unique interactors (from a total of 178 significant interactors identified employing both antibodies, 63.5% were unique to mαORF1p, and 24.7% were unique to LαORF1p). Among them, it is worth noting that a large number of zinc finger proteins (ZNFs) were found in the mαORF1p interactome but were absent in the LαORF1p interactome. In fact, ZNFs were identified as PC-specific L1-ORF1p interactors in the mαORF1p interactome (33 factors, corresponding to 24.6% of the factors identified from mαORF1p) and in the LαORF1p interactome (3 factors, corresponding to 4.6% of factors identified from LαORF1p). Intriguingly, we found a large enrichment of ZNFs in the mαORF1p interactome (11-fold higher in the mαORF1p interactome); notably, these ZNFs were either unique to the mαORF1p dataset (30 out 33 ZNFs were only found in this dataset, corresponding to a 90.9% of all factors), or were share with the LαORF1p interactome (3 ZNFs were also found in LαORF1p. Thus, we speculate that each antibody might capture L1-ORF1p and L1-RNPs in specific cellular contexts (*VIII. DISCUSSION*).

Table 7. PA-1 PC common identified factors using mαORF1 and LαORF1 beads.

	Gene names	(Goodier *et al.*, 2013)	(Taylor *et al.*, 2013)	(Moldovan and Moran, 2015)	(Taylor *et al.*, 2018)	(Vuong *et al.*, 2019)
1	ZNF326	-	Yes	-	-	-
2	ZCCHC3	-	Yes*	-	Yes*	-
3	DHX30	-	-	-	-	-
4	NES	-	-	-	-	-
5	VIM	-	Yes	-	-	-
6	AKAP8	-	Yes	-	-	-
7	CTSA	-	-	-	-	-
8	MOV10	Yes*	Yes*	Yes*	Yes*	-
9	DDX28	-	Yes	-	-	-
10	EXOC7	-	-	-	-	-
11	PABPC1	Yes*	Yes*	-	Yes*	-
12	TUBB6	-	-	-	-	-
13	MRPL21	-	Yes	-	-	-
14	DDX3X	-	Yes	-	-	-
15	TUBB4B	-	Yes	-	Yes*	-
16	CTNNB1	-	-	-	-	-
17	DNAJA3	-	-	-	-	-
18	EPPK1	-	-	-	-	-
19	TRIOBP	-	-	-	-	-
20	ZNF483	-	-	-	-	-
21	ZNF219	-	-	-	-	-

	Significant interactors	Percentage over total (%)
PA-1 PC mαORF1 unique interactors	113	63.5
PA-1 PC LαORF1 unique interactors	44	24.7
PA-1 PC LαORF1 and PC mαORF1 common interactors	21	11.8
Total number of significant interactors identified	178	

	ZNF identified (significant)
PA-1 PC mαORF1 unique interactors	30
PA-1 PC LαORF1 unique interactors	0
PA-1 PC LαORF1 and PC mαORF1 common interactors	3
Total number of significant interactors identified	33

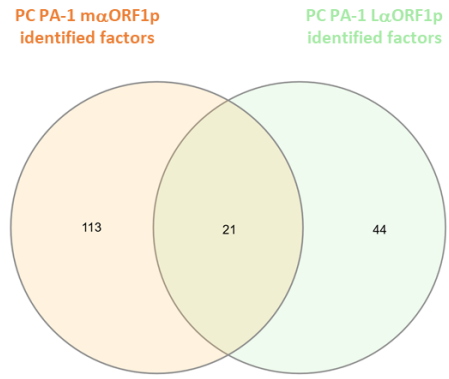


Table 7. PA-1 PC L1-interactors identified using mαORF1 and LαORF1 beads. *Asterisks indicate that the factor was significant in that condition. "Yes" indicates that the protein was identified in the study, but did not reach the significance threshold established by authors.*

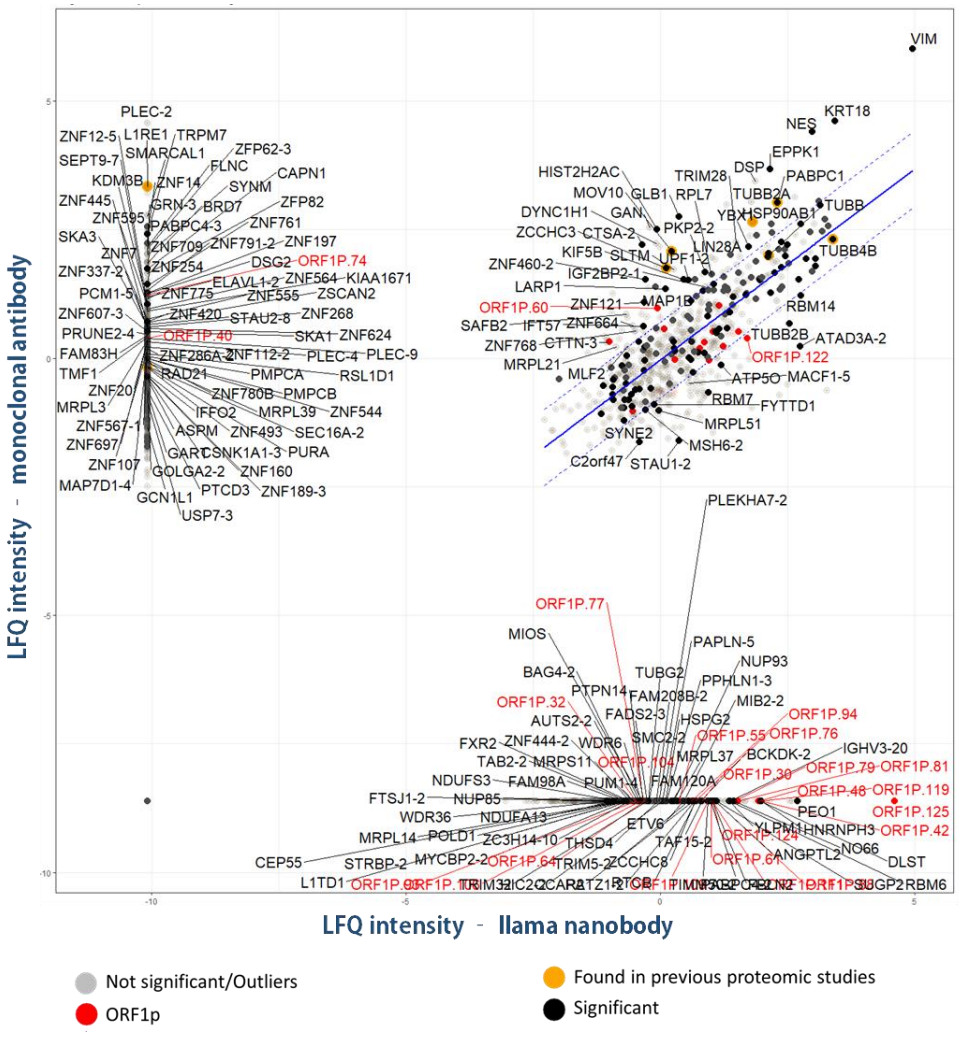
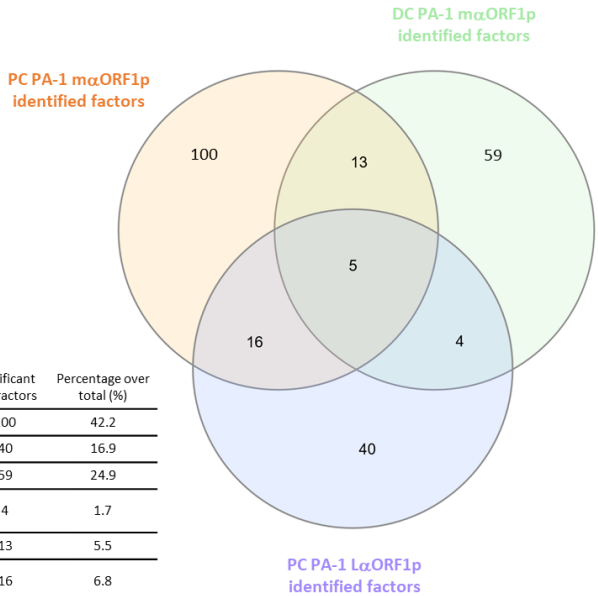


Figure 52. Scatter plot of PC interactors detected with monoclonal antibody (mαORF1p) vs nanobody (LαORF1p). *Proteins are represented according to their Avg. LFQ intensity (VI. MATERIALS AND METHODS, 22. Mass spectrometry data analysis). Upper right quadrant, proteins which showed statistical significance in both mαORF1p and LαORF1p; left quadrant, proteins unique and significant in the interactome captured with mαORF1p; bottom right quadrant, proteins unique and significant in the interactome captured with LαORF1p. Highlighted in black are significant factors (t-test); red, human_L1-ORF1p; yellow, factors found in previous studies.*

Following the same rationale, we next compared the L1-ORF1p interactome of isogenic DCs and PCs, both captured with mαORF1p (**Figure 53**). Notably, we found a relatively small group of specific L1-interactors that were shared between PCs and DCs (ZCCHC3, ZNF326, VIM, NES and DHX30) (5 factors representing 3.7%, 7.7% and 6.1% of factors identified as significant in PC-mαORF1p, PC-LαORF1p, and DC-mαORF1p, respectively). Among these, one factor (ZCCHC3) was found in previous proteomic studies using ectopically expressed L1s and transformed cell lines (Taylor et al, 2013; Taylor *et al*, 2018*)*. Additionally, 9 specific L1-interactors were found in PC-mαORF1p and DC-mαORF1p (representing 6.7% and 11.1% of factors identified as significant in PC-mαORF1p and DC-mαORF1p, respectively) from which one of them (YBX1) had been previously identified in Goodier *et al*, 2013 and Taylor *et al*, 2013. Similarly, 4 factors were common to PC-LαORF1p and DC-mαORF1p (**Table 8**); these include *bona fide* L1-RNP interactors such as MOV10, and PABPC1 and the other 2 were also found in previous studies (MATR3, Goodier *et al*, 2013, Taylor *et al*, 2013; Moldovan & Moran, 2015; Taylor *et al*, 2018; and DDX5, Goodier *et al*, 2013, Taylor *et al*, 2013; Moldovan & Moran, 2015; Taylor *et al*, 2018*)*. Interestingly, we also found a relatively large group of specific L1-interactors that were unique to PCs or DCs [(100 (74.6%), 40 (61.5%), and 59 (72.8%) factors identified in PC-mαORF1p, PC- LαORF1p, and DC-mαORF1p, respectively, were unique to each dataset)]. Remarkably, many of these factors are unique to our study, and have never been identified in previous studies.

Table 8. PA-1 PC and DC common identified factors.

	Gene names	DC mαORF1	PC mαORF1	PC LαORF1	(Goodier *et al.*, 2013)	(Taylor *et al.*, 2013)	(Moldovan and Moran, 2015)	(Taylor *et al.*, 2018)
1	ZCCHC3	Yes*	Yes*	Yes*	-	Yes*	-	Yes*
2	ZNF326	Yes*	Yes*	Yes*	-	Yes*	-	-
3	VIM	Yes*	Yes*	Yes*	-	Yes*	-	-
4	NES	Yes*	Yes*	Yes*	-	-	-	-
5	DHX30	Yes*	Yes*	Yes*	-	-	-	-
6	PLEC	Yes*	Yes*	No	-	-	-	-
7	MRPL44	Yes*	Yes*	No	-	Yes*	-	-
8	YBX1	Yes*	Yes*	No	Yes*	Yes*	-	-
9	RPS9	Yes*	Yes*	No	-	Yes*	-	-
10	RPS2	Yes*	Yes*	No	-	Yes*	-	-
11	RPS18	Yes*	Yes*	No	-	Yes*	-	-
12	RPL4	Yes*	Yes*	No	-	Yes*	-	-
13	MRPL47	Yes*	Yes*	No	-	Yes*	-	-
14	MRPL22	Yes*	Yes*	No	-	Yes*	-	-
15	GAN	Yes*	Yes*	No	-	-	-	-
16	SEC16A	Yes*	Yes*	No	-	-	-	-
17	ZNF189	Yes*	Yes*	No	-	-	-	-
18	RPS13	Yes*	No	Yes*	-	Yes*	-	-
19	MATR3	Yes*	No	Yes*	Yes*	Yes*	Yes*	-
20	KRT18	Yes*	No	Yes*	-	-	-	-
21	DDX5	Yes*	No	Yes*	Yes*	Yes*	-	-



	Significant interactors	Percentage over total (%)
PA-1 PC mαORF1 unique interactors	100	42.2
PA-1 PC LαORF1 unique interactors	40	16.9
DC mαORF1 unique interactors	59	24.9
PA-1 DC mαORF1 and PC LαORF1 common interactors	4	1.7
PA-1 DC and PC mαORF1 common interactors	13	5.5
PA-1 PC LαORF1 and PC mαORF1 common interactors	16	6.8
PA-1 PC mαORF1, PC LαORF1 and DC mαORF1 common interactors	5	2.1
Total number of significant interactors identified	237	

Table 8. L1-interactors found in PA-1 pluripotent (PC) and differentiated (DC) cells. *"Yes" indicates that the protein was identified in the study and asterisks indicate that the factor was significant in that condition. "No" means that the protein was not identified in that condition.*

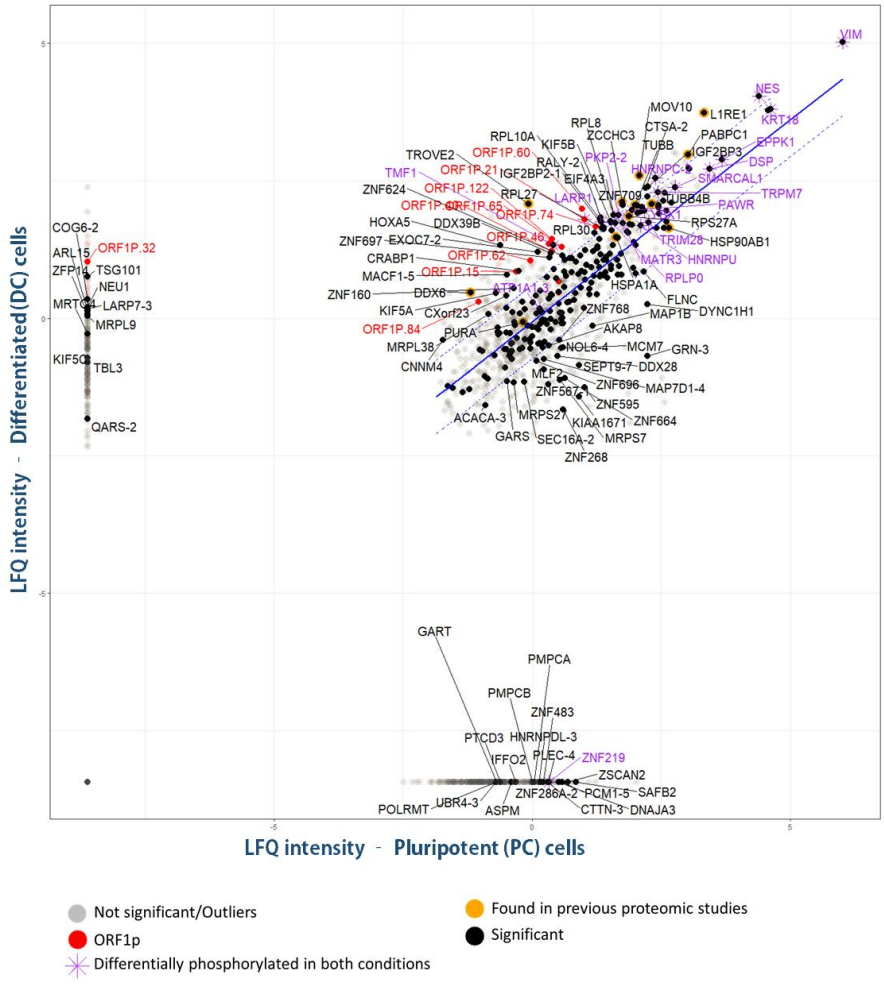


Figure 53. L1-ORF1p interactors detected in PC and DC samples using the mouse anti-ORF1p monoclonal antibody (mαORF1p). *Proteins are represented according to their Avg. LFQ intensity (VI. MATERIALS AND METHODS, 22. Mass spectrometry data analysis). Upper right quadrant, proteins which showed statistical significance in both PA-1 PC and PA-1 DC; left quadrant, proteins unique and significant in the PA-1 PC interactome; bottom right quadrant, proteins unique and significant in the PA-1 DC interactome. Highlighted in black are significant factors (t-test); red, human_L1-ORF1p; yellow, factors found in previous studies; purple, factors differentially phosphorylated.*

2.5. A list of PC-specific L1-ORF1p interactors potentially associated with L1-regulation in PCs.

Our working hypothesis to dissect the mechanism of L1-silencing indicates that the host factor/s involved would be pluripotent-specific and could recognize TPRT intermediates, specifically the Y-like structure (RNA:cDNA heteroduplex attached to genomic DNA) generated during the initial steps of L1 retrotransposition. Furthermore, we suggest that these host factors could first interact with L1 during translation and/or during the assembly of L1-RNPs. Here, we aim to use our proteomic datasets to identify factors interacting with L1-ORF1p that could be involved in L1-silencing or in L1-regulation in PCs, analyzing host factors identified as unique to PC PA-1s and using different criteria to prioritize possible candidates that would be subsequently analyzed in the context of L1-silencing and retrotransposition.

Briefly, to determine whether a host factor is significantly enriched in PC PA-1s or in DC PA-1s *(see VI. MATERIALS AND METHODS, 22. Mass spectrometry data analysis),* data from both PC interactomes [PC PA-1 (mαORF1p) and PC PA-1 (LαORF1p)] were compared with the DC PA-1 (mαORF1p) interactome, generating a list of factors mostly present in PC PA-1 cells. On a complementary approach, we applied the same analysis to identify factors showing differential phosphorylation between PA-1 PCs and DCs. While other post-translational modifications (PTM) could be analyzed, we restricted our analyses to phosphorylation, as we could robustly explored this PTM in our proteomic datasets. Thus, while it is possible that other PTMs could influence L1-silencing, the proteomic approach used in this Thesis prohibited us from exploring additional PTMs with robustness and reliability.

Next, we determined whether any of our factors has been previously identified in other LINE-1 interactome studies. To do that, we used R Studio scripts and libraries such as dplyr (Hadley *et al,* 2018*)* and its merge and join functions, to compare factors based on their uniprotID in the list obtained and the lists extracted from previous proteomic studies (mentioned in *III. INTRODUCTION*). However, this approach yielded hardly any coincidence.

Additional filtering criteria were also explored; we performed a search for potential matches between the proteins in our list and the interactome of RNA/DNA hybrids, generated in a study by Cristini *et al,* (2018). Notably, some of our PC-specific factors (SAFB2, Class I; HDAC2, HNRNPU, YBX1, LARP1, Class II; DNAJA3, RBM39, HSP90AB1, Class III) were found to be among those described as first, second or third-order interactors of RNA/DNA hybrids.

In sum, the use of these filtering criteria was used to generate a first list of host factors specific to PCs or differentially expressed/phosphorylated among PCs and DCs (a list with 30 host factors) (**Table 9**).

Furthermore, since our lab generated RNAseq datasets from PC and DC PA-1s, we also analyzed expression levels of our candidates. These analyses revealed that 6 out of 24 factors (SAFB2, ZNF219, ZNF483, LIN28A, RBM39, TRIM28) were expressed at a higher level in PCs, and that 3 of 24 factors (MCM7, SPTBN1, KRT18) had higher expression levels in DCs.

Table 9. List of 30 host factors of interest.

#	uniprotID	Gene names	logfold	Significant in PC vs DC comparison	PA1PC maORF1p Significant	PA1PC LaORF1p Significant	PA1DC maORF1p Significant	Differentially expressed phosphosite	Found in previous studies	Interacts with RNA/DNA hybrids	RNAseq	Subcellular location [CC]
1	Q7Z7L9	ZSCAN2	-8.674	Yes	Yes	NA	NA	-	Mita et al, 2020	-	Similar expression	Nucleus
2	Q96EY1	DNAJA3	-8.330	No	Yes	Yes	No	-	-	Class III interactor	-	Mitochondrion matrix. Cytoplasm, cytosol, Cell junction, synapse, postsynaptic cell membrane, Peripheral membrane protein. Note=Recruited to the postsynaptic cell membrane of the neuromuscular junction through interaction with MUSK.
3	P28799-3	GRN	-8.090	Yes	Yes	NA	No	-	-	-	Similar expression	Secreted, Lysosome. Note: Endocytosed by SORT1 and delivered to lysosomes. Targeted to lysosome by PSAP via M6PR and LRP1, in both biosynthetic and endocytic pathways.
4	Q14151	SAFB2	-8.037	Yes	Yes	No	NA	-	Vuong et al. 2019 Mita et al, 2020	Class I interactor	Higher expression in PC (Significant)	Cytoplasm. Nucleus.
5	Q9P2Y4	ZNF219	-7.345	Yes	Yes	Yes	NA	Yes (Phosphorylated in PC)	Mita et al, 2020	-	Higher expression in PC (Significant)	Nucleus.
6	Q8TF39	ZNF483	-7.048	Yes	Yes	Yes	NA	-	-	-	Higher expression in PC (Significant)	Nucleus
7	O75439	PMPCB	-6.486	Yes	Yes	NA	NA	-	-	-	Similar expression	Mitochondrion matrix.
8	P42704	LRPPRC	-6.093	Yes	NA	No	No	Yes (Phosphorylated in DC)			Similar expression	
9	Q14587	ZNF268	-5.654	Yes	Yes	NA	No	-	Vuong et al. 2019	-	Similar expression	[Isoform 1]: Nucleus /[Isoform 2]: Nucleus. Cytoplasm.
10	Q8IZT6	ASPM	-5.072	Yes	Yes	NA	NA	-	-	-	Similar expression	Nucleus, cytoplasm, spindle (cytoskeleton).
11	Q8IYB9	ZNF595	-4.809	Yes	Yes	NA	No	-	-	-	Similar expression	Nucleus
12	Q5T4S7-3	UBR4	-4.627	Yes	Yes	No	NA	-	-	-	Similar expression	Membrane; Multi-pass membrane protein. Cytoplasm. Cytoplasm, cytoskeleton. Nucleus. Note=Concentrates at the leading edge of membrane structures involved in actin motility.

Table 9. List of 30 host factors of interest.

#	uniprotID	Gene names	logfold	Significant in PC vs DC comparison	PA1PC maORF1p Significant	PA1PC LaORF1p Significant	PA1DC maORF1p Significant	Differentially expressed phosphosite	Found in previous studies	Interacts with RNA/DNA hybrids	RNAseq	Subcellular location [CC]
13	Q8N184-1	ZNF567	-4.203	Yes	Yes	NA	No	-	-	-	Similar expression	Nucleus.
14	Q8IXZ2	ZC3H3	-3.522	Yes	No	NA	No	Yes (Phosphorylated in DC)	Mita et al, 2020	-	Similar expression	Nucleus.
15	Q92769-3	HDAC2	-3.362	No	No	No	No	Yes (Phosphorylated in PC)	-	Class II Interactor	Similar expression	Nucleus, cytoplasm.
16	P06748	NPM1	-2.958	Yes	No	No	No	-	Goodier et al, 2013	-	-	Nucleus, cytoskeleton.
17	Q01082	SPTBN1	-2.828	Yes	No	No	No	-	-	-	Higher expression in DC (Significant)	Cytoskeleton.
18	Q2Q1W2	TRIM71	-2.686	Yes	No	Yes	No	-	-	-	-	P-body.
19	P33993	MCM7	-2.283	Yes	No	Yes	No	-	Liu et al, 2017	-	Higher expression in DC (Significant)	Nucleus.
20	Q9H9Z2	LIN28A	-1.853	Yes	Yes	No	No	-	-	-	Higher expression in PC (Significant)	Cytoplasm, rough endoplasmic reticulum, P-body, Stress granule, nucleolus.
21	Q14498-2	RBM39	-1.61	Yes	No	NA	No	Yes (Phosphorylated in PC)	-	Class III interactor	Higher expression in PC (Significant)	Nucleus speckle.
22	P08238	HSP90AB1	-1.61	Yes	Yes	No	No		Taylor et al, 2013; Taylor et al, 2018	Class III interactor	Similar expression	Cytoplasm, Melanosome, Nucleus, Secreted, Cell membrane, Dynein axonemal particle.
23	P17014-5	ZNF12	-1.33	Yes	Yes	NA	No	-	-	-	Similar expression	Nucleus.
24	P05783	KRT18	-0.672	No	Yes	No	Yes	Yes	-	-	Higher expression in DC (Significant)	Perinuclear region, nucleolus.

