L1 retrotransposon activity in muscle cells

Paula Peressini Lopez

UNIVERSITÉ CÔTE D'AZUR

ÉCOLE DOCTORALE
SCIENCES DE LA VIE ET DE LA SANTÉ

THÈSE DE DOCTORAT

Activité du rétrotransposon L1 dans les cellules musculaires

Paula PERESSINI LÓPEZ

Institute de Recherche sur le Cancer et le Vieillissement – IRCAN

Présentée en vue de l'obtention du grade de docteur en Sciences de la vie et la Santé
d'Université Côte d'Azur

Mention : interactions moléculaires et cellulaires

Dirigée par : Gaël Cristofari / Chloé Féral
Soutenue le : 29 Juin 2020

Devant le jury, composé de :
Président du jury
Eric Röttinger, Directeur de Recherche CNRS, IRCAN
Rapporteurs
Chantal Vaury-Zwiller, Directrice de Recherche CNRS, Université Clermont Auvergne
Pierre-Antoine Defossez, Directeur de Recherche CNRS, Université Paris 7
Examinatrice
Deborah Bourc'his, Directrice de Recherche INSERM, Institut Curie

Activité du rétrotransposon L1 dans les cellules musculaires

L1 retrotransposon activity in muscle cells

Jury :

Président du jury

Eric Röttinger, Directeur de Recherche CNRS, IRCAN

Rapporteurs

Chantal Vaury-Zwiller, Directrice de Recherche CNRS, Université Clermont Auvergne

Pierre-Antoine Defossez, Directeur de Recherche CNRS, Université Paris 7

Examinatrice

Deborah Bourc'his, Directrice de Recherche INSERM, Institut Curie

Résumé de thèse

Près de la moitié du génome humain provient d'éléments transposables (TE). Parmi eux, l'élément LINE-1 ou L1 (Long INterspersed Element-1) forme la seule famille d'éléments transposables actuellement active et autonome chez l'Homme. Bien que des centaines de milliers de copies soient dispersées dans le génome humain, seules 80 à 100 d'entre elles sont encore compétentes pour la rétrotransposition, c'est-à-dire capables de se reproduire par un mécanisme de "copier-coller" via un ARN intermédiaire et une étape de transcription inverse. L'activité des L1s peut avoir des conséquences délétères, en particulier par mutagenèse insertionnelle. Elle est néanmoins étroitement régulée au niveau transcriptionnel et post-transcriptionnel. Inversement, des facteurs d'hôtes spécifiques sont nécessaires pour accomplir le cycle réplicatif des L1s. Lorsqu'elles se produisent dans la lignée germinale ou dans l'embryon précoce, les insertions de L1 peuvent être transmises à la génération suivante. La rétrotransposition des L1s a également été décrite dans certains tissus somatiques, comme dans les tumeurs épithéliales et dans le cerveau, à la fois dans les cellules progénitrices neurales et dans les neurones différenciés. Néanmoins, les niveaux d'expression des L1 compétents pour la rétrotransposition, et leur mobilisation, dans d'autres tissus somatiques restent incertains.

Ici, nous avons étudié l'activité des rétrotransposons L1 dans les cellules musculaires squelettiques humaines et murines. Nous montrons que la protéine du L1 la plus abondante, ORF1p, qui est essentielle à la rétrotransposition, est indétectable dans nos conditions expérimentales, dans des échantillons murins ou humains de muscle squelettique, alors qu'elle est facilement détectable dans les cellules cancéreuses ou dans les testicules. De même, elle n'est pas détectée dans les myoblastes immortalisés d'origine murine ou humaine. En revanche, nous avons découvert que le L1 est capable de rétrotransposition dans les myoblastes humains et murins lorsqu'elle est exprimée à partir d'un plasmide ou d'une copie intégrée avec un promoteur constitutif ou inductible, respectivement. En conclusion, si l'expression du L1 est inférieure à la limite de détection dans le muscle, les myoblastes sont bien permissifs à la rétrotransposition, ce qui indique que ces cellules expriment tous les facteurs cellulaires nécessaires pour réaliser ce processus, et n'expriment pas de facteurs de restriction significatifs qui bloqueraient la rétrotransposition.