Table 9. List of 30 host factors of interest.

#	uniprotID	Gene names	logfold	Significant in PC vs DC comparison	PA1PC maORF1p Significant	PA1PC LaORF1p Significant	PA1DC maORF1p Significant	Differentially expressed phosphosite	Found in previous studies	Interacts with RNA/DNA hybrids	RNAseq	Subcellular location [CC]
25	Q13263	TRIM28	-0.184	No	Yes	No	No	Yes	Liu et al, 2017	Mita et al, 2020	Higher expression in PC (Significant)	Nucleus.
26	Q00839	HNRNPU	0.086	No	No	No	Yes	Yes	Goodier et al, 2013; Taylor et al, 2013; Taylor et al, 2018	Class II Interactor	Similar expression	Nucleus, Nucleus matrix, Chromosome, Nucleus speckle, centrosome, kinetochore, spindle, spindle pore, Midbody, Cytoplasm, Cell surface, Cytoplasmic granule.
27	Q9NZC9	SMARCAL1	0.322	No	Yes	NA	No	Yes (Phosphorylated in PC)	-	-	Similar expression	Nucleus
28	P67809	YBX1	0.700	No	Yes	No	Yes	Yes (Both in PC and DC)	Taylor et al, 2013, Goodier et al, 2013	Class II Interactor	Similar expression	Cytoplasm. Nucleus. Cytoplasmic granule. Secreted. Secreted, extracellular exosome. Cytotoxic stress and DNA damage enhance translocation to the nucleus. Localized in cytoplasmic mRNP granules containing untranslated mRNAs. Secreted by mesangial and monocytic cells after inflammatory challenges.
29	Q3V6T2-4	CCDC88A	1.706	No	No	NA	No	Yes (Phosphorylated in DC)	-	-	-	Cell membrane, cytosol, Cytoplasmic vesicle, lamellipodium, cilium basal, body, centriole.
30	Q6PKG0	LARP1	2.479	No	Yes	No	No	Yes (Phosphorylated in PC)	Goodier et al, 2013; Taylor et al, 2013; Moldovan & Moran, 2015; Taylor et al, 2018 Mita et al, 2020	Class II Interactor	-	Cytoplasm, Cytoplasmic granule.

PA1PC mαORF1p Significant, Significant in mαORF1p versus mIgG replicates comparison; PA1PC LαORF1p Significant, Significant in LαORF1p versus mIgG replicates comparison; PA1DC maORF1p Significant, Significant in mαORF1p versus mIgG replicates comparison. Classes were stablished in Agnese *et al, (2018)* according to their statistical significance (Class I, top 25%; Class II, next 50%; and Class III, bottom 25%).

Next, I extracted any information available for these 30 factors, using Pubmed, GeneCards and UniProt databases. Using this information, we annotated known functions (when available), their cellular localization, and the context in which they are expressed. These analyses allowed us to rank our factors according to the number of evidences supporting a possible role in relation to the mechanism of L1-silencing. As a result, we selected 14 factors as top L1-silencing candidates, which are summarized in the following table (**Table 10**).

Table 10. List of potential PA-1 PC interactors involved in the de novo L1 silencing mechanism												
#	uniprotID	Gene names	logfold	Significant in PC vs DC comparison	PA1PC mαORF1p Significant	PA1PC LαORF1p Significant	PA1DC mαORF1p Significant	Differentially expressed phosphosite	Found in previous studies	Interacts with RNA/DNA hybrids	RNAseq	Subcellular location [CC]
1	Q7Z7L9	ZSCAN2	-8.674	Yes	Yes	NA	NA	-	Mita et al, 2020	-	Higher expression in PC (Significant)	Nucleus
2	Q96EY1	DNAJA3	-8.33	No	Yes	Yes	No	-	-	Class III interactor	-	Mitochondrion matrix. Cytoplasm, cytosol, Cell junction, synapse, postsynaptic cell membrane, Peripheral membrane protein. Note=Recruited to the postsynaptic cell membrane of the neuromuscular junction through interaction with MUSK.
3	P28799-3	GRN	-8.09	Yes	Yes	NA	No	-	-	-	Similar expression	Secreted, Lysosome. Note: Endocytosed by SORT1 and delivered to lysosomes. Targeted to lysosome by PSAP via M6PR and LRP1, in both biosynthetic and endocytic pathways.
4	Q14151	SAFB2	-8.037	Yes	Yes	No	NA	-	Vuong et al. 2019 Mita et al, 2020	Class I interactor	Higher expression in PC (Significant)	Cytoplasm. Nucleus.
5	Q9P2Y4	ZNF219	-7.345	Yes	Yes	Yes	NA	Yes (Phosphorylated in PC)	Mita et al, 2020	-	Higher expression in PC (Significant)	Nucleus
6	Q8TF39	ZNF483	-7.048	Yes	Yes	Yes	NA	-	-	-	Higher expression in PC (Significant)	Nucleus
7	Q14587	ZNF268	-5.654	Yes	Yes	NA	No	-	Vuong et al. 2019	-	Similar expression	[Isoform 1]: Nucleus / [Isoform 2]: Nucleus. Cytoplasm.
8	Q8IYB9	ZNF595	-4.809	Yes	Yes	NA	No	-	-	-	Similar expression	Nucleus
9	Q5T4S7-3	UBR4	-4.627	Yes	Yes	No	NA	-	-	-	Similar expression	Membrane; Multi-pass membrane protein. Cytoplasm. Cytoplasm, cytoskeleton. Nucleus. Note=Concentrates at the leading edge of membrane structures involved in actin motility.
10	Q8N184-1	ZNF567	-4.203	Yes	Yes	NA	No	-	-	-	Similar expression	Nucleus
11	Q92769-3	HDAC2	-3.362	No	No	No	No	Yes (Phosphorylated in PC)	-	Class II Interactor	Similar expression	Nucleus, cytoplasm.

Table 10. List of potential PA-1 PC interactors involved in the de novo L1 silencing mechanism												
#	uniprotID	Gene names	logfold	Significant in PC vs DC comparison	PA1PC mαORF1p Significant	PA1PC LαORF1p Significant	PA1DC mαORF1p Significant	Differentially expressed phosphosite	Found in previous studies	Interacts with RNA/DNA hybrids	RNAseq	Subcellular location [CC]
12	P33993	MCM7	-2.283	Yes	No	Yes	No	-	Liu et al, 2017	-	Higher expression in DC (Significant)	Nucleus
13	Q9NZC9	SMARCAL1	0.322	No	Yes	NA	No	Yes (Phosphorylated in PC)	-	-	Similar expression	Nucleus
14	P67809	YBX1	0.7	No	Yes	No	Yes	Yes (Both in PC and DC)	Taylor et al, 2013, Goodier et al, 2013	Class II Interactor	Similar expression	Cytoplasm. Nucleus. Cytoplasmic granule. Secreted. Secreted, extracellular exosome. Note=Predominantly cytoplasmic in proliferating cells. Cytotoxic stress and DNA damage enhance translocation to the nucleus. Localized in cytoplasmic mRNP granules containing untranslated mRNAs. Shuttles between nucleus and cytoplasm. Localized with DDX1, MBNL1 and TIAL1 in stress granules upon stress. Secreted by mesangial and monocytic cells after inflammatory challenges.

PA1PC mαORF1p Significant, Significant in mαORF1p versus mIgG replicates comparison; PA1PC LαORF1p Significant, Significant in LαORF1p versus mIgG replicates comparison; PA1DC maORF1p Significant, Significant in mαORF1p versus mIgG replicates comparison.

Below, I will briefly describe each of the 14 L1-silencing candidates:

1) Q7Z7L9 (ZSCAN2) – Zinc finger and SCAN domain-containing protein 2.

The ZSCAN2 gene belongs to the Krüeppel C2H2 type family of proteins, and as other members, contains a KRAB (Krüeppel-associated box) domain and several copies of a characteristic zinc finger protein motif. It is associated with the Fanconi Anemia pathway of DNA repair and studies in mice revealed that it is expressed during embryonic development. Moreover, it was found as a retrotransposition activator in Mita *et al*, 2020. Additionally, it is expressed in adult mice testes, where it might be involved in gene regulation processes of germ cells. Selected to be specific to PA-1 PC (-8.67 logfold), significant and unique in PA-1 PC mαORF1.

Molecular function: DNA binding; DNA-binding transcription factor activity; metal ion binding.

2) Q96EY1 (DNAJA3) – DNAJ homolog subfamily A member 3.

DNAJA3 belongs to the DNAJ/Hsp40 family of proteins and it has multiple functions; it is involved in protein folding, mediates signal transduction (for proliferation, survival and apoptosis processes), and acts as a tumor suppressor. May modulate the transcriptional activity of INF gamma. It is expressed in mitochondria and involved in synapsis of neurons. Selected due to its PA-1 PC specificity (-8.33 logfold), being unique and significant in the affinity capture performed with LαORF1p and mαORF1p beads, and for being a Class III interactor of RNA/DNA hybrids.

Molecular function: ATP binding; GTPase regulator activity; Hsp70 protein binding; IkappaB kinase complex binding; interferon-gamma receptor binding; metal ion binding; NF-kappaB binding; protein-containing complex binding; protein kinase binding; transcription factor binding; unfolded protein binding.

3) P28799 (GRN_HUMAN) – Progranulin.

Related to lysosomal function and might act as a cytokine. In fact, it is a secreted protein acting as a growth factor (inflammation, wound healing and cell proliferation). Notably, it modulates inflammation in neurons, increasing their survival, growth and integrity. Interacts with the HIV-1 Tat protein and inhibits transcription from the HIV-1 promoter. Selected for being found mainly in PA-1 PCs (-8.09 logfold), being significant in PA-1 PC mαORF1p, and although it was found in PA-1 DCs, it didn't reach statistical significance.

Molecular function: chaperone binding; cytokine activity; growth factor activity; RNA binding.

4) Q14151 (SAFB2_HUMAN) - Scaffold attachment factor B2.

SAFB2 binds to the S/MAR region of DNA (scaffold/matrix attachment region), represses estrogen receptor alpha and regulates cell cycle, apoptosis, differentiation, and regulates immune system genes. Contains RNA-binding motifs. Selected due to being specific and unique to PA-1 PC (-8.04 logfold), being present in the interactomes obtained with mαORF1p and LαORF1, and being significant in mαORF1p. Interacts with RNA/DNA hybrids, being a Class I interactor. In Cristini *et al,* (2018) it was observed how SAFB2 knockdown decreased the overall amount of RNA/DNA hybrids. Remarkably, it was identified in Vuong *et al,* (2019) as an interactor of L1-ORF1p in hESCs, another pluripotent cell line with active L1-silencing (unpublished observations) and as a retrotransposition activator in Mita *et al,* 2020. Consistently, RNAseq analyses revealed that it is expressed at a higher level in PA-1 PCs when compared to PA-1 DCs (data not shown).

Molecular function: identical protein binding; RNA binding; sequence-specific DNA binding.

5) Q9P2Y4 (ZNF219_HUMAN) − Zinc finger protein 219.

ZNF219 is a member of the Krüeppel-type family of proteins. Acts as a transcriptional repressor of HMGN1 (high mobility group nucleosome binding domain 1 protein), and associates with transcriptionally active chromatin. Regulates expression of SNCA, a protein involved in synaptic activity, and it is involved in chondrocyte differentiation. Selected for being PA-1 PC specific (-7.35 logfold) and it was found in mαORF1 and LαORF1 datasets in a significant manner. Presents differential phosphorylation and it seems to be significantly phosphorylated in PA-1 PC. According to RNAseq analyses, it is predominantly expressed in PA-1 PC (significant). In addition, ZNF219 was identified as a retrotransposition activator in Mita *et al*, 2020.

Molecular function: DNA binding; DNA-binding transcription factor activity; DNA-binding transcription repressor activity, RNA polymerase II-specific; histamine receptor activity; metal ion binding; RNA polymerase II cis-regulatory region sequence-specific DNA binding

6) Q8TF39 (ZN483_HUMAN) - Zinc finger protein 483.

As other members of this family, ZN483 is a zinc finger protein with KRAB and SCAN domains. May be involved in transcriptional regulation and it is expressed in brain, among other tissues. Selected as significant in PA-1 PC (logfold -7.05), both in mαORF1p and LαORF1 datasets, and it wasn't found in PA-1 DCs. Moreover, it is expressed at a higher level in PA-1 PCs when compared to PA-1 DCs.

Molecular function: DNA binding; DNA-binding transcription factor activity; metal ion binding.

7) Q14587 (ZNF268_HUMAN) − Zinc Finger Protein 268.

As the first candidate (i.e., ZSCAN2), ZNF268 belongs to the Krüeppel C2H2 family of zinc finger proteins, and is involved in the development of ovarian cancer. It is part of the TNF-alpha-induced NF-kappa-B signaling pathway. Selected for being found mainly in PA-1 PCs (-5.65 logfold); it was found in the PA-1 PC mαORF1p dataset (significant) and in the PA-1 DC dataset (also captured with mαORF1p), but it didn't reach statistical significance. Identified in Vuong *et al,* (2019) as an interactor of L1-ORF1p in hESCs.

Molecular function: DNA binding; DNA-binding transcription factor activity; metal ion binding.

8) Q8IYB9 (ZNF595_HUMAN) − Zinc finger protein 595.

ZNF595 is another member from the Krüeppel C2H2 family of zinc finger proteins. It potentially acts as a transcription factor regulating different cellular and developmental processes. Mainly found in PA-1 PCs (-4.81 logfold), where it was found in the PC PA-1 mαORF1p dataset (significant); it was also found in the DC PA-1 mαORF1p dataset, but it didn't reach statistical significance.

Molecular function: DNA binding; metal ion binding.

9) Q5T4S7 (UBR4_HUMAN) − E3 ubiquitin-protein ligase UBR4.

UBR4 encodes for a protein that interacts with RB1 (retinoblastoma-associated protein), and RB1 is involved in the formation and maintenance of heterochromatin, by stabilizing histone methylation, recruiting histone methyltransferases leading to epigenetic transcriptional repression. UBR4 bring histone methyltransferases and histone deacetylases to genomic DNA, promoting epigenetic transcriptional silencing. Found only in the PC PA-1 mαORF1 dataset (-4.63 logfold), as a statistically significant interactor.

Molecular function: calmodulin binding; ubiquitin-protein transferase activity; zinc ion binding.

10) Q8N184 (ZNF567_HUMAN) − Zinc finger protein 567.

ZNF567 might potentially act as a transcriptional regulator, and it is expressed in testis and endometrium. Selected because it was found mainly in PC PA-1 (-4.2 logfold), on the mαORF1 dataset and as a significant interactor; it was also found in the DC PA-1 mαORF1 dataset, but as a not significant factor.

Molecular function: calmodulin binding; ubiquitin-protein transferase activity; zinc ion binding

11) Q92769 (HDAC2_HUMAN) - Histone deacetylase 2.

HDAC2 is responsible for the deacetylation of lysine residues in the N-terminal region of histones H2A, H2B, H3 and H4, inducing epigenetic silencing. It can form silencing complexes when associated with different proteins, including YY1, and can regulate cell cycle progression and developmental processes. Notably, it is highly expressed in testis and endometrium, among other tissues. In this case, there is no significant difference between PA-1 PC and DC mαORF1p datasets (-3.36 logfold), and it is not a statistically significant L1 interactor; in contrast, it was found as a statistically significant L1-ORF1p interactor in the 4 biological replicates of PC PA-1 LαORF1p captures. Beside forming silencing complexes and promoting epigenetic silencing by promoting histone deacetylation, it was selected because it interacts with RNA/DNA hybrids, and it was classified as a Class II interactor; it shows differential phosphorylation and it is significantly phosphorylated in PC PA-1.

Molecular function: chromatin binding; deacetylase activity; enzyme binding; heat shock protein binding; histone deacetylase activity; histone deacetylase binding; NAD-dependent histone deacetylase activity (H3-K14 specific); NF-kappaB binding; promoter-specific chromatin binding; protein deacetylase activity; RNA binding; RNA polymerase II repressing transcription factor binding; sequence-specific DNA binding; transcription factor binding

12) P33993 (MCM7_HUMAN) − DNA replication licensing factor.

MCM7 is part of the minichromosome maintenance protein complex (MCM), acting as a putative helicase required for DNA replication. Selected because it was found as a statistically significant L1-interactor in the PA-1 PC vs DC comparison (logfold -2.283, see **Table 10**); furthermore, MCM7 was captured using the LαORF1 nanobody in PC PA-1. Notably, MCM7 was classified as a strong activator of L1 retrotransposition in a recent genome wide CRISPR/Cas9 screening (Liu *et al,* 2018*)*. Furthermore, according to RNAseq analysis, MCM7 expression is higher in PC PA-1s when compared with DC PA-1s.

Molecular function: ATP binding; DNA helicase activity; DNA replication origin binding; single-stranded DNA binding

13) Q9NZC9 (SMAL1/SMARCAL1) − SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A-like protein 1.

SMARCAL1 is an ATP-dependent annealing helicase involved in stably reannealing unwound DNA during replication, performing the reverse function of conventional helicases. It is a member of the SWI/SNF family of proteins, known to regulate the transcription of some genes by altering the chromatin structure around them. It is highly expressed in the ovary and testicles. Selected for being significant in the PC PA-1 mαORF1p dataset; it shows differential phosphorylation.

Molecular function: annealing helicase activity; ATP binding; DNA-dependent ATPase activity; helicase activity

14) P67809 (YBX1) - Y-box-binding protein 1.

DNA and RNA binding protein involved in several biological processes, such as repression of translation, splicing, DNA repair, transcription regulation, and RNA stabilization. It is highly expressed in testes. Selected because it was found as a significant L1-ORF1p interactor in the PC PA-1 mαORF1p dataset, because it was classified as a Class II interactor of RNA/DNA hybrids (Liu *et al,* 2018*)*, and because it was previously found in two previous proteomic studies (Taylor *et al,* 2013*;* Goodier *et al,* 2013*)*. It shows differential phosphorylation between PA-1 PCs and DCs.

Molecular function: C5-methylcytidine-containing RNA binding; chromatin binding; DNA binding; DNA-binding transcription activator activity, RNA polymerase II-specific; DNA-binding transcription factor activity; DNA-binding transcription factor activity, RNA polymerase II-specific; double-stranded DNA binding; GTPase binding; miRNA binding; RNA binding; RNA polymerase II cis-regulatory region sequence-specific DNA binding; single-stranded DNA binding

VIII. DISCUSSION

1. Section I: RNase H2 role in LINE-1 retrotransposition

As explained in the previous sections, the study of the relation among LINE-1 and RNase H2 was proposed for two reasons:

1) **RNase H2 had previously been suggested to control LINE-1 retrotransposition** (Volkman & Stetson, 2014), **in a manner similar to other genes involved in AGS disease.** Results obtained in previous studies pointed to RNase H2 as another factor that could restrict L1 mobility, as observed with other genes involved in AGS (TREX1, Stetson *et al, 2008;* Thomas *et al, 2017*; SAMHD1, Zhao *et al, 2013*; ADAR1, Orecchini *et al, 2017*);

2) **Our interest in discovering possible factors interacting with the TPRT intermediates,** specifically the LINE-1 mRNA:cDNA hybrid, where some RNase H activity might be involved (Piskareva & Schmatchenko, 2006);

Therefore, considering the function of RNase H2 against RNA:DNA hybrids in the cell, we decided to use cell lines lacking RNase H2 to test whether RNase H2 activity could affect LINE-1 retrotransposition. However, the first retrotransposition assays carried out in RNase H2 KO cells (*VII. RESULTS, 1.1. L1 retrotransposition is compromised in RNase H2 knock out (KO) mutant cell lines*, **Figures 14-17**) revealed that LINE-1 retrotransposition was not augmented; on the contrary, L1 retrotransposition was severely reduced in RNase H2 KO cells (i.e., exactly the opposite of what was proposed). In addition, even when RNase H2 Knock Out (KO) cells massively miss-incorporate ribonucleotides in genomic DNA, inducing DNA damage, ENdonuclease-Independent (ENi) L1 retrotransposition doesn't occur at a high frequency in these DNA damaged sites (**Figures 13 and 14**).

Based on this observation, we considered the possibility that RNase H2 exerted a positive, rather than a restrictive function on LINE-1 mobilization, and we proposed new experiments to interrogate the possible role of RNase H2 on retrotransposition.

In the experiments performed, we demonstrated how other retrotransposons lacking an RNase H domain also rely on cellular RNase H2 activity to mobilize, whereas retrotransposons encoding for a domain with RNase H activity do not seem to require cellular RNase H2 to efficiently retrotranspose (*VII. RESULTS, 1.2. RNase H2 facilitates the mobilization of non-LTR retroelements but is not required for LTR retrotransposons and DNA transposons*, **Figures 18-20**). Moreover, overexpression of RNase H2 in wild type cells leaded to an increase in L1 retrotransposition (*VII. RESULTS, 1.3. Overexpression of RNase H2 increase L1*

retrotransposition, **Figures 21 and 22**). However, considering that RNase H2 presents activity toward embedded ribonucleotides, we wondered whether the observed reduction in retrotransposition could be due to an increase in the mutation rate of the reporter cassette from the engineered L1 caused by the lack of embedded ribonucleotide RNase H2 removal activity (*VII. RESULTS, 1.5. RNase H2 KO cells do not exhibit an increased mutation rate on retrotranscribed L1 DNAs*, **Figure 26**). We analyzed the sequence of retrotranscribed L1 sequences, from both parental and RNase H2 KO HeLa cells, and didn't observe significant differences in the mutation rate observed. Our results prove that the lack of cellular RNase H2 activity doesn't increase the mutation rate at dinucleotide repeats present in L1-retrotranscribed sequences, at least on HeLa cells.

In 2018, while we were preparing to publish our results, Choi *et al,* published a paper suggesting that RNase H2 might act as a LINE-1 restriction factor, using shRNAs to reduce RNase H2 expression. However, in our study we used multiple independent CRISPR/Cas9-edited null clones for RNase H2, employing three different cell lines and several LINE retrotransposition reporters, providing comprehensive evidence to support a role for RNase H2 on LINE-1 retrotransposition. Furthermore, our results were consistent with other recently published, suggesting that RNase H2 is indeed necessary for L1 retrotransposition (Bartsch *et al, 2017).*

In addition, further experiments where we observed complementation of the RNase H2 deficiency by overexpression of RNase H1 or an increased in L1 retrotransposition upon overexpression of the RNase H2 separation-of-function (SoF) mutant support the hypothesis that RNase H2 activity is indeed required for LINE-1 mobilization (*VII. RESULTS, 1.4. Complementation of RNase H2 KO cell lines with the RNASEH2A subunit, but not with a "separation of function" mutant version, recovers L1 retrotransposition,* ***Figures 23-25****; 1.6. Overexpression of a separation-of-function (SoF) mutant RNASEH2A rescues L1 retrotransposition,* ***Figure 27****; 1.7. Overexpression of RNase H1 in RNase H2 KO cells partially rescues L1 retrotransposition,* **Figures 28 and 29**). However, we noticed that RNase H2 SoF mutant produced longer RNA products compared with wild-type RNase H2 (**Figure 23B**), suggesting that it might exhibit an altered scission pattern on the RNA strand of heteroduplexes. Thus, RNase H2-SoF mutant failure to fully restore L1 retrotransposition suggest that efficient L1 retrotransposition requires the complete removal of the L1-RNA from L1-mRNA:L1-cDNA hybrids.

These results lead us to propose a model where RNase H2 would degrade the RNA from the L1-mRNA:L1-cDNA hybrid produced during TPRT, allowing efficient and productive LINE-1 retrotransposition (**Figure 29D**).

This model makes even more sense in light of a recent study in which a PCNA-interacting protein motif, the PIP motif, was identified in L1-ORF2p (Taylor *et al, 2013)*, and the interaction between L1-ORF2p and PCNA was shown to be required for efficient retrotransposition. Given that the RNASEH2B subunit also contains a functional PIP domain that allows RNase H2 to interact with PCNA (Chon *et al, 2009;* Bubeck *et al, 2011)* and according with our results *(VII. RESULTS, 1.8. RNase H2 may interact with LINE-1 through PCNA*, **Figure 30***)* we speculate that PCNA might act as an anchor protein connecting L1-ORF2p to RNase H2 during retrotransposition, as PCNA directs RNase H2 activity to replication and repair foci (Bubeck *et al, 2011;* Kind *et al, 2014)*. On the contrary, neither the mobilization of a DNA-Transposon nor the mobilization of an LTR-retrotransposon that contains an active RNase H domain (mouse MusD elements) is influenced by cellular RNase H2 activity. However, RNase H1 lacks a PIP-domain and we speculate that L1-ORF2p doesn't interact directly with RNase H1; therefore, nuclear RNase H1 could process LINE-1 RNA:cDNA hybrids, but by a simple diffusion mechanism. This hypothesis is consistent with: i) LINE-1 retrotransposition is not fully inhibited on RNase H2 KO cells (retrotransposition levels oscillate between 5-20% depending on the cellular background and L1 reporter used) and ii) complementation of RNase H2 null cells with RNase H1 cannot fully rescue the L1 retrotransposition deficit. It remains to be determined whether the interaction between PCNA and RNase H2 is actually required to promote retrotransposition.

Our model, in which cellular RNase H activity promotes LINE-1 retrotransposition by degrading the RNA of L1-mRNA:L1-cDNA hybrids generated during TPRT (**Figure 29D**), would explain how LINE-1 elements can function without an RNase H domain (Malik *et al, 1999;* Olivares *et al, 2002)*. We provide several lines of evidence supporting this model:

i. reduction of LINE-1 retrotransposition in cells lacking RNase H2;

ii. rescue of this defect by complementation with wild-type RNASEH2A;

iii. overexpression of RNase H1 resulting in partial recovery of retrotransposition through its RNA:DNA hybrid hydrolyzing activity;

iv. increased retrotransposition following overexpression of both wild-type and separation-of-function RNase H2A.

Overall, **we demonstrated that RNase H activity against RNA:DNA hybrids, mainly provided by cellular RNase H2, is important for efficient and productive LINE-1 retrotransposition.** This implies that RNase H2 is involved

in the degradation of the LINE-1 mRNA from the L1-mRNA:L1-cDNA hybrid and such degradation enables the synthesis of the second strand and ultimately its insertion into the genome.

Additionally, we favor a model in which a role for the conserved PIP domain found in most vertebrate LINEs (Taylor *et al, 2013)* is to ensure a functional association with RNase H2, which would degrade the RNA of RNA:DNA hybrids generated during retrotransposition. Consistent with this model, except a few examples found in lower eukaryotes and plants, LINEs lack RNase H activity, and rely on cellular RNase H to complete a round of retrotransposition (Malik *et al, 1999;* Olivares *et al, 2002)*. However, beside degradation, displacement of the RNA from RNA:DNA hybrids during second strand cDNA synthesis would also allow completion of integration during LINE retrotransposition. However, strand displacement activity during nucleic acid polymerization is highly unusual among RTs/polymerases (Lanciault & Champoux, 2004; Kelleher & Champoux, 1998). Interestingly, R2Bm, a site-specific LINE element from the silkworm *Bombyx mori*, lack RNase H activity and PIP motifs, but code for a highly processive Reverse Transcriptase that can displace RNA annealed to ssDNA without losing processivity (Kurzynska-Kokorniak *et al, 2007)*. In sum, the multiple mechanisms exploited by LINEs to degrade/displace RNAs from RNA:DNA hybrids generated during retrotransposition indicate that this is a key aspect to ensure their integration.

Given the role of RNase H2 in AGS, and our proposed model, our results suggest that L1 RNA:DNA hybrids accumulated in patient´s cell could influence the pathophysiology of AGS. As mutations in genes encoding RNase H2 are a frequent cause of AGS (Crow & Rehwinkel, 2009), our findings are relevant with respect to the possible sources of the immunostimulatory nucleic acids thought to cause autoinflammation. Two sources for such cytoplasmic nucleic acids have been proposed: DNA damage or retroelements (Crow & Manel, 2015). In particular, active LINE-1s are expressed and highly active in the central nervous system (Muotri *et al*, 2005; Coufal *et al*, 2009) and strong experimental evidence suggests that TREX1, SAMHD1 and ADAR1 act as LINE-1 restriction factors, since its absence leads to an increase in LINE-1 retrotransposition (Stetson *et al, 2008;* Zhao *et al, 2013;* Orecchini *et al, 2017;* Thomas *et al, 2017)*. In contrast and according to our model, those AGS patients carrying homozygous inactivating mutations in RNase H2, wouldn't be characterized for accommodating higher L1 retrotransposition levels in their genomes.

Although in a previous study they were unable to detect an elevated L1 retrotransposition in the hippocampus of an AGS patient with mutations in SAMHD1 (Upton *et al, 2015)*, more recent work using TREX1-deficient neuronal cells generated from human embryonic stem cells has implicated the accumulation of LINE-1-derived single-stranded DNAs (ssDNAs) in type I IFN production and neurotoxicity (Thomas *et al, 2017)*, consistent with the role of active LINE-1s in the pathophysiology of AGS (Garcia Pérez & Alarcón-Riquelme, 2017). This work reinforces the hypothesis that by-products of active retrotransposition, beyond the accumulation of LINE-1 insertions *per se*, might be relevant for the pathology of AGS patients (Upton *et al, 2015)*, with TREX1 routinely degrading single-stranded LINE-1s (ssDNA) from retrotransposition intermediates (Stetson *et al, 2008;* Garcia Pérez & Alarcón-Riquelme, 2017; Thomas *et al, 2017)*.

Recently, an increase in DNA damage and a cGAS-STING-dependent increase in interferon-stimulated genes (ISG) expression was demonstrated in mouse embryonic fibroblasts (MEFs) lacking RNase H2 (Mackenzie *et al, 2016)*, and a novel mechanism linking genomic instability to inflammation has been established, with micronuclei providing a source of cytoplasmic DNA capable of activating cGAS (Mackenzie *et al, 2017)*. However, the direct relevance of this to AGS has yet to be determined. In this study, neither DNA damage nor the ISG response was mitigated by RNase H1 overexpression (Mackenzie *et al, 2016)*, in contrast to the partial rescue of the retrotransposition defect observed in RNase H2-lacking HeLa cells overexpressing RNase H1 (**Figure 28 and 29**). Thus, it is formally possible that genomic instability is the underlying cause of autoinflammation in AGS associated with RNase H2 mutations.

Despite the above data, since we have observed that RNase H2 activity is required for LINE-1 retrotransposition and as cells with AGS-causing mutations in RNase H2 would have reduced levels of productive retrotransposition (*VII. RESULTS. 1.9. RNASEH2A mutations characterized in AGS patients also reduce L1 retrotransposition,* **Figure 31**), it would be possible that the accumulation of LINE-1 mRNA:cDNA hybrids in the nucleus could be a source of immunostimulatory nucleic acids. However, it is currently unclear how such hybrids, that are covalently bound to the genome, would ultimately access the cytoplasm and activate recognition receptors. An additional question is whether physiological levels of retrotransposition byproducts are sufficient to elicit the observed inflammatory response.

However, alternative explanations are of course possible; there is increasing evidence that DNA double-strand break repair may be RNA-mediated *(*Keskin *et*

al, 2014; Ohle *et al, 2016;* Michelini *et al, 2017)*. Thus, an intriguing possibility would be that RNA:DNA hybrids may play a more active role in the LINE-1 retrotransposition process, for example, by recruiting the DNA repair machinery and RNase H2. In addition, RNase H2 deficiency may also result in larger genomic rearrangements (Reijns *et al*, 2012*)* that could affect LINE-1 retrotransposition. Although we cannot rule out that this type of genomic instability interferes with retrotransposition, or even that non-productive retrotransposition contributes to the increased genomic rearrangements in RNase H2 null cells, it seems unlikely that this is the reason for the reduced LINE- 1 retrotransposition that we observed, particularly because RNase H2 overexpression results in a higher rate of retrotransposition. Further work is still needed to determine the relative importance of retroelement activity and genomic instability in AGS.

In conclusion, this Thesis contributes to the mechanistic understanding of LINE-1 retrotransposition, as we demonstrate that cellular RNase H2 plays an important role in LINE retrotransposition, explaining how LINE elements lacking an RNase H domain can retrotranspose effectively in cells. Furthermore, our data add a new layer of complexity to the understanding of AGS pathophysiology, as we demonstrate that not all AGS proteins are LINE restriction factors, as previously observed and suggested.

2. Section II: Study of the LINE-1 interactome

As described in sections III. and IV. (see *III. INTRODUCTION* and *IV. HYPOTHESIS, Theoretical background supporting the hypothesis*), it is now well established that **LINE-1s can mobilize in our pluripotent genome**, specifically during early stages of human embryonic development, resulting in the accumulation of new and heritable L1 insertions *(III. INTRODUCTION, 3. LINE-1 retrotransposition in pluripotent cells)* (Van den Hurk *et al,* 2007*;* Garcia-Perez *et al,* 2007, 2010; Klawitter *et al,* 2016*;* Kano *et al,* 2009*;* Wissing *et al,* 2012*)*. Due to their mutagenic potential in our heritable genome, and to prevent the accumulation of high rates of LINE-1 insertions, **cells have evolved different mechanisms to inhibit their propagation**. A major mechanism used to regulate LINE-1 retrotransposition involves silencing their expression, through **methylation of L1 promoters** *(III. INTRODUCTION, 4.1.1. DNA methylation of the L1 promoter)* (Thayer *et al*, 1993; Bourc'his & Bestor, 2004; Coufal *et al*, 2009). Indeed, silencing L1 expression would also prevent the mobilization of non-autonomous SINE retrotransposons such as Alu and SVA. However, during early stages of human embryonic development, the genome of totipotent embryo cells is **hypomethylated in a genome wide manner, including L1 promoters, and this is in fact exploited by LINE-1 to accumulate *de novo* insertions** that can be transmitted to the next generation, ensuring their evolutionary success over evolution (Zeng & Chen, 2019; Garcia-Perez *et al,* 2016*)*. In addition, it has been observed that during LINE-1 mobilization, specifically during the early steps of the TPRT process *(III. INTRODUCTION, 2.2. The LINE-1 retrotransposition mechanism)*, a Y-like structure consisting of a cDNA:RNA hybrid that is covalently bound to the genome is generated, and its formation on CpG islands prevents methylation of the underlying DNA sequence (Lin & Scott, 2012; Ginno *et al,* 2013*)*.

Because pluripotent cells can´t exploit DNA methylation to control L1 expression/retrotransposition, alternative mechanisms might operate in these cells to regulate retrotransposition. In this context, **a novel epigenetic mechanism that silence *de novo* LINE-1 insertions** in a sequence-independent manner, most likely directed against TPRT intermediates (Y-like structure), was found to **act exclusively in pluripotent human cells (PCs)**. **Chromatin modifying factors**, such as histone deacetylases, **are involved in maintaining L1 silencing in pluripotent cells (PCs)** (Garcia-perez *et al,* 2007*;* Garcia-Perez *et al,* 2010*;* Wissing *et al,* 2012*) (4.1.2. Histone deacetylation as a possible L1 restriction mechanism)*. However, **we know very little of L1-silencing in PCs, and it is currently unknown how PCs could recognize *de novo* L1 insertions,** or what the silencing mechanism is. An important aspect of **L1-silencing is its attenuation in differentiated cells (DCs)**,

and knowing that this mechanism is not active as cells differentiate can provide key mechanistic clues (Garcia-Perez *et al*, 2010*)*.

Previous studies of the L1 interactome provided a broad view of host factors that interact with L1-RNPs. However, most studies used **transformed cell lines overexpressing engineered L1s with epitope-tagged L1 proteins (L1-ORF1p and L1-ORF2p)** (Goodier *et al*, 2013*;* Mandal *et al*, 2013*;* Taylor *et al*, 2013*;* Moldovan & Moran, 2015; Taylor *et al*, 2018*)***.** The **same is true for the few genome wide Loss of Function studies** that found dozens of host factors that act to promote or repress L1 retrotransposition (Liu *et al*, 2018*;* Mita *et al*, 2020*)*. In fact, the use of transformed cell lines implies that any impact on L1 biology might be relevant only to cellular niches similar to that found in human cancers, and to some level to somatic cells. Although information obtained using transformed cell lines and L1 overexpression vectors has proven to be very useful to uncover L1 biology processes, it is very likely that the composition of the L1 interactome and the regulation of L1 in physiological cell models might be fundamentally different.

So far, only one L1 interactome study has been conducted using physiologically relevant pluripotent human cells, using human embryonic stem cells (hESCs) (Vuong *et al,* 2019*)*. This is the first study that has analyzed the L1 interactome in pluripotent non-transformed cell lines, and that used cells mimicking a cellular niche where endogenous L1s are expressed and retrotranspose in humans. However, a major limitation of this study is that advanced proteomic techniques (i.e., as cryomilling, I-DIRT, etc) were not employed, and most of the identified factors were found in the control capture (IgG control coupled beads) and in the L1-ORF1p captures. Thus, the overall relevance of this study is very limited.

A clear conclusion from previous L1 interactome studies is that, although several factors that regulate retrotransposition were found, there was little overlap among identified factors (see *III. INTRODUCTION, 6.1. Proteomic study of the LINE-1 interactome*, **Table 3**). Indeed, the **cell type used to perform such studies seems to influence the nature of the L1 interactomes detected** (Vuong *et al,* 2019*)*. Data in our laboratory has demonstrated that **endogenous active L1s are expressed in PA-1 cells**, as revealed by Reverse Transcriptase activity assays (endogenous LEAP, not shown) and by the characterization of *de novo* L1 insertions accumulated in these cells (data not shown). Therefore, in this work **I decided to compare the interactome of isogenic pluripotent PA-1 (PCs PA-1) and differentiated PA-1 cells (DCs PA-1),** a comparison potentially leading to the identification of interactors involved in LINE-1 silencing and regulation. An important aspect of this work is the use of advanced proteomic methods, and to do that we collaborated with the lab of Dr. John LaCava (Rockefeller University, US), to study

the L1 interactome in pluripotent and differentiated PA-1 cells by capturing L1-RNPs and their interactors in the ectopic and endogenous contexts. Each approach has advantages and limitations, expecting to obtain potentially different results; under endogenous conditions, the captured LINE-1 proteins could belong to fossil or incomplete copies still able to be expressed but remaining inactive (Lander *et al*, 2001*)*. On the other hand, the overexpression of ectopic L1-encoded proteins can activate restriction pathways that would change the composition of the L1-interactome, and overexpression of L1 encoded proteins can create unspecific interactions.

On a first approach we therefore attempted to capture the LINE-1 interactome in PA-1 cells ectopically expressing a human RC-L1, using 3xFLAG-tagged L1-ORF2p expression vectors to compare the interactomes obtained under ectopic conditions with those of previous studies and with the endogenous LINE-1 interactome. However, the implemented mass spectrometry approaches to characterize L1-RNPs and interactors developed and optimized by Dr LaCava's group (LaCava *et al*, 2015; LaCava *et al*, 2016*)* required significant amounts of endogenously expressed L1 proteins. Thus, as described in section 2.1 of *VII. RESULTS*, transfected PA-1 cells expressed L1-ORF2p-3xFLAG significantly below the expression levels of HEK293T$_{LD}$ cells overexpressing pMT302 (Dai *et al*, 2012; Taylor *et al*, 2013*)*, not being possible to obtain a high enough yield when capturing L1-RNPs employing dynabeads coupled to an anti-FLAG antibody, prohibiting us from implementing protocols optimized using HEK293T$_{LD}$ cells (*VII. RESULTS, 2.1. Capture of ectopic L1-RNPs in PCs and DCs*, **Figures 32-34**).

To solve this limitation, we next decided to generate PA-1 Tet-On cell lines by integrating the rtTA-Advanced element (analogous to the *tet* element from HEK293T$_{LD}$) (*VII. RESULTS, 2.1.1. Generation of inducible TetOn Advanced (Clontech) PA-1 cell lines to overexpress LINE-1 proteins*, **Figures 35-37**), in order to create inducible cell lines to increase the expression levels of LINE-1 encoded proteins. As described in section *2.1.1.* of *VII. RESULTS*, we successfully generated a small panel of PA-1 Tet-On cell lines, but when testing the expression levels of the rtTA-Advanced protein, most PA-1 Tet-On cell lines expressed very low levels of the rtTA-Advanced protein, and rtTA expression levels were much lower than those expressed by HEK293T$_{LD}$ (**Figure 37**). Because rtTA-Advanced protein expression levels are directly related with the achievable levels of inducible expression, none of the cell lines tested overexpressed enough L1 encoded proteins. We speculate that PA-1 cells silenced rtTA-Advanced expression, and even if cells retained resistance to blasticidin or neomycin (conferred by the vectors encoding the rtTA-Advanced), the PA-1-derived cell lines were unable to induce expression of L1

encoded proteins to levels that would allow implementing the optimized proteomic protocols established by the LaCava lab.

In consequence, as the proteomic analyses planned required a minimum level of L1 expression not achieved by engineered systems, we decided to characterize the interactome of endogenous L1s, capturing endogenously expressed L1-RNPs (*VI. RESULTS, 2.2. Capture of endogenous L1-RNPs using PC and DC PA-1 cells*). To do that, we first optimized the affinity capture process to identify co-immunoprecipitated interactors by mass spectrometry (*VI. RESULTS, 2.2.1. Optimization of endogenous L1-ORF1p affinity capture,* **Figures 38-41**), and we then studied the LINE-1 interactome of endogenous elements in PC and DC PA-1 cells (*VI. RESULTS, 2.3. Capturing L1-RNPs and their interactors on PA-1 PC and DC*).

A key aspect of the proteomic approaches employed (**Figure 42**) is the maintenance of L1-RNPs' native interactors, so that we could be able to detect interactions formed *in vivo* from those that occurred post-lysis, to distinguish between real interactors and putative contaminants (Domanski *et al,* 2012*)*. During the optimization process, we performed western-blot of L1-ORF1p from the affinity captures and concluded that at least 500 mg of cells would be required for proteomic experiments, to make sure that the obtained material had detectable levels of L1 proteins and their co-purified interactors (*VI. RESULTS, 2.3. Capturing L1-RNPs and their interactors on PC and DC PA-1 cells*, see Sypro-stained gels from **Figures 43, 46 and 49**).

After optimization, we next purified endogenously expressed L1-RNPs in PCs using affinity capture with antibodies directed against L1-ORF1p (*VI. RESULTS, 2.3. Capturing L1-RNPs and their interactors on PC and DC PA-1 cells*), and **using two independent antibodies:** a mouse monoclonal anti-ORF1p antibody (mαORF1p), and a llama-derived anti-ORF1p nanobody (LαORF1p), both as coupled magnetic dynabeads. However, the interactome of isogenic PA-1 DCs was characterized using only the mouse monoclonal anti-ORF1p beads (mαORF1p). To obtain the desired yield from PC and DC PA-1 cells, we used a large amount of material (4 gr of cells per experiment); as negative controls of affinity captures, we used dynabeads coupled to mouse Ig G or to an EGFP nanobody (LαG94-10 nanobody) in experiments using the mouse monoclonal anti-ORF1p antibody (mαORF1p) or the llama-derived anti-ORF1p nanobody (LαORF1p), respectively.

Upon bioinformatic analyses, we generated a list of L1-interactors that were specific to PCs or DCs; additionally, we compared the factors identified in PCs using the two antibodies. As expected, both in PCs and DCs, we found several *bona fide* L1-interactors that were found in previous studies (Goodier *et al,* 2013*;* Mandal

et al, 2013; Taylor *et al*, 2013; Moldovan & Moran, 2015; Taylor *et al*, 2018), demonstrating the reliability of our proteomic approach. Our results revealed that **7.9% and 12.3% of the L1-interactors identified in PC and DC PA-1 cells, respectively, were found in previous proteomic studies,** even if they used ectopically expressed L1s and transformed cell lines (see *VII. RESULTS, sections 2.3.1., 2.3.2., and 2.3.3*, **Tables 4, 5 and 6**). Well-known L1-interactors, such as MOV10, PABPC1, ZCCHC3 or TUBB4B, were identified in PC PA-1s using both the monoclonal antibody (mαORF1p) and the nanobody (LαORF1p) against L1-ORF1p. Other factors, such as MATR3, or DDX5, were identified both in PC PA-1s using the nanobody and in DC PA-1s using the monoclonal antibody against L1-ORF1p (see *VII. RESULTS, 2.4. Comparison of L1-ORF1p interactomes from isogenic PCs and DCs,* **Tables 7 and 8**). Surprisingly, there is little overlap with factors previously found in the study conducted by Vuong *et al*, 2019 using hESCs. However, the methodology used, as well as the method used to analyze data, could be responsible for these differences; in fact, in any comparison, we only considered factors identified as significant, while Vuong *et al*, 2019 didn't include significance in their analyses, making it difficult to directly compare with their data.

However, we also found factors that have not been described to date as L1 interactors. Indeed, a major finding from our proteomic experiments is that most interactors found, in PCs and DCs, are novel interactors. The interactome of L1-ORF1p in PC PA-1 cells revealed that most L1-interactors **(92 %, 164 from a total of 178 significant interactors identified in PC PA-1 cells) haven't been found in previous L1 proteomic studies**, conducted with ectopically expressed L1s and transformed cell lines. Similarly, the interactome of L1-ORF1p on isogenic DC PA-1 cells also revealed that most L1-interactors **(87.6 %, 71 from 81 significant interactors identified in DC PA-1 cells)** are also novel within L1 proteomic studies. There are several reasons that could explain our results: i) the proteomic techniques employed (i.e., cryomilling, rapid affinity capture; see *VI. MATERIALS AND METHODS*) allow the preservation of native intermolecular interactions, allowing us to detect weak interactors that would be detached using more aggressive physical/chemical techniques; and ii) the use of transformed cells overexpressing engineered L1 constructs in previous studies may lead to a stoichiometric imbalance in protein complexes, promiscuous interactions or viral restriction pathway activation. Thus, because in our approach we are detecting the interactome of endogenous L1-ORF1p/L1-RNPs in pluripotent human cells, and using advanced proteomic methods, uncovering novel interactors might not be that surprising.