Dans l'ensemble, nos résultats suggèrent que l'activité somatique des L1s pourrait ne pas être restreinte au cerveau ou aux cellules cancéreuses, mais pourrait également avoir lieu dans les muscles dans des conditions environnementales ou pathologiques qui déclencheraient leur expression.

Mots-clés: rétrotransposon, muscle squelettique, vieillissement, insertion

Abstract

Almost half of the human genome derives from transposable elements (TE). Among them, the Long INterspersed Element-1 (LINE-1 or L1) forms the only currently active and autonomous transposable element family in humans. Although hundreds of thousands L1 copies are dispersed in the human genome, only 80-100 of them are still retrotransposition competent, i.e. able to replicate by a “copy-and-paste” mechanism via an RNA intermediate and a reverse transcription step. On the one hand, L1 activity can have deleterious consequences, such as insertional mutagenesis, and is tightly regulated at the transcriptional or post-transcriptional levels. However, specific host factors are necessary for completion of L1 replication cycle. When occurring in the germline or in the early embryo, L1 insertions can be transmitted to the next generation. Somatic retrotransposition has been also described in epithelial tumors and in the brain, both in neural progenitor cells and differentiated neurons. Nevertheless, the extent of L1 expression and mobilization in other somatic tissues remains unclear.

Here, we investigated the activity of L1 retrotransposons in human and mouse skeletal muscle cells. We show that the most abundant L1 protein, ORF1p, which is essential to retrotransposition, is undetectable under our experimental conditions, in mouse or human muscle samples, while it is readily detected in cancer cells or in testis. Similarly, it was undetected in immortalized mouse or human myoblasts. However, we found that L1 is capable of retrotransposition in human and mouse myoblasts when expressed from a plasmid or from an integrated copy with a constitutive or inducible promoter, respectively. In conclusion, while L1 expression is under the limit of detection in muscle, myoblasts are permissive to retrotransposition, indicating that these cells express all the cellular factors necessary to achieve this process, and do not express significant restriction factors that would prevent retrotransposition.

Altogether, our findings suggest that somatic L1 activity could not be confined to the brain or cancer cells, but could also occur in muscles under environmental or pathological conditions that would unleash L1 expression.

Keywords: retrotransposon, skeletal muscle, aging, insertion

« *For a research worker the unforgotten moments of his life are those rare ones which come after years of plodding work, when the veil over nature's secret seems suddenly to lift* »

Gerty Cori

Acknowledgements

Firstly, I would like to thank both of my PhD supervisors, Chloé and Gaël, because this work could not be possible without their guidance. Thanks to Chloé for being always available and eager to help, for having her door open and for being so supportive. Thank you to Gaël for all those “do you have a minute?” that often turned into 2 hours of discussion, for pushing me to always improve myself and for helping me get to the finish line. To both of them, thank you for choosing me and for giving me this opportunity.

I thank the Signalife program for providing the funding and administrative support for this project. I would also extend my gratitude to the members of the jury for kindly agreeing to review my work.

Next, I have to thank my team, since each one of them has been part of my PhD experience in one way or the other. Julian, my first lab partner, for being a force that moved this project forward from the beginning with his enthusiasm and motivation (I missed our laughs in the cell culture). Thank you to Emy, for taking over and for pushing with me until the end. To Aurelien, for his help in this project and the innumerable times he lent me his advice. To Vivien, for cloning the constructs and for always being so available and kind. Thanks to Arpita, for her involvement in this project and for those talks on the corridor, for always knowing how and when to motivate me. To Sophie, for her kindness. To Claude, because this team would not be the same without him, neither would my experience here. Thank you for being such a generous person. Finally, to my dear Ramona, for being one of the highlights of this journey. I’m incredibly glad I got to meet someone so wonderful.