To characterize the L1 interactome of PCs in detail, we compared factors identified with the monoclonal antibody and with the nanobody against L1-ORF1p, and, surprisingly, we found that both interactomes had many unique factors which had never been described before. From **a total of 178 significant interactors identified** employing both antibodies, **63.5% were only captured by the monoclonal antibody (mαORF1p)**, and **24.7% only by the nanobody (LαORF1p)** (see *VII. RESULTS, 2.4. Comparison of L1-ORF1p interactomes from isogenic PCs and DCs*, **Table 7, Figure 52**). Moreover, among the newly described PC PA-1 interactors, we identified **many new Zinc Finger Proteins (ZNFs)**, equivalent to a 18.5% (33 out of 178) of all the significant factors identified in PC PA-1s, 30 of them detected only using the monoclonal antibody (mαORF1p). While inherent characteristics of mouse monoclonal IgG antibodies and llama nanobodies might explain the marked differences found between both L1-interactomes, their size is perhaps the main differential characteristic. We speculate that, due to their lower size, nanobodies might be able to better interact with large macromolecular complexes, as L1-ORF1p could be literally hidden behind a large constellation of proteins and/or RNAs, avoiding the access of large size antibodies (i.e., mouse monoclonal). According to this possibility, when comparing with mouse monoclonal interactors, I anticipate that a large number of nanobody interactors might directly interact with L1-ORF1p, rather than through RNA and/other proteins; similarly, it is possible that the cellular localization of large macromolecular L1-RNPs presumably identified with nanobodies might be different from those identified using conventional antibodies, which are mostly found as large aggregates in the cytoplasm. Notably, Gene Ontology (GO) analyses of the unique factors identified by each of the antibodies (monoclonal or nanobody), revealed that 62% and 72% of the LαORF1 and mαORF1 associated factors, respectively, were associated to nuclear functions. However, Biological process and Molecular function categories provided little information, as the function of most of the factors identified are currently unknown (**Figures 54 and 55**). Thus, pluripotent human PA-1 cells express a constellation of L1-RNPs, differing in the composition of L1-ORF1p interactomes, as revealed by our affinity capture experiments using mouse monoclonal (mαORF1p) and llama-derived nanobodies (LαORF1p) anti-human_L1-ORF1p. While more research is needed to further clarify why the composition of the L1-ORF1p-interactome is influenced by the antibody used in captures, these analyses uncovered a novel aspect of L1-biology, which indicates that the composition of L1-RNPs is highly variable even within the same cell type.

Regarding the many zinc finger proteins associated to the PC PA-1 L1-interactome, it is worth noting that all (a total of 33) belong to the C2H2 type. These proteins are known to interact with DNA, playing an important role in processes such as embryonic development, cell proliferation and differentiation, cell cycle regulation and apoptosis. However, little is known about their role in interacting with RNA and/or other proteins (Fedotova *et al*, 2017; Brayer *et al*, 2008; Brayer & Segal, 2008; Swamynathan, 2010). For this reason, this finding may suggest that the interaction of numerous zinc finger proteins with LINE-1 mRNA or LINE-1 proteins, might play a regulatory role. On the other hand, 22 of the 33 identified ZNFs also belong to the Krüeppel-associated box (KRAB) zinc finger family of proteins. As discussed in the introduction (*III. INTRODUCTION, 4.1. Epigenetic control of L1 expression: regulation at the chromatin level*), it has been previously demonstrated that LINE-1 expression is epigenetically regulated by KRAB proteins: KAP1 (TRIM28) (KRAB-associated protein 1) mediates heterochromatin formation and is recruited at LINE-1s by zinc finger proteins containing KRAB motifs (KRAB-ZNFs). However, according to the proposed model (Ecco *et al*, 2017; Imbeault *et al*, 2017), KRAB-ZNFs would bind to DNA via their zinc fingers and recruit KAP1, which would assemble a repressor complex. Could the identified ZNFs belong to such silencing mechanism? The identified ZNFs would be interacting with L1-ORF1p or associated interactors, or perhaps even with L1-mRNA, which doesn't coincide with the current model. Whether the ZNFs interacting with L1-RNPs on PCs act using the currently known silencing mechanism mediated by KAP1 needs to be established. Alternatively, these ZNFs that interact with L1-RNPs in the cytoplasm of PCs might regulate L1 expression/retrotransposition by a new and currently unknown mechanism/s. However, I speculate that these ZNFs could be recruited prior to the generation of the new insertion by direct interaction with L1 proteins, influencing L1 integration and/or L1-silencing when L1-RNPs access the nucleus during retrotransposition. Indeed, and while the interaction of ZNFs with L1-RNPs in the cytosol might looks paradoxical, we recently demonstrated that the interaction of L1-RNPs with PCNA, which regulates TPRT in the nucleus, actually occurs in the cytoplasm (data not shown). Thus, and while ultimately the regulation of L1 by ZNFs might occur in the

nucleus, they could first interact in the cytoplasm, likely during/shortly after L1-RNP assembly.

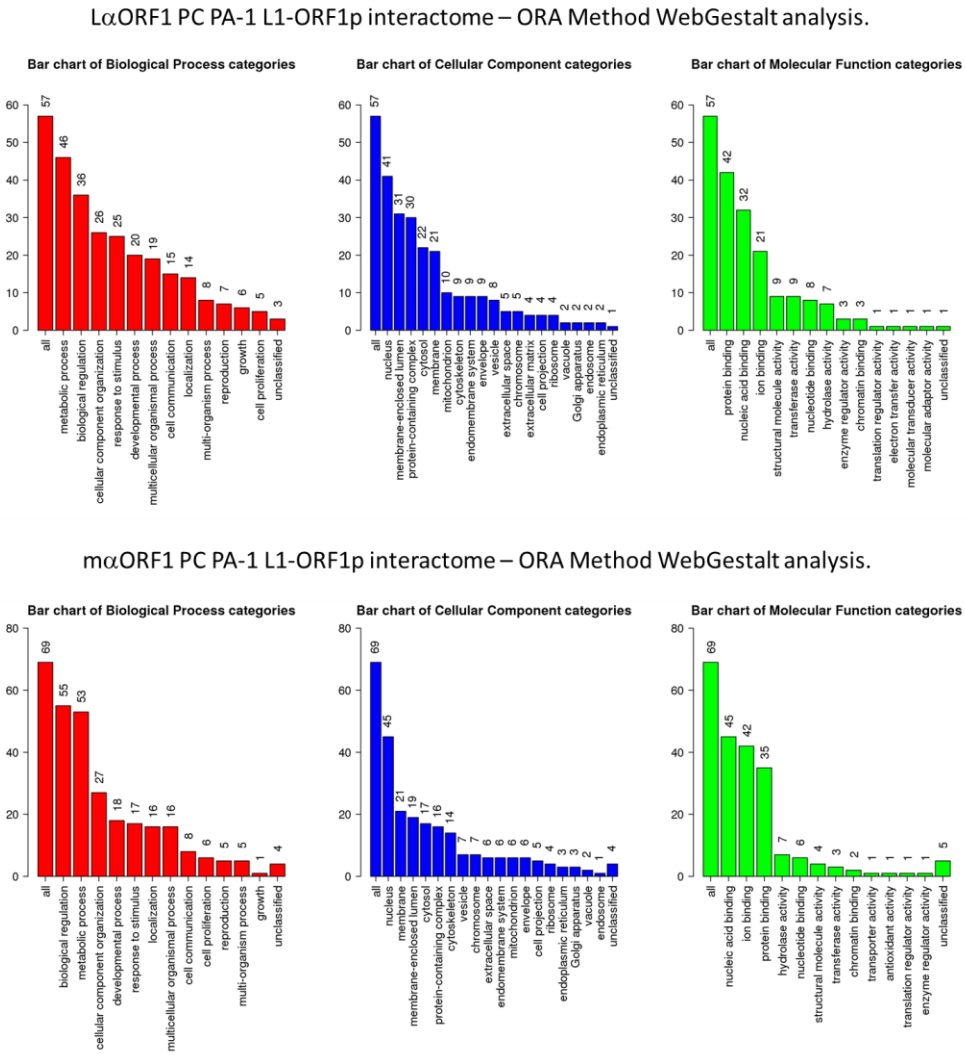


Figure 54. GO analysis of mαORF1p and LαORF1p detected L1-interactomes from PC PA-1. *ORA enrichment method from WebGestalt was employed. Each Biological Process, Cellular Component and Molecular Function category is represented by a red, blue and green bar, respectively. The height of the bar represents the number of factors in the list and also in the category.*

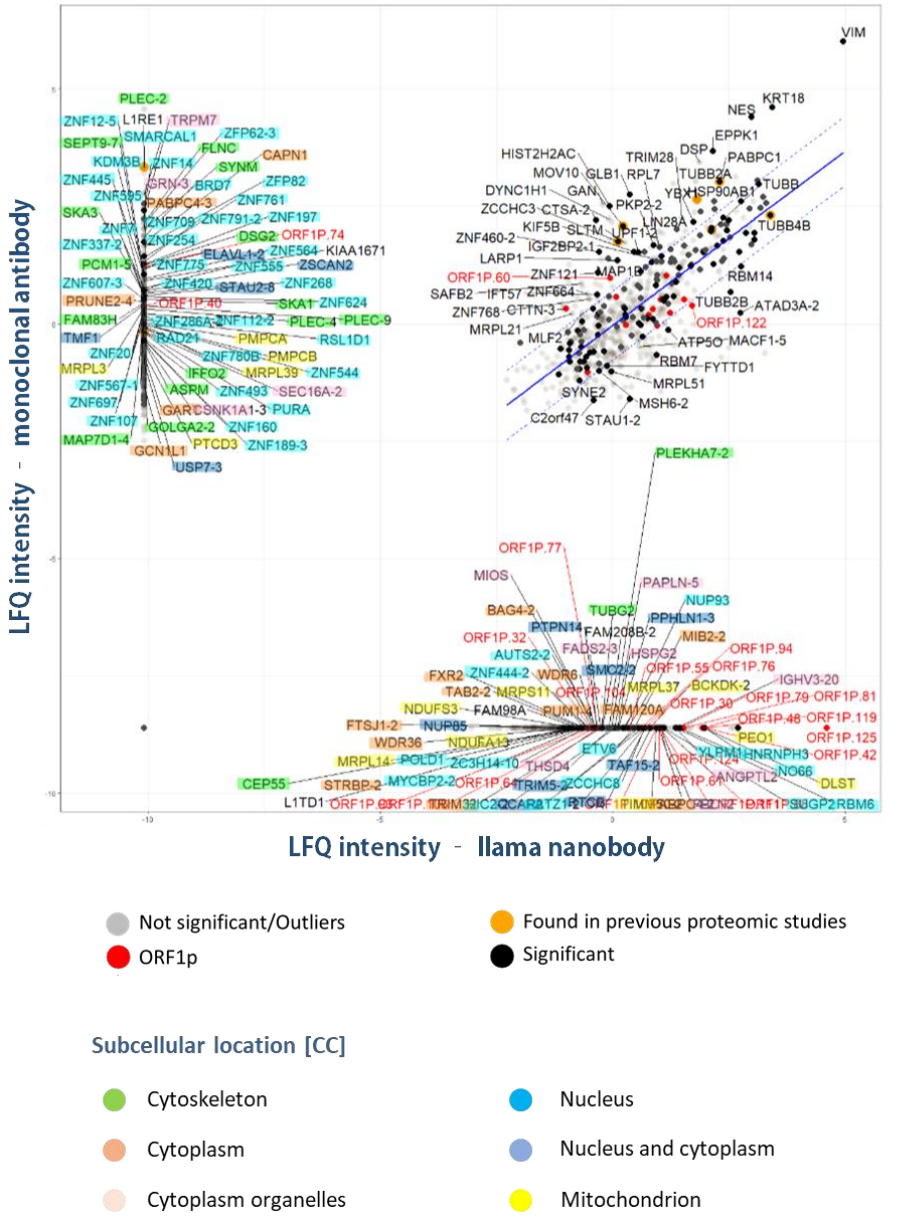


Figure 55. Subcellular localization of PC PA-1 interactors. *Scatter plot of PC interactors detected with monoclonal antibody (mαORF1p) vs nanobody (LαORF1p). Proteins are represented according to their Avg. LFQ intensity (VI. MATERIALS AND METHODS, 22. Mass spectrometry data analysis). Upper right quadrant proteins which showed statistical significance in both mαORF1p and LαORF1p; left quadrant, proteins unique and significant in the interactome captured with mαORF1p; bottom right quadrant, proteins unique and significant in the interactome captured with LαORF1p. Highlighted in black are significant factors (t-test); red, human_L1-*

ORF1p; yellow, factors found in previous studies. Color code indicates the main subcellular location [CC] associated with mαORF1p or LαORF1p unique factors.

To further identify differences on L1-regulation among isogenic PC and DC PA-1s, we next compared their L1-interactomes; these analyses revealed that a large number of L1-interactors are specific to either PCs or DCs (65.8% and 24.9% of the total identified interactors (n=237) are found only in PCs and DCs, respectively), further stressing how LINE-1s are differentially regulated in pluripotent and differentiated cells. Thus, the characterization of host factors interacting with endogenously expressed L1-RNPs using pluripotent and differentiated cells **revealed that the L1 interactome is quite dynamic**, and on average **>85% of pluripotent-specific L1-interactors change with differentiation** (156 out of 178 PC PA-1 specific interactors) (see *VII. RESULTS, 2.4. Comparison of L1-ORF1p interactomes from isogenic PCs and DCs*, **Table 8, Figure 53**).

Subsequently, we analyzed in detail L1-interactors found in PC PA-1s that were absent in DC PA-1s. I speculate that the host factors linked to L1 silencing will most likely appear as PC-interactors, and these would be reduced or absent in the list of DC L1-interactors. We used the compiled information to establish a list of host factors linked to L1-silencing in PCs. Following the selection procedure mentioned in Results and Methods (see *VII. RESULTS, 2.5. A list of PC-specific L1-ORF1p interactors potentially associated with L1-silencing in PCs*, **Table 9**), I first selected 30 PC PA-1 L1-ORF1p associated interactors. Remarkably, several of the selected interactors were involved in related paths, according to STRING clustering analyses, and some were functionally related (**Figure 56**). In fact, STRING analyses identified pathways such as "KRAB-ZNF/KAP Complex" (ZNF595, ZNF12, ZNF483, ZNF268, TRIM28; annotated in Reactome Pathways), "Viral carcinogenesis" (HDAC2, DNAJA3, UBR4; manually curated metabolic and signaling pathway imported from KEGG) and different pathways related with mRNA stability and formation of spliceosomal complexes, where YBX1 and HNRNPU were included. In addition, I performed a search for each of the factors in the BioPlex Explorer human interactome database (Huttlin *et al*, 2015, 2017, 2021), to confirm interactions. However, there was only real evidence for interactions between ZNF219 and HDAC2 in this database. Other interactors were connected by different studies, not specific about the interactome or conducted in other organisms different from *Homo sapiens* (too many studies to be listed – but the link to the STRING search is provided in **Figure 56**). Thus, further research into the role of these factors and the pathways in which they are involved is needed to draw conclusions.

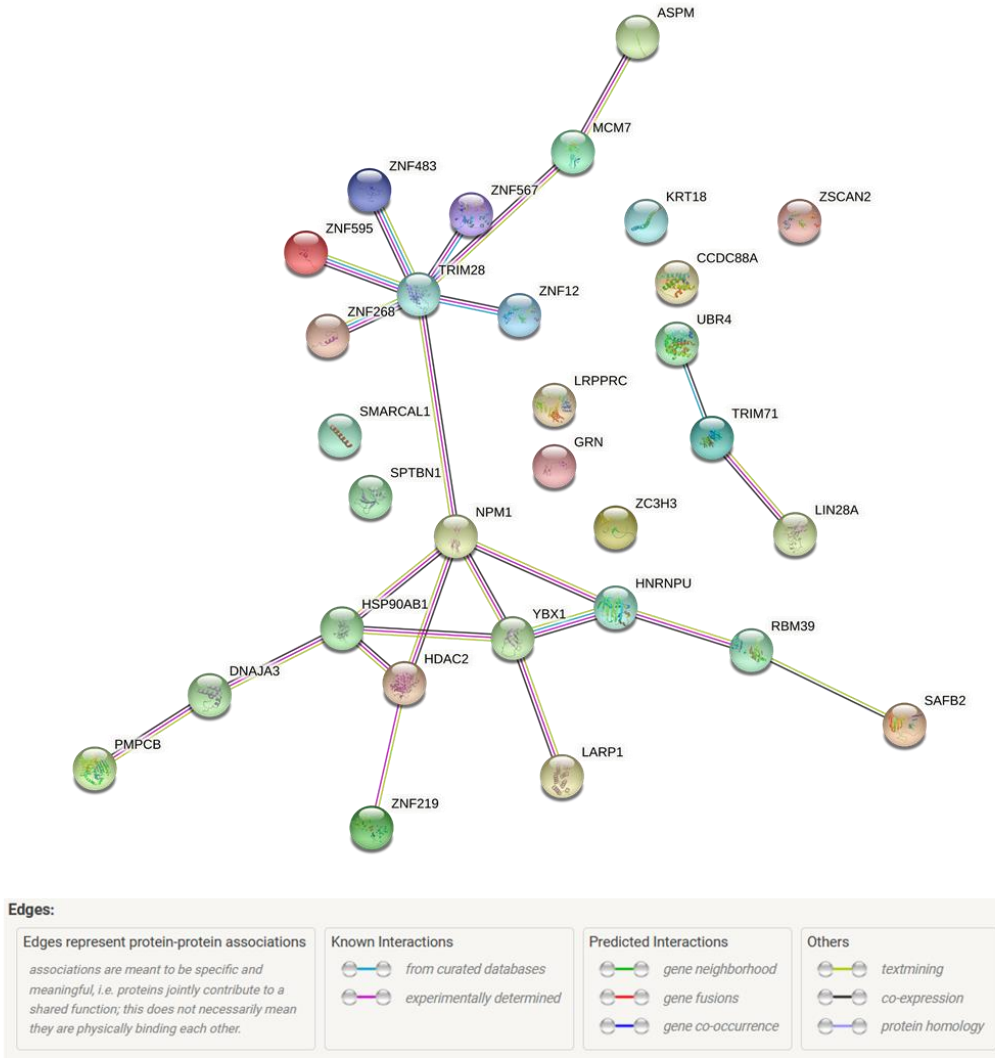


Figure 56. Pathways shared among the 30 factors selected according to STRING. *Description in the main text. Link to search: https://version-11-0b.string-db.org/cgi/network?networkId=blRnAEgKmaSA.*

From the 30 previously selected factors, following the criteria described in *VII. RESULTS, section 2.5.* to rank factors, we generated a list of potential LINE-1 interactors that may be involved in L1-silencing. Most of the 14 factors selected were specific to PC interactomes and could also interact with RNA:DNA hybrids (see *VII. RESULTS, 2.5. A list of PC-specific L1-ORF1p interactors potentially associated with L1-silencing in PCs*, **Table 10**).

In ongoing and future work, we plan to validate the interaction of these candidates with endogenous L1-ORF1p, by co-IP/western blotting. Moreover, in these experiments we plan to study if the interactions are RNA dependent (using RNase A/T1) and/or nucleic acid dependent (using Benzonase (Sigma)); similarly, confocal microscopy will be used to further analyze their co-localization. Engineered RC-L1s containing epitope tags in L1-ORF1p and L1-ORF2p will be also used to analyze whether the interaction occur with L1-ORF1p, L1-ORF2p or both (i.e., the context of an L1-RNP).

To explore the functional role of these factors in L1 regulation and L1-silencing, we are currently using Gain and Loss of function approaches (GOF and LOF) in DCs and PCs, respectively. In ongoing LOF experiments, we are generating CRISPR-KO PC models, while in GOF experiments we plan to OverExpress (OE) candidate cDNAs in DCs (both transient and stably). Next, using established assays, we will determine whether L1-silencing is attenuated in any of the PC KO models generated; a key control in these experiments is to ensure that the generated KO models retain their pluripotent status, as L1-silencing is attenuated in DCs, using methods established in the lab. On a complementary approach, we will determine whether L1-silencing can be established in DCs upon overexpressing candidate interactors; although less likely, an important control in these experiments will be to test whether OE of cDNAs increase the pluripotent status of DCs.

In sum, the combination of these assays will allow us to interrogate whether any of the 14 candidate factors selected are involved in L1 silencing in PCs, either in the initiation or maintenance of silencing as originally described (Garcia-Perez *et al*, 2010*)*. Indeed, to further determine whether any of the selected factors is involved in initiation and/or maintenance of L1-silencing, we plan to use pk-5 and pc-39 cells, which are two pluripotent PA-1 clonal lines that contain a silenced L1-EGFP insertion (described in Garcia-Perez *et al*, 2010*)*; due to active maintenance of L1-silencing, EGFP is poorly expressed from the retrotransposed L1-EGFP present in the genome of pk-5 and pc-39 cells, although expression can be induced in 100% of the cells after treating with IHDACs for less than 18h. Thus, we will use the same CRISPR/Cas9 tools to inactivate expression of candidate genes on pluripotent pk-5 and pc-39 cell lines; if any of the candidate genes is involved in the maintenance of L1-silencing, we expect to detect expression of EGFP from the L1-EGFP insertion present in pk-5 and pc-39 cells in nearly all cells (using IHDACs as positive controls in these assays). However, if any of the factor is involved in the initiation of L1-silencing, we expect no changes on EGFP expression from the

L1-EGFP insertion present in pk-5 and pc-39 cells. To complement these experiments, we will use the PA-1 PC KO models to determine the kinetics of L1-retrotransposition/silencing, using an RC-L1 tagged with the EGFP retrotransposition indicator cassette (i.e., *mEGFPI* cassette) and IHDAC treatment; if the candidate is involved in initiation of L1-silencing, we expect to detect robust EGFP expression from early time points, using IHDAC treated cells as control; differentiated PA-1 PC KO cells will be used as an additional positive control, as no silencing is observed in DCs and EGFP expression can be detected at early time points as described (Garcia-Perez *et al,* 2010*)*; however, if the host factor is involved in maintenance of L1-silencing, we expect to detect a delay in EGFP expression, again using IHDAC treated cells and differentiated PA-1 PC KO cells as controls. Importantly, the same experimental design can be used on PA-1 DC stable OE models, allowing us to unambiguously determine whether any of the selected candidates is involved in initiation or maintenance of L1-silencing.

To corroborate that any of the factors tested is indeed involved in L1-silencing, initiation or maintenance, we will use the PA-1 PC KO and the PA-1 DC stable OE models to generate a small pool of subclonal lines containing a *de novo* L1 insertion tagged with EGFP (i.e., using L1s tagged with the *mEGFPI* cassette). Using the PA-1 PC KO models, if any factor is involved in the initiation or maintenance of L1-silencing, we expect that EGFP will be robustly expressed from the integration site present in the sub-clonal lines; naïve PA-1 cells will be used in parallel, where we determined that >90% of the subclonal lines with *de novo* L1 insertions don't express EGFP, due to L1-silencing (see Garcia-Perez et a., 2010). On the contrary, using the PA-1 DC stable OE models, if any factor is involved in the initiation of L1-silencing (and likely maintenance), we expect lack of EGFP expression from the integration site present in the sub-clonal lines.

In summary, by using assays and methods already optimized in the lab, we aim to identify whether any of the PC-specific factors selected are involved in the initiation or maintenance of L1 silencing. Follow-up experiments will analyze further mechanistic details of L1-silencing, depending on the nature of the identified factor.

IX. CONCLUSIONS

The main conclusions of this Thesis are:

1.1. Cellular RNase H2 activity is strictly required for the mobilization of LINE elements lacking a functional RNase H domain (human L1Hs and zebrafish Zfl2 2), in a cell line independent manner, being dispensable for the mobilization of active LTR-retrotransposons that code their own RNase H domain, or for DNA-Transposons.

1.2. RNASEH2A inactivating mutations characterized in Aicardi Goutières Syndrome patients would prevent L1 retrotransposition, at least using complementation assays with RNase H2 KO cells and *in vitro* retrotransposition assays. In Aicardi Goutières Syndrome patients, lack of RNase H2 activity might result in the accumulation of abortive DNA:RNA LINE-1 hybrids, which could activate the immune system in response to the accumulation of these L1-derived endogenous nucleic acids.

1.3. RNase H2 facilitates LINE-1 retrotransposition by potentially removing the RNA from L1-RNA:cDNA hybrids, acting on "Y-like" structures generated during initial steps of the retrotransposition process.

2.1. Pluripotent human PA-1 cells express a constellation of L1-RNPs, differing in the composition of L1-ORF1p interactomes, as revealed by affinity capture experiments using anti-human_L1-ORF1p mouse monoclonal antibodies and llama-derived nanobodies.