I would like to thank the team of Chloé, for their help in this project. Specially to Laetitia, for kindly teaching me several techniques, and to Laurence for her help with cell culture. I would like to show my gratitude to Pr. Sabrina Sacconi, for providing the samples and her expertise. I also wish to thank the cytometry and imagery platforms from IRCAN for their help in this project, specially to Ludo for being so available and helpful.

I thank the Gilson team for being amazing neighbors with contagious good vibes. Special thanks to Sol for all her help with the muscle cells, reagents and insights. Thank you to the Liti team for their great people and better coffee room.

I would like to express my deepest gratitude to everyone I met during these wonderful (almost) 4 years for helping shape this experience. Thank you to Tomás and Gaia, for taking me in at the very beginning and for showing me the amazing souls they are (os quiero!). To Julien, for being such a great friend and for all those movie nights that helped me get through the week. To Lorenzo for his hospitality and all the Spritz. To Matteo, for

being so kind. To Joan Pau, for his friendship (t'estimo!). To Sanya and Hereroa, for all the coffees and laughs after lunch. I would like to specially thank TGIFs, for existing and saving many of our lives.

I am incredibly thankful for having had the luck of sharing a home with two wonderful people: Alice and Juliette. Thank you to Alice, for putting up with me, for her empathy, for her laugher and her good heart. Thank you to Juliette, for her contagious permanent joyful mood, for her kindness and sensitivity. I could not have asked for better roommates.

Finally, I would like to thank my family for their unconditional support and for putting up with not seeing me for so long, definitely the biggest sacrifice of this experience. Thank you for being always one call away, no matter what. Thank you for always pushing me to move forward and to believe I can do whatever I put my mind to. Most importantly, thank you for your unlimited love and care.

And last, but never least, a thousand thankyous to Pablo. You have sacrificed as much as I did in this process and have always been there to love and support me. Thank you for putting up with my ups and downs, for all the trips, the calls despite the awful internet connection and specially, the laughs. Let's never stop laughing together.

Abbreviations

cDNA: Complementary DNA

CMV: Cytomegalovirus

DNA: Deoxyribonucleic acid

Dox: Doxycycline

ECM: Extracellular matrix

EN: Endonuclease

L1HS: Human-specific L1

LINE-1 (L1): Long Interspersed Nuclear element-1

LTR: Long terminal repeat

MyLC: Myosin Light Chain

MyHC: Myosin Heavy Chain

MyoD: myoblast determination protein 1

MYOG: Myogenin

ORF: Open reading frame

PCR: Polymerase chain reaction

RNA Pol II: RNA polymerase II

RC-L1: Retrotransposition-competent L1

RLE: Restriction-like endonuclease

RNA: Ribonucleic acid

RNase H: Ribonuclease H

RNP: Ribonucleoprotein particle

RT: Reverse transcriptase

SB: Sleeping Beauty

SINE: Short Interspersed Element

SVA: SINE-VNTR-Alu

TE: Transposable element

TPRT: Target-Primed Reverse Transcription

TSD: Target site duplication

UTR: Untranslated region

VLP: Virus-like particle

Table of contents

1. Introduction to mobile DNA

1.Introduction to mobile DNA: Discovery and definitions

During the summer of 1944, at the Cold Spring Harbor Laboratory, Barbara McClintock began studying the mosaicism of corn kernels and its genetic inheritance. She described two loci that were involved in the coloration of the kernels, naming them *Dissociator (Ds)* and *Activator (Ac)*. Towards 1948, she discovered that Ds and Ac were sequences that could change their genomic location, what we know today as Transposable Elements (TEs). Importantly, she found that the location where these sequences were inserted was relevant since they could change the expression of the surrounding genes, and therefore called them “*controlling elements*” [1].

Until then, DNA was thought to be relatively static, and faithfully replicated from one generation to another, beside sporadic mutations. The discovery of these “controlling elements” implied a paradigmatic change: genomes became much more plastic. The discovery of TEs in the genetic material of living organisms has led to intense debates relative to their potential function. Initially, mobile DNA was considered to be “junk” DNA. Nowadays, it is no longer regarded as “junk”, but rather parasitic DNA, that in some cases can be mutualistic. Mobile DNA appears in every form of life, from bacteria, to plants, to animals. The proportion of TE sequences in each genome varies, and there is a direct correlation between the size of genomes and the percentage of TEs they contain [2,3].

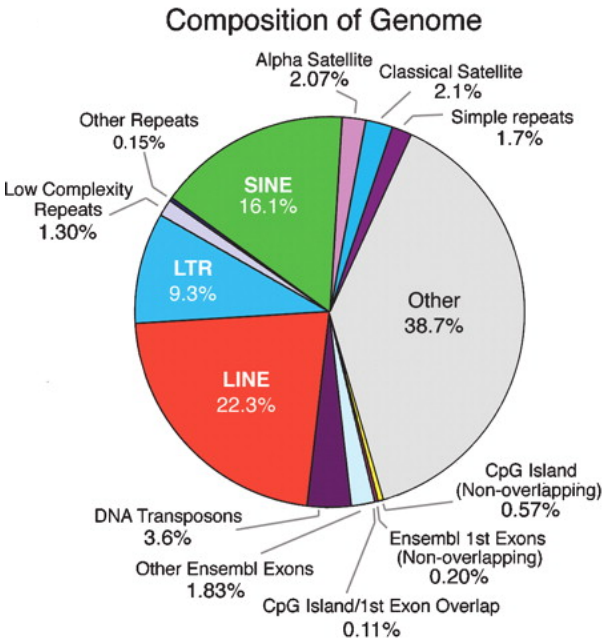


Figure 1. Composition of the genome. Showing the distribution of transposable elements (From [5]).

Regarding the human genome, it is interesting to highlight that TEs account for 45% of it [2,3] (Figure 1). Similarly, TEs represent 37.5% of the mouse genome [4].

This, together with their presence in the genetic material of nearly every type of living form suggests that mobile DNA is indeed more important than originally believed.

1.1. Mobile DNA comes in different forms

Mobile DNA can be classified into two main categories regarding the nature of the sequences mobilized: DNA Transposons and Retrotransposons.

1.1.1. DNA transposons mobilize by a "cut-and-paste" mechanism

DNA transposons usually mobilize by a process commonly known as cut-and-paste, carried out by an enzyme called DNA transposase. Some DNA transposons contain two inverted terminal repeats (ITRs) located 5' and 3' relative to a transposase gene (Figure 2). If present, the transposase can bind the ITRs, excise the sequence between them and re-insert it into another genomic location. Depending on transposon family, a DNA scarce, in the form of target-site duplication or deletion can be left or not at the source or at the target locus, sometimes caused by the repair mechanisms of the host cell after transposase action [6].

DNA transposons are active in most orders, but their activity has ceased in mammals, except in bats [7]. Examples of active DNA transposons include the Ac/Ds elements (maize), PiggyBac (moth), IS (insertion sequence) elements (bacteria) or P-elements (*Drosophila*) [8-12]. DNA transposons form a smaller part of the human genome as compared to the rest of TEs (only 3%) [3]. Therefore, we will not detail their biology in the present thesis.

A notable example of DNA transposons is *Sleeping Beauty*, a transposon from the Tc1/mariner family found in salmon. It contains two ITRs flanking the transposase sequence. Sleeping Beauty was an inactive fish transposon that was reactivated by restoring

its coding sequence by Izsvak and colleagues in 1997 and is currently used as a transgenesis tool that allows the efficient integration of a sequence of interest in the genome of many Vertebrates [13]. The engineered system functions by co-transfection of the sequence of interest, flanked by ITRs, along with a plasmid expressing the transposase. The transposase will recognize the ITRs in trans and will excise them, together with the internal sequence of interest while remaining bound to them. Once this step is completed, the transposase will integrate the bound DNA into a new genomic location [14].