2.2. The interactome of L1-ORF1p on human pluripotent PA-1 cells revealed that most L1-interactors (92 %, 164 from a total of 178 significant interactors identified in PC PA-1 cells) haven't been found in previous L1 proteomic studies, conducted with ectopically expressed L1s and transformed cell lines (HEK293T and HeLa).

2.3. The characterization of host factors interacting with endogenously expressed L1-RNPs using pluripotent and differentiated cells revealed that the L1 interactome is quite dynamic, and on average >85% of pluripotent-specific L1-interactors change with differentiation (156 out of 178 PC PA-1 specific interactors).

2.4. LINE-1 retrotransposition intermediates (i.e., L1-RNPs) interact with a large number of ZNF proteins specifically in pluripotent cells.

Other minor conclusions:

- RNase H2 Knock Out (KO) cells massively miss-incorporate ribonucleotides in genomic DNA, inducing DNA damage. However, ENdonuclease-Independent (ENi) L1 retrotransposition doesn't occur at a high frequency in these DNA damaged sites.
- Endogenous LINE-1 expression levels are not influenced by cellular RNase H2 activity, at least using HeLa cells as a cellular model.
- In the absence of RNase H2, cellular RNase H1 can partially facilitate L1 retrotransposition, at least on complementation assays of RNase H2 KO cells.
- Using transient complementation and a Separation of Function RNASEH2A mutant subunit, we found that the L1-RNA has to be completely degraded from RNA:cDNA hybrids to facilitate retrotransposition.
- Lack of cellular RNase H2 activity doesn't increase the mutation rate at dinucleotide repeats present in L1-retrotranscribed sequences, at least on HeLa cells.
- Wild type human LINE-1s are severely compromised for retrotransposition in RNase H2 KO cells; in contrast, LINE-1s containing inactivating mutations in the PCNA interacting protein motif (PIP) of L1-ORF2p, which abrogate interaction with PCNA, can retrotranspose in RNase H2 KO cell lines, although at severely reduced levels.
- Direct and indirect experimental evidence indicate that the interaction of L1-ORF2p with PCNA might facilitate a functional interaction with the RNase H2 complex, as the B subunit also contains a functional PIP motif. In contrast, a potential functional interaction with RNase H1 during retrotransposition would occur by a much less efficient diffusion mechanism, in part due to lack of PIP motifs in RNase H1.
- At difference with other AGS patients, those carrying homozygous inactivating mutations in RNase H2, wouldn't be characterized for accommodating higher L1 retrotransposition levels in their genomes.
- Similarly, the interactome of L1-ORF1p on isogenic differentiated cells also revealed that most L1-interactors (87.6 %, 71 from 81 significant interactors identified in DC PA-1 cells) are also novel within L1 proteomic studies.
- 7.9%, and 12.3% of L1-interactors identified in PCs and DCs, respectively, have been found in previous proteomic studies, even if they used ectopically expressed L1s and transformed cell lines.
- The comparison of isogenic PC and DC interactomes revealed that a large number of L1-interactors are specific to either PCs or DCs (65.8% and 24.9% of the total identified interactors (237) are found only in PCs and DCs, respectively), further stressing how LINE-1s are differentially regulated in pluripotent and differentiated cells.
- Using PC and DC interactomes, and several criteria to rank factors, we generated a list of potential LINE-1 interactors that may be involved in L1-silencing. Most of

the 14 factors selected were specific to PC interactomes and could also interact with RNA:DNA hybrids.

X. CONCLUSIONES

-

Las principales conclusiones de esta Tesis son:

1.1. La actividad RNasa H2 celular es estrictamente necesaria para la movilización de elementos LINE que carecen de un dominio RNasa H funcional (L1Hs humano y Zfl2 2 de pez cebra), independientemente de la línea celular usada, siendo prescindible para la movilización de retrotransposones LTR activos que codifican su propio dominio RNasa H, o para los DNA-transposones.

1.2. Las mutaciones inactivantes en la subunidad RNASEH2A caracterizadas en pacientes con el síndrome de Aicardi Goutières impedirían la retrotransposición de L1, de acuerdo a los ensayos de complementación con células RNasa H2 KO y ensayos de retrotransposición *in vitro*. En los pacientes con el síndrome de Aicardi Goutières, la falta de RNasa H2 podría dar lugar a la acumulación de híbridos L1 ADN:ARN, lo que podría activar el sistema inmunitario en respuesta a la acumulación de estos ácidos nucleicos endógenos derivados de L1.

1.3. La enzima RNasa H2 facilita la retrotransposición de LINE-1, potencialmente eliminando el ARN de los híbridos ARN:ADNc que se forman en las estructuras "tipo Y" generadas durante los pasos iniciales del proceso de retrotransposición.

2.1. Las células pluripotentes humanas PA-1 expresan una constelación de intermediarios de la retrotransposición de L1 o L1-RNPs, que difieren en la composición de sus interactomas, tal y como revelan los experimentos de captura por afinidad utilizando anticuerpos anti-L1-ORF1p monoclonales de ratón así como nanoanticuerpos derivados de llama.

2.2. El interactoma de L1-ORF1p en células pluripotentes humanas PA-1 reveló que la mayoría de los interactores de L1 (92 %, 164 de un total de 178 interactores significativos identificados en las células PC PA-1) no se han encontrado en estudios proteómicos previos de L1 [realizados con L1s ectópicos y líneas celulares transformadas (HEK293T y HeLa)].

2.3. La caracterización de factores celulares que interactúan con L1-RNPs endógenas en células pluripotentes y diferenciadas, reveló que el interactoma de L1 es bastante dinámico, y en promedio >85% de los interactores específicos de células pluripotentes cambian con la diferenciación celular (156 de 178 interactores específicos de PC PA-1).

2.4. Los intermediarios de la retrotransposición de LINE-1 (es decir, L1-RNPs) interactúan con un gran número de proteínas ZNF, específicamente en células pluripotentes.

Conclusiones menores adicionales:

- Las células Knock Out (KO) para RNasa H2 incorporan masivamente ribonucleótidos en sus genomas, induciendo daño en el ADN. Sin embargo, la retrotransposición de L1 independiente de su endonucleasa (ENi) no ocurre con una alta frecuencia en estos sitios dañados del ADN.
- La actividad enzimática humana RNasa H2 no afecta los niveles de expresión de LINE-1s endógenos, al menos utilizando células HeLa como modelo celular.
- En ausencia de RNasa H2, la enzima humana RNasa H1 puede facilitar parcialmente la retrotransposición de L1, al menos en ensayos de complementación de células RNasa H2 KO.
- Utilizando complementación transitoria y un mutante RNASEH2A de separación de función, encontramos que el ARN de L1 tiene que ser completamente degradado de los híbridos ARN:ADNc para facilitar la retrotransposición.
- La ausencia de RNasa H2 no aumenta la tasa de mutación en repeticiones de dinucleótidos presentes en secuencias retrotranscritas por L1, al menos en células HeLa.
- Los LINE-1 humanos de tipo *wild type* no pueden retrotransponer en células KO RNasa H2; en cambio, elementos LINE-1s que contienen mutaciones inactivantes en el dominio de interacción con PCNA (PIP) localizado en L1-ORF2p, y que eliminan la interacción con PCNA, pueden retrotransponer en células RNasa H2 KO, aunque a niveles bajos.
- Las evidencias experimentales directas e indirectas indican que la interacción de L1-ORF2p con PCNA podría facilitar una interacción funcional con el complejo RNasa H2, ya que la subunidad B también contiene un motivo PIP funcional. En cambio, una posible interacción funcional con la RNasa H1 durante la retrotransposición se produciría por un mecanismo de difusión mucho menos eficiente, en parte debido a la falta de motivos PIP en RNasa H1.
- A diferencia de otros pacientes con AGS, los portadores de mutaciones homocigóticas inactivantes en el gen de la RNasa H2 no se caracterizarían por acomodar mayores niveles de retrotransposición de L1 en sus genomas.
- Del mismo modo, el interactoma de L1-ORF1p en células diferenciadas isogénicas reveló que la mayoría de los L1-interactores (87,6%, 71 de los 81 interactores significativos identificados en las células DC PA-1) también son novedosos dentro de los estudios proteómicos de L1.
- El 7,9%, y el 12,3% de los interactores de L1 identificados en PCs y DCs, respectivamente, han sido encontrados en estudios proteómicos anteriores, pese a usar L1s ectópicos y líneas celulares transformadas (HEK293T y HeLa).
- La comparación de los interactomas isogénicos de PC y DC reveló que la mayoría de los interactores de L1 son específicos de PC o DC (el 65,8% y el 24,9% del total de interactores identificados (237) se encuentran sólo en PC y DC,

respectivamente), siendo estas nuevas evidencias experimentales que demuestran la existencia de una regulación específica de LINE-1 en células pluripotentes.

- Utilizando los interactomas de PC y DC, y varios criterios para clasificar los interactores, hemos generado una lista con factores que podrían estar implicados en el silenciamiento de L1. La mayoría de los 14 factores seleccionados son específicos de los interactomas de PCs, y pueden además interactuar con híbridos de ARN:ADN.

BIBLIOGRAPHY

Ali M, Highet LJ, Lacombe D, Goizet C, King MD, Tacke U, Van Der Knaap MS, Lagae L, Rittey C, Brunner HG, Van Bokhoven H, Hamel B, Oade YA, Sanchis A, Desguerre I, Cau D, Mathieu N, Moutard ML, Lebon P, Kumar D, et al (2006) A second locus for Aicardi-Goutières syndrome at chromosome 13q14-21. *J. Med. Genet.* **43:** 444–450 Available at: https://jmg.bmj.com/lookup/doi/10.1136/jmg.2005.031880

Alisch RS (2006) Unconventional translation of mammalian LINE-1 retrotransposons. *Genes Dev.* **20:** 210–224 Available at: http://www.genesdev.org/cgi/doi/10.1101/gad.1380406

An W, Dai L, Niewiadomska A, Yetil A, O'Donnell KA, Han JS & Boeke JD (2011) Characterization of a synthetic human LINE-1 retrotransposon ORFeus-Hs. *Mob. DNA* **2:** 2 Available at: http://www.mobilednajournal.com/content/2/1/2

Andrews PW, Matin MM, Bahrami AR, Damjanov I, Gokhale P & Draper JS (2005) Embryonic stem (ES) cells and embryonal carcinoma (EC) cells: opposite sides of the same coin. *Biochem. Soc. Trans.* **33:** 1526 Available at: http://www.biochemsoctrans.org/bst/033/bst0331526.htm

Aravin AA, Sachidanandam R, Bourc'his D, Schaefer C, Pezic D, Toth KF, Bestor T & Hannon GJ (2008) A piRNA Pathway Primed by Individual Transposons Is Linked to De Novo DNA Methylation in Mice. *Mol. Cell* **31:** 785–799 Available at: https://linkinghub.elsevier.com/retrieve/pii/S1097276508006199

Baillie JK, Barnett MW, Upton KR, Gerhardt DJ, Richmond TA, De Sapio F, Brennan P, Rizzu P, Smith S, Fell M, Talbot RT, Gustincich S, Freeman TC, Mattick JS, Hume DA, Heutink P, Carninci P, Jeddeloh JA & Faulkner GJ (2011) Somatic retrotransposition alters the genetic landscape of the human brain. *Nature* **479:** 534–537 Available at: http://dx.doi.org/10.1038/nature10531

Barau J, Teissandier A, Zamudio N, Roy S, Nalesso V, Hérault Y, Guillou F & Bourc'his D (2016) The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. *Science (80-.).* **354:** 909–912 Available at: https://www.sciencemag.org/lookup/doi/10.1126/science.aah5143

Bartsch K, Knittler K, Borowski C, Rudnik S, Damme M, Aden K, Spehlmann ME, Frey N, Saftig P, Chalaris A & Rabe B (2017) Absence of RNase H2 triggers generation of immunogenic micronuclei removed by autophagy. *Hum. Mol. Genet.* **26:** 3960–3972 Available at: http://academic.oup.com/hmg/article/26/20/3960/3976572/Absence-of-RNase-H2-triggers-generation-of

Beck CR, Collier P, Macfarlane C, Malig M, Kidd JM, Eichler EE, Badge RM & Moran J V. (2010) LINE-1 Retrotransposition Activity in Human Genomes. *Cell* **141:** 1159–1170 Available at: https://linkinghub.elsevier.com/retrieve/pii/S009286741000557X

Beck CR, Garcia-Perez JL, Badge RM & Moran J V. (2011) LINE-1 Elements in Structural Variation and Disease. *Annu. Rev. Genomics Hum. Genet.* **12:** 187–215 Available at: http://www.annualreviews.org/doi/10.1146/annurev-genom-082509-141802

Belancio VP, Hedges DJ & Deininger P (2008) Mammalian non-LTR retrotransposons: For better or worse, in sickness and in health. *Genome Res.* **18:** 343–358 Available at: http://www.genome.org/cgi/doi/10.1101/gr.5558208

Belfort, Marlene; Derbyshire, Victoria; Parker, Monica M.; Cousineau, Benoit; Lambowitz AM (2002) Mobile Introns: Pathways and Proteins. In *Mobile DNA II* pp 761–783. American Society of Microbiology Available at: http://www.asmscience.org/content/book/10.1128/9781555817954.chap31

Benabdellah K, Cobo M, Muñoz P, Toscano MG & Martin F (2011) Development of an All-in-One Lentiviral Vector System Based on the Original TetR for the Easy Generation of Tet-ON Cell Lines. *PLoS One* **6:** e23734 Available at: https://dx.plos.org/10.1371/journal.pone.0023734

Benitez-Guijarro M, Lopez-Ruiz C, Tarnauskaitė Ž, Murina O, Mian Mohammad M, Williams TC, Fluteau A, Sanchez L, Vilar-Astasio R, Garcia-Canadas M, Cano D, Kempen MH, Sanchez-Pozo A, Heras SR, Jackson AP, Reijns MA & Garcia-Perez JL (2018) RNase H2, mutated in Aicardi-Goutières syndrome, promotes LINE-1 retrotransposition. *EMBO J.* **37:** e98506 Available at: http://emboj.embopress.org/lookup/doi/10.15252/embj.201798506

Biessmann H, Valgeirsdottir K, Lofsky A, Chin C, Ginther B, Levis RW & Pardue ML (1992) HeT-A, a transposable element specifically involved in 'healing' broken chromosome ends in Drosophila melanogaster. *Mol. Cell. Biol.* **12:** 3910–3918 Available at: http://mcb.asm.org/lookup/doi/10.1128/MCB.12.9.3910

Blaise S, de Parseval N, Benit L & Heidmann T (2003) Genomewide screening for fusogenic human endogenous retrovirus envelopes identifies syncytin 2, a gene conserved on primate evolution. *Proc. Natl. Acad. Sci.* **100:** 13013–13018 Available at: http://www.pnas.org/cgi/doi/10.1073/pnas.2132646100

Bodak M, Cirera-Salinas D, Yu J, Ngondo RP & Ciaudo C (2017) Dicer, a new regulator of pluripotency exit and LINE-1 elements in mouse embryonic stem cells. *FEBS Open Bio* **7:** 204–220 Available at: http://doi.wiley.com/10.1002/2211-5463.12174

Boeke JD, Garfinkel DJ, Styles CA & Fink GR (1985) Ty elements transpose through an RNA intermediate. *Cell* **40:** 491–500 Available at: https://linkinghub.elsevier.com/retrieve/pii/0092867485901977

Bogerd HP (2006) APOBEC3A and APOBEC3B are potent inhibitors of LTR-retrotransposon function in human cells. *Nucleic Acids Res.* **34:** 89–95 Available at: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkj416

Bogerd HP, Wiegand HL, Hulme AE, Garcia-Perez JL, O'Shea KS, Moran J V. & Cullen BR (2006) Cellular inhibitors of long interspersed element 1 and Alu retrotransposition. *Proc. Natl. Acad. Sci.* **103:** 8780–8785 Available at: http://www.pnas.org/cgi/doi/10.1073/pnas.0603313103

Bourc'his D & Bestor TH (2004) Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature* **431:** 96–99 Available at: http://www.nature.com/articles/nature02886

Bourque G, Burns KH, Gehring M, Gorbunova V, Seluanov A, Hammell M, Imbeault M, Izsvák Z, Levin HL, Macfarlan TS, Mager DL & Feschotte C (2018) Ten things you should know about transposable elements. *Genome Biol.* **19:** 199 Available at: https://genomebiology.biomedcentral.com/articles/10.1186/s13059-018-1577-z

Branco MR, Ficz G & Reik W (2012) Uncovering the role of 5-hydroxymethylcytosine in the epigenome. *Nat. Rev. Genet.* **13:** 7–13 Available at: http://www.nature.com/articles/nrg3080

Brayer KJ, Kulshreshtha S & Segal DJ (2008) The protein-binding potential of C2H2 zinc finger domains. *Cell Biochem. Biophys.* **51:** 9–19

Brayer KJ & Segal DJ (2008) Keep your fingers off my DNA: Protein-protein interactions mediated by C2H2 zinc finger domains. *Cell Biochem. Biophys.* **50:** 111–131

Britten RJ & Kohne DE (1968) Repeated Sequences in DNA. *Science (80-.).* **161:** 529–540 Available at: https://www.sciencemag.org/lookup/doi/10.1126/science.161.3841.529

Brouha B, Schustak J, Badge RM, Lutz-Prigge S, Farley AH, Moran J V. & Kazazian HH (2003) Hot L1s account for the bulk of retrotransposition in the human population. *Proc. Natl. Acad. Sci.* **100:** 5280–5285 Available at: http://www.pnas.org/cgi/doi/10.1073/pnas.0831042100

Bubeck D, Reijns MAM, Graham SC, Astell KR, Jones EY & Jackson AP (2011) PCNA directs type 2 RNase H activity on DNA replication and repair substrates. *Nucleic Acids Res.* **39:** 3652–3666 Available at: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkq980

Bunz F (1998) Requirement for p53 and p21 to Sustain G2 Arrest After DNA Damage. *Science (80-.).* **282:** 1497–1501 Available at: https://www.sciencemag.org/lookup/doi/10.1126/science.282.5393.1497

Burke JM, Moon SL, Matheny T & Parker R (2019) RNase L Reprograms Translation by Widespread mRNA Turnover Escaped by Antiviral mRNAs. *Mol. Cell* **75:** 1203-1217.e5 Available at: https://linkinghub.elsevier.com/retrieve/pii/S1097276519305842

Burns KH (2017) Transposable elements in cancer. *Nat. Rev. Cancer* **17:** 415–424 Available at: http://www.nature.com/articles/nrc.2017.35

Castro-Diaz N, Ecco G, Coluccio A, Kapopoulou A, Yazdanpanah B, Friedli M, Duc J, Jang SM, Turelli P & Trono D (2014) Evolutionally dynamic L1 regulation in embryonic stem cells. *Genes Dev.* **28:** 1397–1409 Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4083085/pdf/1397.pdf

De Cecco M, Ito T, Petrashen AP, Elias AE, Skvir NJ, Criscione SW, Caligiana A, Brocculi G, Adney EM, Boeke JD, Le O, Beauséjour C, Ambati J, Ambati K, Simon M, Seluanov A, Gorbunova V, Slagboom PE, Helfand SL, Neretti N, et al (2019) L1 drives IFN in senescent cells and promotes age-associated inflammation. *Nature* **566:** 73–78 Available at: http://www.nature.com/articles/s41586-018-0784-9

Choi J, Hwang S-Y & Ahn K (2018) Interplay between RNASEH2 and MOV10 controls LINE-1 retrotransposition. *Nucleic Acids Res.* **46:** 1912–1926 Available at: http://academic.oup.com/nar/advance-article/doi/10.1093/nar/gkx1312/4791130

Chon H, Sparks JL, Rychlik M, Nowotny M, Burgers PM, Crouch RJ & Cerritelli SM (2013) RNase H2 roles in genome integrity revealed by unlinking its activities. *Nucleic Acids Res.* **41:** 3130–3143 Available at: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkt027

Chon H, Vassilev A, DePamphilis ML, Zhao Y, Zhang J, Burgers PM, Crouch RJ & Cerritelli SM (2009) Contributions of the two accessory subunits, RNASEH2B and RNASEH2C, to the activity and properties of the human RNase H2 complex. *Nucleic Acids Res.* **37:** 96–110 Available at: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkn913

Cost GJ, Feng Q, Jacquier A & Boeke JD (2002) Human L1 element target-primed reverse transcription in vitro. *EMBO J.* **21:** 5899–5910 Available at: http://emboj.embopress.org/cgi/doi/10.1093/emboj/cdf592

Coufal NG, Garcia-Perez JL, Peng GE, Marchetto MCN, Muotri AR, Mu Y, Carson CT, Macia A, Moran J V. & Gage FH (2011) Ataxia telangiectasia mutated (ATM) modulates long interspersed element-1 (L1) retrotransposition in human neural stem cells. *Proc. Natl. Acad. Sci. U. S. A.* **108:** 20382–20387 Available at: http://www.pnas.org/cgi/doi/10.1073/pnas.1100273108

Coufal NG, Garcia-Perez JL, Peng GE, Yeo GW, Mu Y, Lovci MT, Morell M, O'Shea KS, Moran J V. & Gage FH (2009) L1 retrotransposition in human neural progenitor cells. *Nature* **460:** 1127–1131 Available at: https://www.nature.com/articles/nature08248

Craig N (2002) Mobile DNA: an Introduction. In *Mobile DNA II. ASM Press.*, Craig N Craigie R Gellert M & Lambowitz A (eds) pp 3–11. Washington, DC: American Society of Microbiology Available at: http://www.asmscience.org/content/book/10.1128/9781555817954.chap1

Crichton JH, Dunican DS, MacLennan M, Meehan RR & Adams IR (2014) Defending the genome from the enemy within: mechanisms of retrotransposon suppression in the mouse germline. *Cell. Mol. Life Sci.* **71:** 1581–1605 Available at: http://link.springer.com/10.1007/s00018-013-1468-0

Cristea IM, Williams R, Chait BT & Rout MP (2005) Fluorescent proteins as proteomic probes. *Mol. Cell. Proteomics* **4:** 1933–1941 Available at: https://linkinghub.elsevier.com/retrieve/pii/S1535947620300293

Cristini A, Groh M, Kristiansen MS & Gromak N (2018) RNA/DNA Hybrid Interactome Identifies DXH9 as a Molecular Player in Transcriptional Termination and R-Loop-Associated DNA Damage. *Cell Rep.* **23:** 1891–1905 Available at: https://doi.org/10.1016/j.celrep.2018.04.025

Cristofari G, Ficheux D & Darlix J-L (2000) The Gag-like Protein of the Yeast Ty1 Retrotransposon Contains a Nucleic Acid Chaperone Domain Analogous to Retroviral Nucleocapsid Proteins. *J. Biol. Chem.* **275:** 19210–19217 Available at: http://www.jbc.org/lookup/doi/10.1074/jbc.M001371200

Crow YJ, Hayward BE, Parmar R, Robins P, Leitch A, Ali M, Black DN, Van Bokhoven H, Brunner HG, Hamel BC, Corry PC, Cowan FM, Frints SG, Klepper J, Livingston JH, Lynch SA, Massey RF, Meritet JF, Michaud JL, Ponsot G, et al (2006a) Mutations in the gene encoding the 3′-5′ DNA exonuclease TREX1 cause Aicardi-Goutières syndrome at the AGS1 locus. *Nat. Genet.* **38:** 917–920

Crow YJ, Leitch A, Hayward BE, Garner A, Parmar R, Griffith E, Ali M, Semple C, Aicardi J, Babul-Hirji R, Baumann C, Baxter P, Bertini E, Chandler KE, Chitayat D, Cau D, Déry C, Fazzi E, Goizet C, King MD, et al (2006b) Mutations in genes encoding ribonuclease H2 subunits cause Aicardi-Goutières syndrome and mimic congenital viral brain infection. *Nat. Genet.* **38:** 910–916 Available at: https://doi.org/10.1038/ng1842

Crow YJ & Manel N (2015) Aicardi-Goutières syndrome and the type I interferonopathies. *Nat. Rev. Immunol.* **15:** 429–40