Figure 2. Structure of a DNA transposon. The transposase gene is flanked by the inverted repeats (ITR).

1.1.2. Retrotransposons create new insertions through a "copy-and-paste" process

In contrast to DNA transposons, retrotransposons rely on an RNA intermediate and a reverse transcription step to generate a DNA molecule that can integrate into a new locus [15], a process commonly known as "copy-and-paste".

1.1.2.1. Autonomous retrotransposons encode their own enzymatic machinery

Autonomous retrotransposons are those that encode in their sequences all the necessary enzymatic machinery for their retrotransposition.

Autonomous retrotransposons can be classified in two main groups depending on whether they contain long-terminal repeats (LTRs) or not (Figure 3). LTR-retrotransposons contain LTRs at each end and are between 300 to 1000 base pairs (bp) long. They encode two essential genes: *gag* which encodes the structural proteins that form the virus-like particles (VLPs), and *pol* which codes for proteins with reverse transcriptase, integrase and protease activities. The 5' LTR functions as a promoter that drives the transcription of the genomic RNA. Upon transcription and translation, VLPs can assemble [17,18]. They contain the genomic RNA of the retrotransposon that will be reverse transcribed into DNA.

LTR-retrotransposons represent approximately 8% of the human genome and are widely represented and active in other organisms like plants and yeast [3]. The main families of LTR retrotransposons are *Bel/Pao*, *Ty1*/*copia*, ERVs and *Ty3/Gypsy* [19]. *Bel/Pao* elements are present in 40 species of metazoan, and around 160 families have been reported [20,21]. *Ty1* and *Ty3* are active in *Saccharomyces cerevisiae*, while *Copia* and *Gypsy* are present in *Drosophila melanogaster* among others [22-25].