Crow YJ & Rehwinkel J (2009) Aicardi-Goutieres syndrome and related phenotypes: linking nucleic acid metabolism with autoimmunity. *Hum. Mol. Genet.* **18:** R130–R136 Available at: https://academic.oup.com/hmg/article-lookup/doi/10.1093/hmg/ddp293

Curcio MJ & Belfort M (2007) The beginning of the end: Links between ancient retroelements and modern telomerases. *Proc. Natl. Acad. Sci.* **104:** 9107–9108 Available at: http://www.pnas.org/cgi/doi/10.1073/pnas.0703224104

Dai L, LaCava J, Taylor MS & Boeke JD (2014) Expression and detection of LINE-1 ORF-encoded proteins. *Mob. Genet. Elements* **4:** e29319 Available at: http://www.tandfonline.com/doi/abs/10.4161/mge.29319

Dai L, Taylor MS, O'Donnell KA & Boeke JD (2012) Poly(A) Binding Protein C1 Is Essential for Efficient L1 Retrotransposition and Affects L1 RNP Formation. *Mol. Cell. Biol.* **32:** 4323–4336 Available at: http://mcb.asm.org/cgi/doi/10.1128/MCB.06785-11

Dakhore S, Nayer B & Hasegawa K (2018) Human Pluripotent Stem Cell Culture: Current Status, Challenges, and Advancement. *Stem Cells Int.* **2018:** 1–17 Available at: https://www.hindawi.com/journals/sci/2018/7396905/

Denli AM, Narvaiza I, Kerman BE, Pena M, Benner C, Marchetto MCN, Diedrich JK, Aslanian A, Ma J, Moresco JJ, Moore L, Hunter T, Saghatelian A & Gage FH (2015) Primate-Specific ORF0 Contributes to Retrotransposon-Mediated Diversity. *Cell* **163:** 583–893

Derbyshire KM & Grindley NDF (1986) Replicative and conservative transposition in bacteria. *Cell* **47:** 325–327 Available at: https://linkinghub.elsevier.com/retrieve/pii/0092867486905866

Dewannieux M, Esnault C & Heidmann T (2003) LINE-mediated retrotransposition of marked Alu sequences. *Nat. Genet.* **35:** 41–48 Available at: http://www.nature.com/articles/ng1223

Dmitriev SE, Andreev DE, Terenin IM, Olovnikov IA, Prassolov VS, Merrick WC & Shatsky IN (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. *Mol. Cell. Biol.* **27:** 4685–4697 Available at: http://mcb.asm.org/cgi/doi/10.1128/MCB.02138-06 [Accessed November 23, 2018]

Domanski M, Molloy K, Jiang H, Chait B, Rout M, Jensen T & LaCava J (2012) Improved methodology for the affinity isolation of human protein complexes expressed at near endogenous levels. *Biotechniques* Available at: https://www.future-science.com/doi/10.2144/000113864

Doolittle RF, Johnson MS & McClure MA (1989) Origins and Evolutionary Relationships of Retroviruses. *Q. Rev. Biol.* **64:** 1–30 Available at: https://www.journals.uchicago.edu/doi/10.1086/416128

Dornan D, Eckert M, Wallace M, Shimizu H, Ramsay E, Hupp TR & Ball KL (2004) Interferon Regulatory Factor 1 Binding to p300 Stimulates DNA-Dependent Acetylation of p53. *Mol. Cell. Biol.* **24:** 10083–10098

Dou Y, Kalmykova S, Pashkova M, Oghbaie M, Jiang H, Molloy KR, Chait BT, Rout MP, Fenyö D, Jensen TH, Altukhov I & LaCava J (2020) Affinity proteomic dissection of the human nuclear cap-binding complex interactome. *Nucleic Acids Res.* **48:** 10456–10469 Available at: https://academic.oup.com/nar/article/48/18/10456/5909927

Doucet AJ, Hulme AE, Sahinovic E, Kulpa DA, Moldovan JB, Kopera HC, Athanikar JN, Hasnaoui M, Bucheton A, Moran J V. & Gilbert N (2010) Characterization of LINE-1 ribonucleoprotein particles. *PLoS Genet.* **6:** 1–19 Available at: https://dx.plos.org/10.1371/journal.pgen.1001150

Doucet AJ, Wilusz JE, Miyoshi T, Liu Y & Moran J V. (2015) A 3′ Poly(A) Tract Is Required for LINE-1 Retrotransposition. *Mol. Cell* **60:** 728–741 Available at: https://linkinghub.elsevier.com/retrieve/pii/S1097276515007790

Ecco G, Imbeault M & Trono D (2017) KRAB zinc finger proteins. *Dev.* **144:** 2719–2729 Available at: https://journals.biologists.com/dev/article/144/15/2719/48057/KRAB-zinc-finger-proteins

Eickbush TH (1997) MOLECULAR BIOLOGY: Telomerase and Retrotransposons: Which Came First? *Science (80-.).* **277:** 911–912 Available at: https://www.sciencemag.org/lookup/doi/10.1126/science.277.5328.911

Engels WR & Preston CR (1981) Identifying P factors in Drosophila by means of chromosome breakage hotspots. *Cell* **26:** 421–428 Available at: https://linkinghub.elsevier.com/retrieve/pii/0092867481902117

Ergün S, Buschmann C, Heukeshoven J, Dammann K, Schnieders F, Lauke H, Chalajour F, Kilic N, Strätling WH & Schumann GG (2004) Cell Type-specific Expression of LINE-1 Open Reading Frames 1 and 2 in Fetal and Adult Human Tissues. *J. Biol. Chem.* **279:** 27753–27763 Available at: http://www.jbc.org/lookup/doi/10.1074/jbc.M312985200

Esnault C (2002) A Tetrahymena thermophila ribozyme-based indicator gene to detect transposition of marked retroelements in mammalian cells. *Nucleic Acids Res.* **30:** 49e – 49 Available at: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/30.11.e49

Esnault C, Maestre J & Heidmann T (2000) Human LINE retrotransposons generate processed pseudogenes. *Nat. Genet.* **24:** 363–367 Available at: http://www.nature.com/articles/ng0400_363

Faulkner GJ, Kimura Y, Daub CO, Wani S, Plessy C, Irvine KM, Schroder K, Cloonan N, Steptoe AL, Lassmann T, Waki K, Hornig N, Arakawa T, Takahashi H, Kawai J, Forrest ARR, Suzuki H, Hayashizaki Y, Hume DA, Orlando V, et al (2009) The regulated retrotransposon transcriptome of mammalian cells. *Nat. Genet.* **41:** 563–571 Available at: https://www.nature.com/articles/ng.368

Fedotova AA, Bonchuk AN, Mogila VA & Georgiev PG (2017) C2H2 zinc finger proteins: The largest but poorly explored family of higher eukaryotic transcription factors. *Acta Naturae* **9:** 47–58

Feng Q, Moran J V., Kazazian HH & Boeke JD (1996) Human L1 Retrotransposon Encodes a Conserved Endonuclease Required for Retrotransposition. *Cell* **87:** 905–916 Available at: https://linkinghub.elsevier.com/retrieve/pii/S0092867400819972

Feusier J, Watkins WS, Thomas J, Farrell A, Witherspoon DJ, Baird L, Ha H, Xing J & Jorde LB (2019) Pedigree-based estimation of human mobile element retrotransposition rates. *Genome Res.* **29:** 1567–1577 Available at: http://genome.cshlp.org/lookup/doi/10.1101/gr.247965.118

Flasch DA, Macia Á, Sánchez L, Ljungman M, Heras SR, García-Pérez JL, Wilson TE & Moran J V. (2019) Genome-wide de novo L1 Retrotransposition Connects Endonuclease Activity with Replication. *Cell* **177:** 837-851.e28 Available at: https://linkinghub.elsevier.com/retrieve/pii/S0092867419302338

Freeman JD, Goodchild NL & Mager DL (1994) A modified indicator gene for selection of retrotransposition events in mammalian cells. *Biotechniques* **17:** 46, 48–9, 52 Available at: http://www.ncbi.nlm.nih.gov/pubmed/7946311

Freeman P, Macfarlane C, Collier P, Jeffreys AJ & Badge RM (2011) L1 hybridization enrichment: a method for directly accessing de novo L1 insertions in the human germline. *Hum. Mutat.* **32:** 978–988 Available at: https://onlinelibrary.wiley.com/doi/10.1002/humu.21533

Garcia-perez JL, Doucet AJ, Bucheton A, Garcia-perez JL, Doucet AJ, Bucheton A, Moran J V & Gilbert N (2007) Distinct mechanisms for trans -mediated mobilization of cellular RNAs by the LINE-1 reverse transcriptase Distinct mechanisms for trans -mediated mobilization of cellular RNAs by the LINE-1 reverse transcriptase. **:** 602–611

Garcia-Perez JL, Marchetto MCN, Muotri AR, Coufal NG, Gage FH, O'Shea KS & Moran J V. (2007) LINE-1 retrotransposition in human embryonic stem cells. *Hum. Mol. Genet.* **16:** 1569–1577

Garcia-Perez JL, Morell M, Scheys JO, Kulpa DA, Morell S, Carter CC, Hammer GD, Collins KL, O'Shea KS, Menendez P & Moran J V. (2010) Epigenetic silencing of engineered L1 retrotransposition events in human embryonic carcinoma cells. *Nature* **466:** 769–773 Available at: http://dx.doi.org/10.1038/nature09209

Garcia-Perez JL, Widmann TJ & Adams IR (2016) The impact of transposable elements on mammalian development. *Development* **143:** 4101–4114 Available at: https://dev.biologists.org/content/143/22/4101.long

Garcia Pérez JL & Alarcón-Riquelme ME (2017) The TREX1 Dinosaur Bites the Brain through the LINE. *Cell Stem Cell* **21:** 287–288 Available at: https://linkinghub.elsevier.com/retrieve/pii/S1934590917303284

Garfinkel DJ, Boeke JD & Fink GR (1985) Ty element transposition: Reverse transcriptase and virus-like particles. *Cell* **42:** 507–517 Available at: https://linkinghub.elsevier.com/retrieve/pii/0092867485901084

Gazquez-Gutierrez A, Witteveldt J, R. Heras S & Macias S (2021) Sensing of transposable elements by the antiviral innate immune system. *RNA*: rna.078721.121 Available at: http://rnajournal.cshlp.org/lookup/doi/10.1261/rna.078721.121

Genovesio A, Kwon Y-J, Windisch MP, Kim NY, Choi SY, Kim HC, Jung S, Mammano F, Perrin V, Boese AS, Casartelli N, Schwartz O, Nehrbass U & Emans N (2011) Automated Genome-Wide Visual Profiling of Cellular Proteins Involved in HIV Infection. *J. Biomol. Screen.* **16:** 945–958 Available at: http://journals.sagepub.com/doi/10.1177/1087057111415521

Geurts AM, Yang Y, Clark KJ, Liu G, Cui Z, Dupuy AJ, Bell JB, Largaespada DA & Hackett PB (2003) Gene transfer into genomes of human cells by the sleeping beauty transposon system. *Mol. Ther.* **8:** 108–117 Available at: http://dx.doi.org/10.1016/S1525-0016(03)00099-6

Gilbert N, Lutz-Prigge S & Moran J V. (2002) Genomic deletions created upon LINE-1 retrotransposition. *Cell* **110:** 315–325

Ginno PA, Lim YW, Lott PL, Korf I & Chédin F (2013) GC skew at the 5' and 3' ends of human genes links R-loop formation to epigenetic regulation and transcription termination. TL - 23. *Genome Res.* **23 VN-r:** 1590–1600 Available at: /Users/irishelenejonkers/Documents/ReadCube Media/Genome Research 2013 Ginno-1.pdf%5Cnhttp://dx.doi.org/10.1101/gr.158436.113

Goodier JL, Cheung LE & Kazazian HH (2012) MOV10 RNA Helicase Is a Potent Inhibitor of Retrotransposition in Cells. *PLoS Genet.* **8:** e1002941 Available at: https://dx.plos.org/10.1371/journal.pgen.1002941

Goodier JL, Cheung LE & Kazazian HH (2013) Mapping the LINE1 ORF1 protein interactome reveals associated inhibitors of human retrotransposition. *Nucleic Acids Res.* **41:** 7401–7419

Goodier JL & Kazazian HH (2008) Retrotransposons Revisited: The Restraint and Rehabilitation of Parasites. *Cell* **135:** 23–35 Available at: https://linkinghub.elsevier.com/retrieve/pii/S0092867408011793

Goodier JL, Mandal PK, Zhang L & Kazazian HH (2010) Discrete subcellular partitioning of human retrotransposon RNAs despite a common mechanism of genome insertion. *Hum. Mol. Genet.* **19:** 1712–1725 Available at: https://academic.oup.com/hmg/article-lookup/doi/10.1093/hmg/ddq048

Goodier JL, Pereira GC, Cheung LE, Rose RJ & Kazazian HH (2015) The Broad-Spectrum Antiviral Protein ZAP Restricts Human Retrotransposition. *PLoS Genet.* **11:** e1005252 Available at: https://dx.plos.org/10.1371/journal.pgen.1005252

Goodier JL, Zhang L, Vetter MR & Kazazian HH (2007) LINE-1 ORF1 Protein Localizes in Stress Granules with Other RNA-Binding Proteins, Including Components of RNA Interference RNA-Induced Silencing Complex. *Mol. Cell. Biol.* **27:** 6469–6483 Available at: https://journals.asm.org/doi/10.1128/MCB.00332-07

Goodwin TJD, Ormandy JE & Poulter RTM (2001) L1-like non-LTR retrotransposons in the yeast Candida albicans. *Curr. Genet.* **39:** 83–91 Available at: http://link.springer.com/10.1007/s002940000181

Gudas LJ & Wagner JA (2011) Retinoids regulate stem cell differentiation. *J. Cell. Physiol.* **226:** 322–330 Available at: http://doi.wiley.com/10.1002/jcp.22417

Hadley W, Romain F, Lionel H & Kirill M (2018) dplyr: A Grammar of Data Manipulation. Available at: https://cran.r-project.org/package=dplyr

Hamdorf M, Idica A, Zisoulis DG, Gamelin L, Martin C, Sanders KJ & Pedersen IM (2015) MiR-128 represses L1 retrotransposition by binding directly to L1 RNA. *Nat. Struct. Mol. Biol.* **22:** 824–831 Available at: http://www.nature.com/articles/nsmb.3090

Hancks DC, Goodier JL, Mandal PK, Cheung LE & Kazazian HH (2011) Retrotransposition of marked SVA elements by human L1s in cultured cells. *Hum. Mol. Genet.* **20:** 3386–3400 Available at: https://academic.oup.com/hmg/article-lookup/doi/10.1093/hmg/ddr245

Hancks DC & Kazazian HH (2012) Active human retrotransposons: variation and disease. *Curr. Opin. Genet. Dev.* **22:** 191–203 Available at: http://www.ncbi.nlm.nih.gov/pubmed/22406018

Hancks DC & Kazazian HH (2016) Roles for retrotransposon insertions in human disease. *Mob. DNA* **7:** 9 Available at: http://dx.doi.org/10.1186/s13100-016-0065-9

Harmsen MM & De Haard HJ (2007) Properties, production, and applications of camelid single-domain antibody fragments. *Appl. Microbiol. Biotechnol.* **77:** 13–22 Available at: http://link.springer.com/10.1007/s00253-007-1142-2

Heberle H, Meirelles GV, da Silva FR, Telles GP & Minghim R (2015) InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. *BMC Bioinformatics* **16:** 169 Available at: http://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-015-0611-3

Helman E, Lawrence MS, Stewart C, Sougnez C, Getz G & Meyerson M (2014) Somatic retrotransposition in human cancer revealed by whole-genome and exome sequencing. *Genome Res.* **24:** 1053–1063 Available at: http://genome.cshlp.org/cgi/doi/10.1101/gr.163659.113

Heras SR, Macias S, Cáceres JF & Garcia-Perez JL (2014) Control of mammalian retrotransposons by cellular RNA processing activities. *Mob. Genet. Elements* **4:** e28439 Available at: http://www.tandfonline.com/doi/abs/10.4161/mge.28439

Heras SR, Macias S, Plass M, Fernandez N, Cano D, Eyras E, Garcia-Perez JL & Cáceres JF (2013) The Microprocessor controls the activity of mammalian retrotransposons. *Nat. Struct. Mol. Biol.* **20:** 1173–1181 Available at: http://www.nature.com/articles/nsmb.2658

Hiller B, Achleitner M, Glage S, Naumann R, Behrendt R & Roers A (2012) Mammalian RNase H2 removes ribonucleotides from DNA to maintain genome integrity. *J. Exp. Med.* **209:** 1419–1426 Available at: https://rupress.org/jem/article/209/8/1419/53763/Mammalian-RNase-H2-removes-ribonucleotides-from

Hohjoh H & Singer MF (1997) Sequence-specific single-strand RNA binding protein encoded by the human LINE-1 retrotransposon. *EMBO J.* **16:** 6034–6043 Available at: https://www.embopress.org/doi/full/10.1093/emboj/16.19.6034

Holmes SE, Singer MF & Swergold GD (1992) Studies on p40, the leucine zipper motif-containing protein encoded by the first open reading frame of an active human LINE-1 transposable element. *J. Biol. Chem.* **267:** 19765–19768 Available at: https://www.jbc.org/article/S0021-9258(19)88618-0/pdf

Howard G, Eiges R, Gaudet F, Jaenisch R & Eden A (2008) Activation and transposition of endogenous retroviral elements in hypomethylation induced tumors in mice. *Oncogene* **27:** 404–408 Available at: http://www.nature.com/articles/1210631

Howe K, Clark MD, Torroja CF, Torrance J, Berthelot C, Muffato M, Collins JE, Humphray S, McLaren K, Matthews L, McLaren S, Sealy I, Caccamo M, Churcher C, Scott C, Barrett JC, Koch R, Rauch G-J, White S, Chow W, et al (2013) The zebrafish reference genome sequence and its relationship to the human genome. *Nature* **496:** 498–503 Available at: http://www.nature.com/articles/nature12111

Huira C. Kopera Doucet J & Moran J V (2016) LEAP: L1 Element Amplification Protocol. **1400:** 339–355 Available at: http://link.springer.com/10.1007/978-1-4939-3372-3

Van den Hurk JAJM, Meij IC, del Carmen Seleme M, Kano H, Nikopoulos K, Hoefsloot LH, Sistermans EA, de Wijs IJ, Mukhopadhyay A, Plomp AS, de Jong PTVM, Kazazian HH & Cremers FPM (2007) L1 retrotransposition can occur early in human embryonic development. *Hum. Mol. Genet.* **16:** 1587–1592 Available at: http://academic.oup.com/hmg/article/16/13/1587/2356012/L1-retrotransposition-can-occur-early-in-human

Huttlin EL, Bruckner RJ, Navarrete-Perea J, Cannon JR, Baltier K, Gebreab F, Gygi MP, Thornock A, Zarraga G, Tam S, Szpyt J, Gassaway BM, Panov A, Parzen H, Fu S, Golbazi A, Maenpaa E, Stricker K, Guha Thakurta S, Zhang T, et al (2021) Dual proteome-scale networks reveal cell-specific remodeling of the human interactome. *Cell* **184:** 3022-3040.e28 Available at: https://linkinghub.elsevier.com/retrieve/pii/S0092867421004463

Huttlin EL, Bruckner RJ, Paulo JA, Cannon JR, Ting L, Baltier K, Colby G, Gebreab F, Gygi MP, Parzen H, Szpyt J, Tam S, Zarraga G, Pontano-Vaites L, Swarup S, White AE, Schweppe DK, Rad R, Erickson BK, Obar RA, et al (2017) Architecture of the human interactome defines protein communities and disease networks. *Nature* **545:** 505–509 Available at: http://www.nature.com/articles/nature22366

Huttlin EL, Ting L, Bruckner RJ, Gebreab F, Gygi MP, Szpyt J, Tam S, Zarraga G, Colby G, Baltier K, Dong R, Guarani V, Vaites LP, Ordureau A, Rad R, Erickson BK, Wühr M, Chick J, Zhai B, Kolippakkam D, et al (2015) The BioPlex Network: A Systematic Exploration of the Human Interactome. *Cell* **162:** 425–440 Available at: https://linkinghub.elsevier.com/retrieve/pii/S0092867415007680

Imbeault M, Helleboid PY & Trono D (2017) KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. *Nature* **543:** 550–554 Available at: http://www.nature.com/articles/nature21683

Inoue H, Nojima H & Okayama H (1990) High efficiency transformation of Escherichia coli with plasmids. *Gene* **96:** 23–28 Available at: https://linkinghub.elsevier.com/retrieve/pii/037811199090336P

Iskow RC, McCabe MT, Mills RE, Torene S, Pittard WS, Neuwald AF, Van Meir EG, Vertino PM & Devine SE (2010) Natural mutagenesis of human genomes by endogenous retrotransposons. *Cell* **141:** 1253–1261 Available at: http://dx.doi.org/10.1016/j.cell.2010.05.020

Ivashkiv LB & Donlin LT (2014) Regulation of type i interferon responses. *Nat. Rev. Immunol.* **14:** 36–49 Available at: http://www.nature.com/articles/nri3581

Ivics Z, Hackett PB, Plasterk RH & Izsvák Z (1997) Molecular Reconstruction of Sleeping Beauty, a Tc1-like Transposon from Fish, and Its Transposition in Human Cells. *Cell* **91:** 501–510 Available at: https://linkinghub.elsevier.com/retrieve/pii/S0092867400804365

Jiang N, Feschotte C, Zhang X & Wessler SR (2004) Using rice to understand the origin and amplification of miniature inverted repeat transposable elements (MITEs). *Curr. Opin. Plant Biol.* **7:** 115–119 Available at: https://linkinghub.elsevier.com/retrieve/pii/S136952660400007X

Jönsson ME, Ludvik Brattås P, Gustafsson C, Petri R, Yudovich D, Pircs K, Verschuere S, Madsen S, Hansson J, Larsson J, Månsson R, Meissner A & Jakobsson J (2019) Activation of neuronal genes via LINE-1 elements upon global DNA demethylation in human neural progenitors. *Nat. Commun.* **10:** 3182 Available at: http://www.nature.com/articles/s41467-019-11150-8

Jurka J (1997) Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons. *Proc. Natl. Acad. Sci.* **94:** 1872–1877 Available at: http://www.pnas.org/cgi/doi/10.1073/pnas.94.5.1872

Kano H, Godoy I, Courtney C, Vetter MR, Gerton GL, Ostertag EM & Kazazian HH (2009) L1 retrotransposition occurs mainly in embryogenesis and creates somatic mosaicism. *Genes Dev.* **23:** 1303–1312 Available at: http://genesdev.cshlp.org/cgi/doi/10.1101/gad.1803909

Kapitonov V V. & Jurka J (2005) RAG1 Core and V(D)J Recombination Signal Sequences Were Derived from Transib Transposons. *PLoS Biol.* **3:** e181 Available at: https://dx.plos.org/10.1371/journal.pbio.0030181

Kazazian HH (2004) Mobile Elements: Drivers of Genome Evolution. *Science (80-.).* **303:** 1626–1632 Available at: https://www.sciencemag.org/lookup/doi/10.1126/science.1089670

Kazazian HH & Moran J V. (1998) The impact of L1 retrotransposons on the human genome. *Nat. Genet.* **19:** 19–24 Available at: http://www.nature.com/articles/ng0598-19

Kelleher CD & Champoux JJ (1998) Characterization of RNA Strand Displacement Synthesis by Moloney Murine Leukemia Virus Reverse Transcriptase. *J. Biol. Chem.* **273:** 9976–9986 Available at: http://dx.doi.org/10.1074/jbc.273.16.9976

Kennedy EM, Amie SM, Bambara RA & Kim B (2012) Frequent Incorporation of Ribonucleotides during HIV-1 Reverse Transcription and Their Attenuated Repair in Macrophages*. *J. Biol. Chem.* **287:** 14280–14288 Available at: https://linkinghub.elsevier.com/retrieve/pii/S0021925820529979