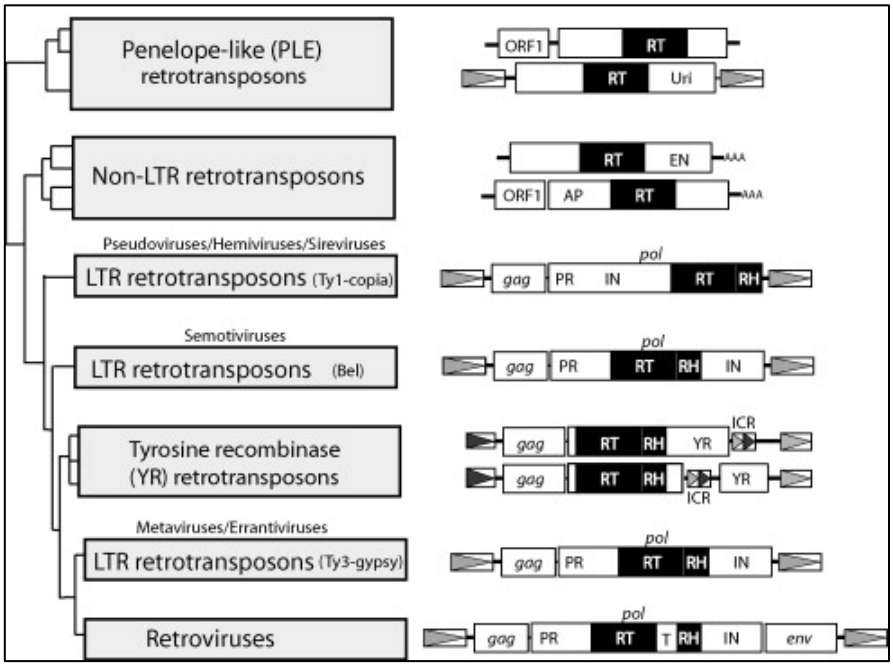


Figure 3. Phylogenetic relationships between retrotransposons based on the reverse transcriptase domain sequence. Different groups of retrotransposons are represented along with their general structure (right). The structure of each retrotransposon is represented by boxes that show the open reading frames. gag, env and pol genes are represented in different groups that either contain them or that contain ORFs that are similar to them. Arrows represent terminal repeats, AAAs represent poly(A) tails. Abbreviations for protein encoding domains: endonuclease (EN), reverse transcriptase (RT), apurinic endonuclease (APE), reverse transcriptase RNase H domain (RH), proteinase (PR), integrase (IN), tether (T), tyrosine recombinase (YR). (Adapted from [16])

Endogenous Retroviruses (ERVs) are related to LTR-retrotransposons (Figure 4). They have a structure similar to them, but they often also contain an additional *env* gene, which codes for the retroviral envelope. ERVs have likely emerged from LTR-retrotransposons by acquisition of an env gene[26,27]. Human ERVs (HERVs) are unable to generate new insertions since all copies have accumulated deleterious mutations during evolution. However, their expression might be involved in certain neurological and autoimmune diseases, as well as in cancer [28-31]. Some families can even form viral particles [32].
In contrast, Mouse Endogenous Retroviruses (MERVs) are still active, even being able to form viral particles [33]. MERVs expression has been reported in the early development of mouse Embryonic Stem cells (mESCs). It has been proposed that they play an important role at these initial stages [34-37].

Non-LTR retrotransposons lack LTRs, they are around 4 to 7 kb long and they retrotranspose through a different mechanism than that of LTR-retrotransposons. Retrotransposition of non-LTR elements begins with the nick of the target site DNA by an endonuclease that is encoded in the ORF of the retrotransposon [38]. Two big groups can be drawn out based on the endonuclease type of non-LTR retrotransposons [39,40]. One group

presents one ORF that encodes for a restriction enzyme-like endonuclease (RLE) in the C-terminal segment and is formed by more ancient non-LTR elements [41]. The second group contains two ORFs, one of them encoding an endonuclease with homology to apurinic/apyrimidinic endonuclease (APE) and constitutes the younger group of these elements [42]. Most of non-LTR retrotransposons integrate in the genome randomly, while others are site-specific, integrating in repetitive sequences that include telomeres, or in ribosomal DNA clusters or multicopy RNA genes [43].

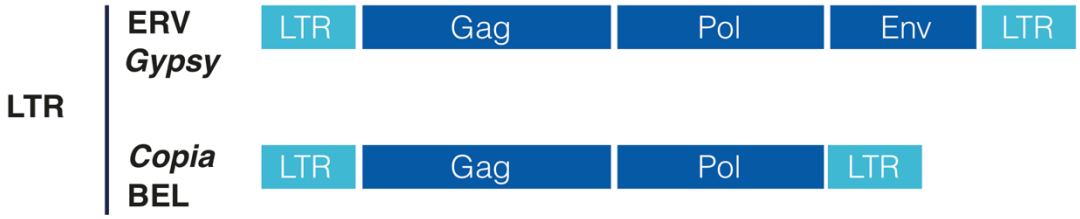


Figure 4. Schematic representation of the structure of LTR retrotransposons. A general structure of LTR retrotransposons is represented. The long terminal repeats (LTRs), characteristic of this type of element, flanks the *gag, pol and env* (for ERV) genes. Gag and pol genes constitute the machinery necessary to mobilize these TEs.

An example of an RLE bearing non-LTR element is R2. It encodes only one open reading frame (ORF) that contains the information for a self-cleaving ribozyme, called R2 protein [44]. R2 elements are present in vertebrates, arthropods and echinoderms, and they insert into the ribosomal cluster of 28S genes [45]. The Long INterspersed Element-1 (LINE-1) is the most studied non-LTR retrotransposon encoding an APE endonuclease. LINE-1 or L1 is the only autonomously active element in the human genome. L1 contains two ORFs, coding for a protein with RNA binding properties and endonuclease/reverse transcriptase activities, respectively [46]. L1 is well represented in the mammalian genomes but appears in many other eukaryotic species [47].

1.1.2.2. Non-autonomous elements use the machinery encoded by other retrotransposons in trans

Some TEs do not encode the machinery necessary for their own mobilization. Instead, these so-called non-autonomous elements rely for their replication on proteins produced by other active TEs.

An example of non-autonomous element in the human and mouse genomes are Short INterspersed Elements (SINEs). In the human and mouse genomes, active families of SINEs principally comprise *Alu* and SVA elements, and B1 and B2 elements, respectively. All of them hijack the retrotransposition machinery of LINE-1 to mobilize their RNA.

Alu elements derive from the 7SL RNA gene and account for 11% of the human genome, being the most represented non-autonomous element with almost a million of copies [3]. Although their presence is important in the genome, most of these copies are currently inactive due to mutations that alter their structure, and only approximately 6000 of them potentially active [48-50]. Alu elements are short sequences of 280 bp in length and contain two 7SL monomers separated by a A-rich linker [51]. The left monomer also contains A and B boxes, the cis-sequences necessary for initiating polymerase III transcription [52] (Figure 5). The right monomer is followed by a poly(A) sequence at its 3' end. Although Alu elements carry an internal Pol III promoter, it is important to highlight that their flanking sequences can greatly influence its transcription and ability to transpose [53]. Alu can be subdivided into J, S and Y families, and among them, the Y and S families are those active in modern humans. Alu has been reported to generate up to 60 disease-causing mutations, as well as rearrangements between nearby copies [54,55].

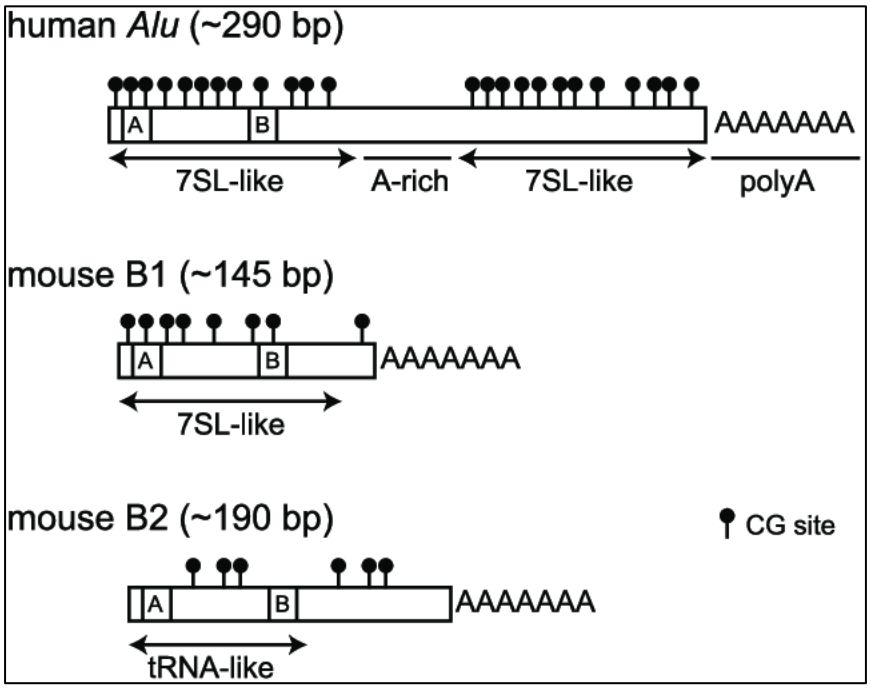


Figure 5. Schematic representation of SINE elements. Human Alu elements are approximately 290 bp long and contain two 7SL monomers separated by an A-rich linker (top), and followed by a poly(A) tail. The left monomer also contains A and B boxes, necessary for polymerase III transcription. B1 elements also contain A and B boxes but are formed by only one 7SL-like monomer. B2 is similar to B1 but contains a tRNA-like monomer instead of a 7SL-like one. The black lollipops represent the position of CpG dinucleotides in Alu, B1 and B2 elements (Adapted from [56]).