Keskin H, Shen Y, Huang F, Patel M, Yang T, Ashley K, Mazin A V. & Storici F (2014) Transcript-RNA-templated DNA recombination and repair. *Nature* **515:** 436–439 Available at: http://www.nature.com/articles/nature13682

Khazina E & Weichenrieder O (2009) Non-LTR retrotransposons encode noncanonical RRM domains in their first open reading frame. *Proc. Natl. Acad. Sci.* **106:** 731–736 Available at: http://www.pnas.org/lookup/doi/10.1073/pnas.0809964106

Kim JM, Vanguri S, Boeke JD, Gabriel A & Voytas DF (1998) Transposable Elements and Genome Organization: A Comprehensive Survey of Retrotransposons Revealed by the Complete Saccharomyces cerevisiae Genome Sequence. *Genome Res.* **8:** 464–478 Available at: http://genome.cshlp.org/lookup/doi/10.1101/gr.8.5.464

Kim N, Huang S -y. N, Williams JS, Li YC, Clark AB, Cho J-E, Kunkel TA, Pommier Y & Jinks-Robertson S (2011) Mutagenic Processing of Ribonucleotides in DNA by Yeast Topoisomerase I. *Science (80-.).* **332:** 1561–1564 Available at: https://www.sciencemag.org/lookup/doi/10.1126/science.1205016

Kimberland ML, Divoky V, Prchal J, Schwahn U, Berger W & Kazazian HH (1999) Full-Length Human L1 Insertions Retain the Capacity for High Frequency Retrotransposition in Cultured Cells. *Hum. Mol. Genet.* **8:** 1557–1560 Available at: https://academic.oup.com/hmg/article-lookup/doi/10.1093/hmg/8.8.1557

Kind B, Muster B, Staroske W, Herce HD, Sachse R, Rapp A, Schmidt F, Koss S, Cardoso MC & Lee-Kirsch MA (2014) Altered spatio-temporal dynamics of RNase H2 complex assembly at replication and repair sites in Aicardi–Goutières syndrome. *Hum. Mol. Genet.* **23:** 5950–5960 Available at: https://academic.oup.com/hmg/article-lookup/doi/10.1093/hmg/ddu319

Klawitter S, Fuchs N V., Upton KR, Muñoz-Lopez M, Shukla R, Wang J, Garcia-Cañadas M, Lopez-Ruiz C, Gerhardt DJ, Sebe A, Grabundzija I, Merkert S, Gerdes P, Pulgarin JA, Bock A, Held U, Witthuhn A, Haase A, Sarkadi B, Löwer J, et al (2016) Reprogramming triggers endogenous L1 and Alu retrotransposition in human induced pluripotent stem cells. *Nat. Commun.* **7:** 10286 Available at: http://www.nature.com/articles/ncomms10286

de Koning APJ, Gu W, Castoe TA, Batzer MA & Pollock DD (2011) Repetitive Elements May Comprise Over Two-Thirds of the Human Genome. *PLoS Genet.* **7:** e1002384 Available at: https://dx.plos.org/10.1371/journal.pgen.1002384

Kopera HC, Moldovan JB, Morrish TA, Garcia-Perez JL & Moran J V. (2011) Similarities between long interspersed element-1 (LINE-1) reverse transcriptase and telomerase. *Proc. Natl. Acad. Sci.* **108:** 20345–20350 Available at: http://www.pnas.org/cgi/doi/10.1073/pnas.1100275108

Kubo S, Seleme M d. C, Soifer HS, Perez JLG, Moran J V., Kazazian HH & Kasahara N (2006) L1 retrotransposition in nondividing and primary human somatic cells. *Proc. Natl. Acad. Sci.* **103:** 8036–8041 Available at: http://www.pnas.org/cgi/doi/10.1073/pnas.0601954103

Kulpa DA & Moran J V. (2005) Ribonucleoprotein particle formation is necessary but not sufficient for LINE-1 retrotransposition. *Hum. Mol. Genet.* **14:** 3237–3248 Available at: http://academic.oup.com/hmg/article/14/21/3237/2355827/Ribonucleoprotein-particle-formation-is-necessary

Kulpa DA & Moran J V. (2006a) Cis-preferential LINE-1 reverse transcriptase activity in ribonucleoprotein particles. *Nat. Struct. Mol. Biol.* **13:** 655–660 Available at: https://www.nature.com/articles/nsmb1107

Kulpa DA & Moran J V (2006b) Cis-preferential LINE-1 reverse transcriptase activity in ribonucleoprotein particles. *Nat. Struct. Mol. Biol.* **13:** 655–660 Available at: http://www.nature.com/articles/nsmb1107

Kurzynska-Kokorniak A, Jamburuthugoda VK, Bibillo A & Eickbush TH (2007) DNA-directed DNA Polymerase and Strand Displacement Activity of the Reverse Transcriptase Encoded by the R2 Retrotransposon. *J. Mol. Biol.* **374:** 322–333 Available at: https://linkinghub.elsevier.com/retrieve/pii/S0022283607012351

LaCava J, Jiang H & Rout MP (2016) Protein Complex Affinity Capture from Cryomilled Mammalian Cells. *J. Vis. Exp.***:** 1–11 Available at: http://www.jove.com/video/54518/protein-complex-affinity-capture-from-cryomilled-mammalian-cells

LaCava J, Molloy KR, Taylor MS, Domanski M, Chait BT & Rout MP (2015) Affinity proteomics to study endogenous protein complexes: Pointers, pitfalls, preferences and perspectives. *Biotechniques* **58:** 103–119

Lanciault C & Champoux JJ (2004) Single Unpaired Nucleotides Facilitate HIV-1 Reverse Transcriptase Displacement Synthesis through Duplex RNA. *J. Biol. Chem.* **279:** 32252–32261 Available at: https://linkinghub.elsevier.com/retrieve/pii/S0021925820775149

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, Fitzhugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, Levine R, McEwan P, McKernan K, et al (2001) Initial sequencing and analysis of the human genome. *Nature* **409:** 860–921 Available at: http://www.nature.com/articles/35057062

Larson PA, Moldovan JB, Jasti N, Kidd JM, Beck CR & Moran J V. (2018) Spliced integrated retrotransposed element (SpIRE) formation in the human genome. *PLOS Biol.* **16:** e2003067 Available at: https://dx.plos.org/10.1371/journal.pbio.2003067

Lebon P, Badoual J, Ponsot G, Goutières F, Hémeury-Cukier F & Aicardi J (1988) Intrathecal synthesis of interferon-alpha in infants with progressive familial encephalopathy. *J. Neurol. Sci.* **84:** 201–208 Available at: https://linkinghub.elsevier.com/retrieve/pii/0022510X88901256

Lee-Kirsch MA, Gong M, Chowdhury D, Senenko L, Engel K, Lee YA, De Silva U, Bailey SL, Witte T, Vyse TJ, Kere J, Pfeiffer C, Harvey S, Wong A, Koskenmies S, Hummel O, Rohde K, Schmidt RE, Dominiczak AF, Gahr M, et al (2007) Mutations in the gene encoding the 3′-5′ DNA exonuclease TREX1 are associated with systemic lupus erythematosus. *Nat. Genet.* **39:** 1065–1067 Available at: http://www.nature.com/articles/ng2091

Lenz J (2016) HERV-K HML-2 diversity among humans. *Proc. Natl. Acad. Sci.* **113:** 4240–4242 Available at: http://www.pnas.org/lookup/doi/10.1073/pnas.1603569113

Levis RW, Ganesan R, Houtchens K, Tolar LA & Sheen F (1993) Transposons in place of telomeric repeats at a Drosophila telomere. *Cell* **75:** 1083–1093 Available at: https://linkinghub.elsevier.com/retrieve/pii/009286749390318K

Li X, Zhang J, Jia R, Cheng V, Xu X, Qiao W, Guo F, Liang C & Cen S (2013) The MOV10 Helicase Inhibits LINE-1 Mobility. *J. Biol. Chem.* **288:** 21148–21160 Available at: https://linkinghub.elsevier.com/retrieve/pii/S0021925820455462

Lin GG & Scott JG (2012) R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Mol. Cell* **100:** 130–134

Lin Y-C, Boone M, Meuris L, Lemmens I, Van Roy N, Soete A, Reumers J, Moisse M, Plaisance S, Drmanac R, Chen J, Speleman F, Lambrechts D, Van de Peer Y, Tavernier J & Callewaert N (2014) Genome dynamics of the human embryonic kidney 293 lineage in response to cell biology manipulations. *Nat. Commun.* **5:** 4767 Available at: http://www.nature.com/articles/ncomms5767

Liu N, Lee CH, Swigut T, Grow E, Gu B, Bassik MC & Wysocka J (2018) Selective silencing of euchromatic L1s revealed by genome-wide screens for L1 regulators. *Nature* **553:** 228–232 Available at: http://www.nature.com/articles/nature25179

Livak KJ & Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2-ΔΔCT method. *Methods* **25:** 402–408 Available at: https://linkinghub.elsevier.com/retrieve/pii/S1046202301912629

Luan DD, Korman MH, Jakubczak JL & Eickbush TH (1993) Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: A mechanism for non-LTR retrotransposition. *Cell* **72:** 595–605 Available at: https://linkinghub.elsevier.com/retrieve/pii/0092867493900785

Macia A, Blanco-Jimenez E & García-Pérez JL (2015) Retrotransposons in pluripotent cells: Impact and new roles in cellular plasticity. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1849:** 417–426 Available at: http://dx.doi.org/10.1016/j.bbagrm.2014.07.007

Macia A, Muñoz-Lopez M, Cortes JL, Hastings RK, Morell S, Lucena-Aguilar G, Marchal JA, Badge RM & Garcia-Perez JL (2011) Epigenetic Control of Retrotransposon Expression in Human Embryonic Stem Cells. *Mol. Cell. Biol.* **31:** 300–316 Available at: http://mcb.asm.org/cgi/doi/10.1128/MCB.00561-10

Macia A, Widmann TJ, Heras SR, Ayllon V, Sanchez L, Benkaddour-Boumzaouad M, Muñoz-Lopez M, Rubio A, Amador-Cubero S, Blanco-Jimenez E, Garcia-Castro J, Menendez P, Ng P, Muotri AR, Goodier JL & Garcia-Perez JL (2017) Engineered LINE-1 retrotransposition in nondividing human neurons. *Genome Res.* **27:** 335–348 Available at: http://genome.cshlp.org/lookup/doi/10.1101/gr.206805.116

Mackenzie KJ, Carroll P, Lettice L, Tarnauskaitė Ž, Reddy K, Dix F, Revuelta A, Abbondati E, Rigby RE, Rabe B, Kilanowski F, Grimes G, Fluteau A, Devenney PS, Hill RE, Reijns MA & Jackson AP (2016) Ribonuclease H2 mutations induce a <scp>cGAS</scp> / <scp>STING</scp> -dependent innate immune response. *EMBO J.* **35:** 831–844 Available at: https://onlinelibrary.wiley.com/doi/10.15252/embj.201593339

Mackenzie KJ, Carroll P, Martin C-A, Murina O, Fluteau A, Simpson DJ, Olova N, Sutcliffe H, Rainger JK, Leitch A, Osborn RT, Wheeler AP, Nowotny M, Gilbert N, Chandra T, Reijns MAM & Jackson AP (2017) cGAS surveillance of micronuclei links genome instability to innate immunity. *Nature* **548:** 461–465 Available at: http://dx.doi.org/10.1038/nature23449

Mager DL & Freeman JD (2000) Novel Mouse Type D Endogenous Proviruses and ETn Elements Share Long Terminal Repeat and Internal Sequences. *J. Virol.* **74:** 7221–7229 Available at: https://journals.asm.org/doi/10.1128/JVI.74.16.7221-7229.2000

Mager DL & Stoye JP (2015) Mammalian Endogenous Retroviruses. In *Mobile DNA III* pp 1079–1100. Washington, DC, USA: ASM Press Available at: http://doi.wiley.com/10.1128/9781555819217.ch47

Malik HS (2000) Poised for Contagion: Evolutionary Origins of the Infectious Abilities of Invertebrate Retroviruses. *Genome Res.* **10:** 1307–1318 Available at: http://www.genome.org/cgi/doi/10.1101/gr.145000

Malik HS, Burke WD & Eickbush TH (1999) The age and evolution of non-LTR retrotransposable elements. *Mol. Biol. Evol.* **16:** 793–805 Available at: https://academic.oup.com/mbe/article-lookup/doi/10.1093/oxfordjournals.molbev.a026164

Mandal PK, Ewing AD, Hancks DC & Kazazian HH (2013) Enrichment of processed pseudogene transcripts in L1-ribonucleoprotein particles. *Hum. Mol. Genet.* **22:** 3730–3748

Mannion NM, Greenwood SM, Young R, Cox S, Brindle J, Read D, Nellåker C, Vesely C, Ponting CP, McLaughlin PJ, Jantsch MF, Dorin J, Adams IR, Scadden ADJ, Öhman M, Keegan LP & O'Connell MA (2014) The RNA-Editing Enzyme ADAR1 Controls Innate Immune Responses to RNA. *Cell Rep.* **9:** 1482–1494 Available at: https://linkinghub.elsevier.com/retrieve/pii/S2211124714009103

Marchetto MCN, Narvaiza I, Denli AM, Benner C, Lazzarini TA, Nathanson JL, Paquola ACM, Desai KN, Herai RH, Weitzman MD, Yeo GW, Muotri AR & Gage FH (2013) Differential L1 regulation in pluripotent stem cells of humans and apes. *Nature* **503:** 525–529 Available at: http://www.nature.com/articles/nature12686

Martienssen R (1998) Transposons, DNA methylation and gene control. *Trends Genet.* **14:** 263–4 Available at: http://www.ncbi.nlm.nih.gov/pubmed/9676527

Martin SL (1991) Ribonucleoprotein particles with LINE-1 RNA in mouse embryonal carcinoma cells. *Mol. Cell. Biol.* **11:** 4804–4807 Available at: https://mcb.asm.org/content/11/9/4804.long

Martin SL & Bushman FD (2001) Nucleic Acid Chaperone Activity of the ORF1 Protein from the Mouse LINE-1 Retrotransposon. *Mol. Cell. Biol.* **21:** 467–475 Available at: http://mcb.asm.org/cgi/doi/10.1128/MCB.21.2.467-475.2001

Mátés L, Chuah MKL, Belay E, Jerchow B, Manoj N, Acosta-Sanchez A, Grzela DP, Schmitt A, Becker K, Matrai J, Ma L, Samara-Kuko E, Gysemans C, Pryputniewicz D, Miskey C, Fletcher B, VandenDriessche T, Ivics Z & Izsvák Z (2009) Molecular evolution of a novel hyperactive Sleeping Beauty transposase enables robust stable gene transfer in vertebrates. *Nat. Genet.* **41:** 753–761 Available at: http://www.nature.com/articles/ng.343

Mathias S, Scott A, Kazazian H, Boeke J & Gabriel A (1991) Reverse transcriptase encoded by a human transposable element. *Science (80-.).* **254:** 1808–1810 Available at: https://www.sciencemag.org/lookup/doi/10.1126/science.1722352

McClintock B (1950) The origin and behavior of mutable loci in maize. *Proc. Natl. Acad. Sci.* **36:** 344–355 Available at: http://www.pnas.org/cgi/doi/10.1073/pnas.36.6.344

McElhinny SAN, Kumar D, Clark AB, Watt DL, Watts BE, Lundström E-B, Johansson E, Chabes A & Kunkel TA (2010) Genome instability due to ribonucleotide incorporation into DNA. *Nat. Chem. Biol.* **6:** 774–781 Available at: http://dx.doi.org/10.1038/nchembio.424

Medstrand P & Mager DL (1998) Human-Specific Integrations of the HERV-K Endogenous Retrovirus Family. *J. Virol.* **72:** 9782–9787 Available at: https://journals.asm.org/doi/10.1128/JVI.72.12.9782-9787.1998

Messerschmidt DM, Knowles BB & Solter D (2014) DNA methylation dynamics during epigenetic reprogramming in the germline and preimplantation embryos. *Genes Dev.* **28:** 812–828

Michelini F, Pitchiaya S, Vitelli V, Sharma S, Gioia U, Pessina F, Cabrini M, Wang Y, Capozzo I, Iannelli F, Matti V, Francia S, Shivashankar G V., Walter NG & Di Fagagna FDA (2017) Damage-induced lncRNAs control the DNA damage response through interaction with DDRNAs at individual double-strand breaks. *Nat. Cell Biol.* **19:** 1400–1411 Available at: http://www.nature.com/articles/ncb3643

Miki Y, Nishisho I, Horii A, Miyoshi Y, Utsunomiya J, Kinzler KW, Vogelstein B & Nakamura Y (1992) Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer. *Cancer Res.* **52:** 643–5 Available at: http://www.ncbi.nlm.nih.gov/pubmed/1310068

Mills RE, Bennett EA, Iskow RC & Devine SE (2007) Which transposable elements are active in the human genome? *Trends Genet.* **23:** 183–191 Available at: https://linkinghub.elsevier.com/retrieve/pii/S0168952507000595

Mita P, Sun X, Fenyö D, Kahler DJ, Li D, Agmon N, Wudzinska A, Keegan S, Bader JS, Yun C & Boeke JD (2020) BRCA1 and S phase DNA repair pathways restrict LINE-1 retrotransposition in human cells. *Nat. Struct. Mol. Biol.* **27:** 179–191 Available at: http://dx.doi.org/10.1038/s41594-020-0374-z

Moldovan JB & Moran J V. (2015) The Zinc-Finger Antiviral Protein ZAP Inhibits LINE and Alu Retrotransposition. *PLoS Genet.* **11:** 1–34

Monot C, Kuciak M, Viollet S, Mir AA, Gabus C, Darlix JL & Cristofari G (2013) The Specificity and Flexibility of L1 Reverse Transcription Priming at Imperfect T-Tracts. *PLoS Genet.* **9:**

Moran J V., DeBerardinis RJ & Kazazian HH (1999) Exon shuffling by L1 retrotransposition. *Science* **283:** 1530–4 Available at: https://www.sciencemag.org/lookup/doi/10.1126/science.283.5407.1530

Moran J V., Holmes SE, Naas TP, DeBerardinis RJ, Boeke JD & Kazazian HH (1996) High Frequency Retrotransposition in Cultured Mammalian Cells. *Cell* **87:** 917–927 Available at: https://linkinghub.elsevier.com/retrieve/pii/S0092867400819984

Morgan TH (1922) Croonian Lecture:—On the mechanism of heredity. *Proc. R. Soc. London. Ser. B, Contain. Pap. a Biol. Character* **94:** 162–197 Available at: https://royalsocietypublishing.org/doi/10.1098/rspb.1922.0053

Morimoto M & Kopan R (2009) rtTA toxicity limits the usefulness of the SP-C-rtTA transgenic mouse. *Dev. Biol.* **325:** 171–178 Available at: http://www.ncbi.nlm.nih.gov/pubmed/19013447

Morrish TA, Garcia-Perez JL, Stamato TD, Taccioli GE, Sekiguchi JA & Moran J V. (2007) Endonuclease-independent LINE-1 retrotransposition at mammalian telomeres. *Nature* **446:** 208–212

Morrish TA, Gilbert N, Myers JS, Vincent BJ, Stamato TD, Taccioli GE, Batzer MA & Moran J V. (2002) DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. *Nat. Genet.* **31:** 159–165 Available at: http://www.nature.com/articles/ng898z

Muckenfuss H, Hamdorf M, Held U, Perković M, Löwer J, Cichutek K, Flory E, Schumann GG & Münk C (2006) APOBEC3 Proteins Inhibit Human LINE-1 Retrotransposition. *J. Biol. Chem.* **281:** 22161–22172 Available at: https://linkinghub.elsevier.com/retrieve/pii/S0021925819477363

Muñoz-Lopez M, Garcia-Cañadas M, Macia A, Morell S & Garcia-Perez JL (2012) Analysis of LINE-1 expression in human pluripotent cells. In *Methods in Molecular Biology* pp 113–125. Available at: http://link.springer.com/10.1007/978-1-61779-794-1_7

Munoz-Lopez M & Garcia-Perez J (2010) DNA Transposons: Nature and Applications in Genomics. *Curr. Genomics* **11:** 115–128 Available at: http://www.eurekaselect.com/openurl/content.php?genre=article&issn=1389-2029&volume=11&issue=2&spage=115

Muñoz-Lopez M, Vilar R, Philippe C, Rahbari R, Richardson SR, Andres-Anton M, Widmann T, Cano D, Cortes JL, Rubio-Roldan A, Guichard E, Heras SR, Sanchez-Luque FJ, Morell M, Aguilar E, Garcia-Cañadas M, Sanchez L, Macia A, Vilches P, Nieto-Perez MC, et al (2019) LINE-1 retrotransposition impacts the genome of human pre-implantation embryos and extraembryonic tissues. *bioRxiv*: 522623 Available at: http://biorxiv.org/content/early/2019/01/18/522623.abstract

Muotri AR, Chu VT, Marchetto MCN, Deng W, Moran J V. & Gage FH (2005) Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature* **435:** 903–910 Available at: http://www.nature.com/articles/nature03663

Muotri AR, Marchetto MCN, Coufal NG, Oefner R, Yeo G, Nakashima K & Gage FH (2010) L1 retrotransposition in neurons is modulated by MeCP2. *Nature* **468:** 443–446 Available at: http://dx.doi.org/10.1038/nature09544

Oeffinger M, Wei KE, Rogers R, DeGrasse JA, Chait BT, Aitchison JD & Rout MP (2007) Comprehensive analysis of diverse ribonucleoprotein complexes. *Nat. Methods* **4:** 951–956 Available at: http://www.nature.com/articles/nmeth1101

Ohle C, Tesorero R, Schermann G, Dobrev N, Sinning I & Fischer T (2016) Transient RNA-DNA Hybrids Are Required for Efficient Double-Strand Break Repair. *Cell* **167:** 1001-1013.e7 Available at: https://linkinghub.elsevier.com/retrieve/pii/S0092867416313824

Olivares M, García-Pérez JL, Carmen Thomas M, Heras SR & Lopezl MC (2002) The non-LTR (long terminal repeat) retrotransposon L1Tc from Trypanosoma cruzi codes for a protein with RNase H activity. *J. Biol. Chem.* **277:** 28025–28030 Available at: http://www.jbc.org/lookup/doi/10.1074/jbc.M202896200

Orecchini E, Doria M, Antonioni A, Galardi S, Ciafrè SA, Frassinelli L, Mancone C, Montaldo C, Tripodi M & Michienzi A (2017) ADAR1 restricts LINE-1 retrotransposition. *Nucleic Acids Res.* **45:** 155–168 Available at: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkw834

Ostertag EM (2000) Determination of L1 retrotransposition kinetics in cultured cells. *Nucleic Acids Res.* **28:** 1418–1423

Peddigari S, Li PWL, Rabe JL & Martin SL (2013) HnRNPL and nucleolin bind LINE-1 RNA and function as host factors to modulate retrotransposition. *Nucleic Acids Res.* **41:** 575–585 Available at: http://academic.oup.com/nar/article/41/1/575/1159054/hnRNPL-and-nucleolin-bind-LINE1-RNA-and-function

Penzkofer T, Jäger M, Figlerowicz M, Badge R, Mundlos S, Robinson PN & Zemojtel T (2017) L1Base 2: more retrotransposition-active LINE-1s, more mammalian genomes. *Nucleic Acids Res.* **45:** D68–D73 Available at: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkw925

Perepelitsa-Belancio V & Deininger P (2003) RNA truncation by premature polyadenylation attenuates human mobile element activity. *Nat. Genet.* **35:** 363–366 Available at: http://www.nature.com/articles/ng1269

Pezic D, Manakov SA, Sachidanandam R & Aravin AA (2014) piRNA pathway targets active LINE1 elements to establish the repressive H3K9me3 mark in germ cells. *Genes Dev.* **28:** 1410–1428 Available at: http://genesdev.cshlp.org/cgi/doi/10.1101/gad.240895.114

Piskareva O, Denmukhametova S & Schmatchenko V (2003) Functional reverse transcriptase encoded by the human LINE-1 from baculovirus-infected insect cells. *Protein Expr. Purif.* **28:** 125–130 Available at: https://linkinghub.elsevier.com/retrieve/pii/S1046592802006551

Piskareva O & Schmatchenko V (2006) DNA polymerization by the reverse transcriptase of the human L1 retrotransposon on its own template in vitro. *FEBS Lett.* **580:** 661–668 Available at: http://doi.wiley.com/10.1016/j.febslet.2005.12.077

Pizarro JG & Cristofari G (2016) Post-Transcriptional Control of LINE-1 Retrotransposition by Cellular Host Factors in Somatic Cells. *Front. Cell Dev. Biol.* **4:** Available at: http://journal.frontiersin.org/Article/10.3389/fcell.2016.00014/abstract

Pokatayev V, Hasin N, Chon H, Cerritelli SM, Sakhuja K, Ward JM, Morris HD, Yan N & Crouch RJ (2016) RNase H2 catalytic core Aicardi-Goutières syndrome–related mutant invokes cGAS–STING innate immune-sensing pathway in mice. *J. Exp. Med.* **213:** 329–336 Available at: http://www.jem.org/lookup/doi/10.1084/jem.20151464

Potenski CJ, Niu H, Sung P & Klein HL (2014) Avoidance of ribonucleotide-induced mutations by RNase H2 and Srs2-Exo1 mechanisms. *Nature* **511:** 251–254 Available at: http://dx.doi.org/10.1038/nature13292

Przyborski SA, Christie VB, Hayman MW, Stewart R & Horrocks GM (2004) Human Embryonal Carcinoma Stem Cells: Models of Embryonic Development in Humans. *Stem Cells Dev.* **13:** 400–408 Available at: https://www.liebertpub.com/doi/10.1089/scd.2004.13.400

Radice AD, Bugaj B, Fitch DHA & Emmons SW (1994) Widespread occurrence of the Tc1 transposon family: Tc1-like transposons from teleost fish. *Mol. Gen. Genet. MGG* **244:** 606–612 Available at: http://link.springer.com/10.1007/BF00282750

Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA & Zhang F (2013) Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8:** 2281–2308 Available at: http://www.nature.com/articles/nprot.2013.143

Ravindran S (2012) Barbara McClintock and the discovery of jumping genes. *Proc. Natl. Acad. Sci.* **109:** 20198–20199 Available at: http://www.pnas.org/cgi/doi/10.1073/pnas.1219372109

Ray DA, Feschotte C, Pagan HJT, Smith JD, Pritham EJ, Arensburger P, Atkinson PW & Craig NL (2008) Multiple waves of recent DNA transposon activity in the bat, Myotis lucifugus. *Genome Res.* **18:** 717–728 Available at: http://www.genome.org/cgi/doi/10.1101/gr.071886.107

Reijns MAM, Bubeck D, Gibson LCD, Graham SC, Baillie GS, Jones EY & Jackson AP (2011) The Structure of the Human RNase H2 Complex Defines Key Interaction Interfaces Relevant to Enzyme Function and Human Disease*. *J. Biol. Chem.* **286:** 10530–10539 Available at: http://dx.doi.org/10.1074/jbc.M110.177394

Reijns MAM & Jackson AP (2014) Ribonuclease H2 in health and disease. *Biochem. Soc. Trans.* **42:** 717–725 Available at: https://portlandpress.com/biochemsoctrans/article/42/4/717/68720/Ribonuclease-H2-in-health-and-disease

Reijns MAM, Rabe B, Rigby RE, Mill P, Astell KR, Lettice LA, Boyle S, Leitch A, Keighren M, Kilanowski F, Devenney PS, Sexton D, Grimes G, Holt IJ, Hill RE, Taylor MS, Lawson KA, Dorin JR & Jackson AP (2012) Enzymatic Removal of Ribonucleotides from DNA Is Essential for Mammalian Genome Integrity and Development. *Cell* **149:** 1008–1022 Available at: https://linkinghub.elsevier.com/retrieve/pii/S0092867412005089

Ribet D, Dewannieux M & Heidmann T (2004) An active murine transposon family pair: Retrotransposition of 'master' MusD copies and ETn trans-mobilization. *Genome Res.* **14:** 2261–2267 Available at: http://www.genome.org/cgi/doi/10.1101/gr.2924904

Rice GI, Bond J, Asipu A, Brunette RL, Manfield IW, Carr IM, Fuller JC, Jackson RM, Lamb T, Briggs TA, Ali M, Gornall H, Couthard LR, Aeby A, Attard-Montalto SP, Bertini E, Bodemer C, Brockmann K, Brueton LA, Corry PC, et al (2009) Mutations involved in Aicardi-Goutières syndrome implicate SAMHD1 as regulator of the innate immune response. *Nat. Genet.* **41:** 829–832 Available at: http://www.nature.com/articles/ng.373

Rice GI, Forte GMA, Szynkiewicz M, Chase DS, Aeby A, Abdel-Hamid MS, Ackroyd S, Allcock R, Bailey KM, Balottin U, Barnerias C, Bernard G, Bodemer C, Botella MP, Cereda C, Chandler KE, Dabydeen L, Dale RC, De Laet C, De Goede CGEL, et al (2013) Assessment of interferon-related biomarkers in Aicardi-Goutières syndrome associated with mutations in TREX1, RNASEH2A, RNASEH2B, RNASEH2C, SAMHD1, and ADAR: a case-control study. *Lancet Neurol.* **12:** 1159–1169 Available at: https://linkinghub.elsevier.com/retrieve/pii/S1474442213702588

Rice GI, Kasher PR, Forte GMA, Mannion NM, Greenwood SM, Szynkiewicz M, Dickerson JE, Bhaskar SS, Zampini M, Briggs TA, Jenkinson EM, Bacino CA, Battini R, Bertini E, Brogan PA, Brueton LA, Carpanelli M, De Laet C, de Lonlay P, del Toro M, et al (2012) Mutations in ADAR1 cause Aicardi-Goutières syndrome associated with a type I interferon signature. *Nat. Genet.* **44:** 1243–1248 Available at: http://www.nature.com/articles/ng.2414

Rice GI, del Toro Duany Y, Jenkinson EM, Forte GMA, Anderson BH, Ariaudo G, Bader-Meunier B, Baildam EM, Battini R, Beresford MW, Casarano M, Chouchane M, Cimaz R, Collins AE, Cordeiro NJ V, Dale RC, Davidson JE, De Waele L, Desguerre I, Faivre L, et al (2014) Gain-of-function mutations in IFIH1 cause a spectrum of human disease phenotypes associated with upregulated type I interferon signaling. *Nat. Genet.* **46:** 503–509 Available at: http://www.nature.com/articles/ng.2933

Richardson SR, Doucet AJ, Kopera HC, Moldovan JB, Garcia-Perez JL & Moran J V. (2015) The Influence of LINE-1 and SINE Retrotransposons on Mammalian Genomes. *Microbiol. Spectr.* **3:** 1165–1208 Available at: http://www.asmscience.org/content/journal/microbiolspec/10.1128/microbiolspec.MDNA3-0061-2014

Richardson SR, Gerdes P, Gerhardt DJ, Sanchez-Luque FJ, Bodea G-O, Muñoz-Lopez M, Jesuadian JS, Kempen M-JHC, Carreira PE, Jeddeloh JA, Garcia-Perez JL, Kazazian HH, Ewing AD & Faulkner GJ (2017) Heritable L1 retrotransposition in the mouse primordial germline and early embryo. *Genome Res.* **27:** 1395–1405 Available at: http://genome.cshlp.org/lookup/doi/10.1101/gr.219022.116

Richardson SR, Narvaiza I, Planegger RA, Weitzman MD & Moran J V. (2014) APOBEC3A deaminates transiently exposed single-strand DNA during LINE-1 retrotransposition. *Elife* **3:** e02008 Available at: https://elifesciences.org/articles/02008

Rodriguez-Martin B, Alvarez EG, Baez-Ortega A, Zamora J, Supek F, Demeulemeester J, Santamarina M, Ju YS, Temes J, Garcia-Souto D, Detering H, Li Y, Rodriguez-Castro J, Dueso-Barroso A, Bruzos AL, Dentro SC, Blanco MG, Contino G, Ardeljan D, Tojo M, et al (2020) Pan-cancer analysis of whole genomes identifies driver rearrangements promoted by LINE-1 retrotransposition. *Nat. Genet.* **52:** 306–319 Available at: http://www.nature.com/articles/s41588-019-0562-0

Rossant J & Papaioannou VE (1984) The relationship between embryonic, embryonal carcinoma and embryo-derived stem cells. *Cell Differ.* **15:** 155–161 Available at: https://linkinghub.elsevier.com/retrieve/pii/004560398490068X

Sanchez-Luque FJ, Kempen M-JHC, Gerdes P, Vargas-Landin DB, Richardson SR, Troskie R-L, Jesuadian JS, Cheetham SW, Carreira PE, Salvador-Palomeque C, García-Cañadas M, Muñoz-Lopez M, Sanchez L, Lundberg M, Macia A, Heras SR, Brennan PM, Lister R, Garcia-Perez JL, Ewing AD, et al (2019) LINE-1 Evasion of Epigenetic Repression in Humans. *Mol. Cell* **75:** 590-604.e12 Available at: https://linkinghub.elsevier.com/retrieve/pii/S109727651930396X

Sassaman DM, Dombroski BA, Moran J V., Kimberland ML, Naas TP, DeBerardinis RJ, Gabriel A, Swergold GD & Kazazian HH (1997) Many human L1 elements are capable of retrotransposition. *Nat. Genet.* **16:** 37–43 Available at: http://www.nature.com/articles/ng0597-37

Sayah DM, Sokolskaja E, Berthoux L & Luban J (2004) Cyclophilin A retrotransposition into TRIM5 explains owl monkey resistance to HIV-1. *Nature* **430:** 569–573 Available at: http://www.nature.com/articles/nature02777

Schneider WM, Chevillotte MD & Rice CM (2014) Interferon-stimulated genes: A complex web of host defenses. *Annu. Rev. Immunol.* **32:** 513–545 Available at: http://www.annualreviews.org/doi/10.1146/annurev-immunol-032713-120231

Scott EC & Devine SE (2017) The role of somatic L1 retrotransposition in human cancers. *Viruses* **9:** 131 Available at: http://www.mdpi.com/1999-4915/9/6/131

Sen SK, Huang CT, Han K & Batzer MA (2007) Endonuclease-independent insertion provides an alternative pathway for L1 retrotransposition in the human genome. *Nucleic Acids Res.* **35:** 3741–3751 Available at: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkm317

Shapiro JA (1969) Mutations caused by the insertion of genetic material into the galactose operon of Escherichia coli. *J. Mol. Biol.* **40:** 93–105

Shukla R, Upton KR, Muñoz-Lopez M, Gerhardt DJ, Fisher ME, Nguyen T, Brennan PM, Baillie JK, Collino A, Ghisletti S, Sinha S, Iannelli F, Radaelli E, Dos Santos A, Rapoud D, Guettier C, Samuel D, Natoli G, Carninci P, Ciccarelli FD, et al (2013) Endogenous retrotransposition activates oncogenic pathways in hepatocellular carcinoma. *Cell* **153:** 101–111

Simmons GM (1992) Horizontal transfer of hobo transposable elements within the Drosophila melanogaster species complex: evidence from DNA sequencing. *Mol. Biol. Evol.* **9:** 1050–1060 Available at: https://academic.oup.com/mbe/article/9/6/1050/1073670/Horizontal-transfer-of-hobo-transposable-elements

Simon M, Van Meter M, Ablaeva J, Ke Z, Gonzalez RS, Taguchi T, De Cecco M, Leonova KI, Kogan V, Helfand SL, Neretti N, Roichman A, Cohen HY, Meer M V., Gladyshev VN, Antoch MP, Gudkov A V., Sedivy JM, Seluanov A & Gorbunova V (2019) LINE1 Derepression in Aged Wild-Type and SIRT6-Deficient Mice Drives Inflammation. *Cell Metab.* **29:** 871-885.e5 Available at: https://doi.org/10.1016/j.cmet.2019.02.014

Singh P, Bourque G, Craig NL, Dubnau JT, Feschotte C, Flasch DA, Gunderson KL, Malik H, Moran J V., Peters JE, Slotkin R & Levin HL (2014) Mobile genetic elements and genome evolution 2014. *Mob. DNA* **5:** 26 Available at: http://mobilednajournal.biomedcentral.com/articles/10.1186/1759-8753-5-26

Skowronski J, Fanning TG & Singer MF (1988) Unit-length line-1 transcripts in human teratocarcinoma cells. *Mol. Cell. Biol.* **8:** 1385–1397 Available at: http://mcb.asm.org/lookup/doi/10.1128/MCB.8.4.1385

Smit AFA (1996) The origin of interspersed repeats in the human genome. *Curr. Opin. Genet. Dev.* **6:** 743–748 Available at: http://www.nature.com/nrc/journal/v16/n8/full/nrc.2016.62.html

Sparks JL, Chon H, Cerritelli SM, Kunkel TA, Johansson E, Crouch RJ & Burgers PM (2012) RNase H2-Initiated Ribonucleotide Excision Repair. *Mol. Cell* **47:** 980–986 Available at: https://linkinghub.elsevier.com/retrieve/pii/S1097276512005990

Speek M (2001) Antisense Promoter of Human L1 Retrotransposon Drives Transcription of Adjacent Cellular Genes. *Mol. Cell. Biol.* **21:** 1973–1985 Available at: https://journals.asm.org/doi/10.1128/MCB.21.6.1973-1985.2001

Sperger JM, Chen X, Draper JS, Antosiewicz JE, Chon CH, Jones SB, Brooks JD, Andrews PW, Brown PO & Thomson JA (2003) Gene expression patterns in human embryonic stem cells and human pluripotent germ cell tumors. *Proc. Natl. Acad. Sci.* **100:** 13350–13355 Available at: http://www.pnas.org/cgi/doi/10.1073/pnas.2235735100

Stetson DB, Ko JS, Heidmann T & Medzhitov R (2008) Trex1 Prevents Cell-Intrinsic Initiation of Autoimmunity. *Cell* **134:** 587–598 Available at: https://doi.org/10.1016/j.cell.2008.06.032

Sugano T, Kajikawa M & Okada N (2006) Isolation and characterization of retrotransposition-competent LINEs from zebrafish. *Gene* **365:** 74–82 Available at: https://linkinghub.elsevier.com/retrieve/pii/S0378111905006542

Swamynathan SK (2010) Krüppel-like factors: three fingers in control. *Hum. Genomics* **4:** 263–270

Swergold GD (1990) Identification, characterization, and cell specificity of a human LINE-1 promoter. *Mol. Cell. Biol.* **10:** 6718–6729 Available at: http://mcb.asm.org/lookup/doi/10.1128/MCB.10.12.6718

Swift S, Lorens J, Achacoso P & Nolan GP (2001) Rapid Production of Retroviruses for Efficient Gene Delivery to Mammalian Cells Using 293 T Cell–Based Systems. *Curr. Protoc. Immunol.* **31:** Available at: https://onlinelibrary.wiley.com/doi/10.1002/0471142735.im1017cs31

Taylor AL (1963) BACTERIOPHAGE-INDUCED MUTATION IN ESCHERICHIA COLI. *Proc. Natl. Acad. Sci.* **50:** 1043–1051 Available at: http://www.pnas.org/cgi/doi/10.1073/pnas.50.6.1043

Taylor MS, Altukhov I, Molloy KR, Mita P, Jiang H, Adney EM, Wudzinska A, Badri S, Ischenko D, Eng G, Burns KH, Fenyö D, Chait BT, Alexeev D, Rout MP, Boeke JD & LaCava J (2018) Dissection of affinity captured LINE-1 macromolecular complexes. *Elife* **7:** 1–40

Taylor MS, LaCava J, Mita P, Molloy KR, Huang CRL, Li D, Adney EM, Jiang H, Burns KH, Chait BT, Rout MP, Boeke JD & Dai L (2013) Affinity Proteomics Reveals Human Host Factors Implicated in Discrete Stages of LINE-1 Retrotransposition. *Cell* **155:** 1034–1048 Available at: http://dx.doi.org/10.1016/j.cell.2013.10.021

Thayer RE, Singer MF & Fanning TG (1993) Undermethylation of specific LINE-1 sequences in human cells producing a LINE-1 -encoded protein. *Gene* **133:** 273–277 Available at: https://linkinghub.elsevier.com/retrieve/pii/037811199390651I

Thomas CA, Tejwani L, Trujillo CA, Negraes PD, Herai RH, Mesci P, Macia A, Crow YJ & Muotri AR (2017) Modeling of TREX1-Dependent Autoimmune Disease using Human Stem Cells Highlights L1 Accumulation as a Source of Neuroinflammation. *Cell Stem Cell* **21:** 319-331.e8 Available at: https://linkinghub.elsevier.com/retrieve/pii/S1934590917302886

Trinkle-Mulcahy L, Boulon S, Lam YW, Urcia R, Boisvert FM, Vandermoere F, Morrice NA, Swift S, Rothbauer U, Leonhardt H & Lamond A (2008) Identifying specific protein interaction partners using quantitative mass spectrometry and bead proteomes. *J. Cell Biol.* **183:** 223–239 Available at: https://rupress.org/jcb/article/183/2/223/45698/Identifying-specific-protein-interaction-partners

Tristán-Ramos P, Rubio-Roldan A, Peris G, Sánchez L, Amador-Cubero S, Viollet S, Cristofari G & Heras SR (2020) The tumor suppressor microRNA let-7 inhibits human LINE-1 retrotransposition. *Nat. Commun.* **11:** Available at: http://dx.doi.org/10.1038/s41467-020-19430-4

Tubio JMC, Li Y, Ju YS, Martincorena I, Cooke SL, Tojo M, Gundem G, Pipinikas CP, Zamora J, Raine K, Menzies A, Roman-Garcia P, Fullam A, Gerstung M, Shlien A, Tarpey PS, Papaemmanuil E, Knappskog S, Van Loo P, Ramakrishna M, et al (2014) Extensive transduction of nonrepetitive DNA mediated by L1 retrotransposition in cancer genomes. *Science (80-.).* **345:** 1251343–1251343 Available at: https://www.sciencemag.org/lookup/doi/10.1126/science.1251343

Tyanova S, Temu T & Cox J (2016) The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat. Protoc.* **11:** 2301–2319 Available at: http://dx.doi.org/10.1038/nprot.2016.136

Ullu E & Melli M (1982) Cloning and characterization of cDNA copies of the 7S RNAs of HeLa cells. *Nucleic Acids Res.* **10:** 2209–2223 Available at: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/10.7.2209

Upton KR, Gerhardt DJ, Jesuadian JS, Richardson SR, Sánchez-Luque FJ, Bodea GO, Ewing AD, Salvador-Palomeque C, van der Knaap MS, Brennan PM, Vanderver A & Faulkner GJ (2015) Ubiquitous L1 Mosaicism in Hippocampal Neurons. *Cell* **161:** 228–239 Available at: https://doi.org/10.1016/j.cell.2015.03.026

Urlinger S, Baron U, Thellmann M, Hasan MT, Bujard H & Hillen W (2000) Exploring the sequence space for tetracycline-dependent transcriptional activators: Novel mutations yield expanded range and sensitivity. *Proc. Natl. Acad. Sci.* **97:** 7963–7968 Available at: http://www.pnas.org/cgi/doi/10.1073/pnas.130192197

Volkman HE & Stetson DB (2014) The enemy within: Endogenous retroelements and autoimmune disease. *Nat. Immunol.* **15:** 415–422 Available at: http://www.nature.com/articles/ni.2872

Vuong LM, Pan S & Donovan PJ (2019) Proteome Profile of Endogenous Retrotransposon-Associated Complexes in Human Embryonic Stem Cells. *Proteomics* **19:**

Ward JR, Vasu K, Deutschman E, Halawani D, Larson PA, Zhang D, Willard B, Fox PL, Moran J V. & Longworth MS (2017) Condensin II and GAIT complexes cooperate to restrict LINE-1 retrotransposition in epithelial cells. *PLoS Genet.* **13:** e1007051 Available at: https://dx.plos.org/10.1371/journal.pgen.1007051

Warkocki Z, Krawczyk PS, Adamska D, Bijata K, Garcia-Perez JL & Dziembowski A (2018) Uridylation by TUT4/7 Restricts Retrotransposition of Human LINE-1s. *Cell***:** 1–12 Available at: https://linkinghub.elsevier.com/retrieve/pii/S0092867418309176

Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M, An P, Antonarakis SE, Attwood J, Baertsch R, Bailey J, Barlow K, Beck S, Berry E, Birren B, Bloom T, Bork P, et al (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature* **420:** 520–562 Available at: http://www.nature.com/nature/journal/v420/n6915/full/nature01262.html

Wei W, Gilbert N, Ooi SL, Lawler JF, Ostertag EM, Kazazian HH, Boeke JD & Moran J V. (2001) Human L1 Retrotransposition: cis Preference versus trans Complementation. *Mol. Cell. Biol.* **21:** 1429–1439 Available at: https://mcb.asm.org/content/21/4/1429

Wildschutte JH, Williams ZH, Montesion M, Subramanian RP, Kidd JM & Coffin JM (2016) Discovery of unfixed endogenous retrovirus insertions in diverse human populations. *Proc. Natl. Acad. Sci.* **113:** E2326–E2334 Available at: http://www.pnas.org/lookup/doi/10.1073/pnas.1602336113

Wilkinson DA, Mager DL & Leong J-AC (1994) Endogenous Human Retroviruses. In *The Retroviridae* pp 465–535. Boston, MA: Springer US Available at: http://link.springer.com/10.1007/978-1-4899-1730-0_9

Wissing S, Mũoz-lopez M, Macia A, Yang Z, Montano M, Collins W, Garcia-perez JL, Moran J V. & Greene WC (2012) Reprogramming somatic cells into ips cells activates line-1 retroelement mobility. *Hum. Mol. Genet.* **21:** 208–218 Available at: https://academic.oup.com/hmg/article-lookup/doi/10.1093/hmg/ddr455

Xie Y, Rosser JM, Thompson TL, Boeke JD & An W (2011) Characterization of L1 retrotransposition with high-throughput dual-luciferase assays. *Nucleic Acids Res.* **39:** e16–e16 Available at: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkq1076

Zeng Y & Chen T (2019) DNA Methylation Reprogramming during Mammalian Development. *Genes (Basel).* **10:** 257 Available at: https://www.mdpi.com/2073-4425/10/4/257

Zeuthen J, Nørgaard JOR, Avner P, Fellous M, Wartiovaara J, Vaheri A, Rosén A & Giovanella BC (1980) Characterization of a human ovarian teratocarcinoma-derived cell line. *Int. J. Cancer* **25:** 19–32 Available at: https://onlinelibrary.wiley.com/doi/10.1002/ijc.2910250104

Zhang A, Dong B, Doucet AJ, Moldovan JB, Moran J V. & Silverman RH (2014) RNase L restricts the mobility of engineered retrotransposons in cultured human cells. *Nucleic Acids Res.* **42:** 3803–3820 Available at: https://academic.oup.com/nar/article/42/6/3803/2435217

Zhao K, Du J, Han X, Goodier JL, Li P, Zhou X, Wei W, Evans SL, Li L, Zhang W, Cheung LE, Wang G, Kazazian HH & Yu XF (2013) Modulation of LINE-1 and Alu/SVA Retrotransposition by Aicardi-Goutières Syndrome-Related SAMHD1. *Cell Rep.* **4:** 1108–1115 Available at: http://dx.doi.org/10.1016/j.celrep.2013.08.019

Figures 42, 43, 46 and 50 were created with BioRender.com.

Venn diagrams were designed using the tool http://www.interactivenn.net/ (Heberle *et al*, 2015).