SVA (SINE-R, VNTR and Alu) elements are less abundant, representing 0.2% of the human genome with 2700 copies [3]. They are about 2000 bp long and their structure is more complex, with CCCTCT repeats followed by an inverted Alu-like sequence, a variable nucleotide tandem repeat or VNTR a sequence that is homologous to HERVs called SINE-R and a poly(A) sequence [57]. As for LINE-1, SVA elements can cause disease and can

carry transductions from the 3' end flanking sequence, since the polyadenylation signal can be bypassed. This is the case for 10% of SVA insertions [58,59].

B1 elements are formed by a single 7SL monomer [60] and are approximately 135 bp long (Figure 5). As Alu, they contain A and B boxes and end with a poly(A) sequence. They represent 2.5% of the mouse genome with 500,000 copies [60,61].
In contrast to Alu and B1, B2 elements derives from a tRNA gene and are represented by 300,000 copies in the mouse genome. They contain both boxes necessary for polymerase III transcription, and a poly(A) tract in their 3' end [62].
Because each TE family has its own particularities, we have chosen to detail here mostly the biology of L1 elements since they represent the main focus of this doctoral work.

1.2. LINE-1 elements account for 17% of the human genome

Currently, LINE-1 are the only active autonomous element in the human genome, accounting for 17% of its sequence [3]. L1s emerged in the genomic material of mammalian species around 160 my ago [63].
Modern human genomes contain ~500.000 L1 copies, but not all of them remain capable of retrotransposition. In fact, only approximately 100 copies per genome seem to encode functional proteins, and among those, only a few are actually able to retrotranspose due to transcriptional and post-transcriptional controls [64-67]. Retrotransposition-competent L1 elements (RC-L1s) able to mobilize at a higher rate in cultured cell assay are often designated as "hot" LINE-1s.

As underlined above, most L1 copies in the human genome remain inactive and correspond to older versions of L1. These older versions of L1 emerged during the evolution of primates, and many lineages of L1 have been described. Although, the only active family in humans belongs to the L1 PA or primate-specific lineage [63,68,69]. 16 subfamilies of the L1PA lineage have been identified by phylogenetic analysis ($L1PA_{17-1}$), and among these 16, only L1PA1 includes the most recent (and active) L1s, those that belong to the L1-Ta subset [64].
In the mouse genome, although the percentage of LINE-1 derived sequences is similar to the human one (18%), 3,000 L1s are potentially active [4].

1.2.1. LINE-1 replicates through Target Primed Reverse Transcription (TPRT)

A full-length LINE-1 is 6kb long. It contains two untranslated regions (UTRs) at its 5' and 3' end, respectively, two ORFs and it ends with a poly(dA) tract (Figure 6). The two ORFs, ORF1 and ORF2, are separated by a small spacer region of 70 bp. The 5' UTR of L1 (920 bp long) contains a sense Polymerase II-promoter that can drive its transcription [70]. In addition, it includes an ORF0 (not depicted in Figure 6), transcribed from an antisense promoter, that has been proposed to facilitate L1 retrotransposition [71]. The 3' UTR is constituted by a sequence rich in guanine, contains a polyadenylation signal and ends with a poly(dA) tail originating from the reverse transcription of L1 mRNA poly(A) tail. Recent studies showed that the 3'UTR of L1 can form G-quadruplex structures that can promote retrotransposition [72,73].

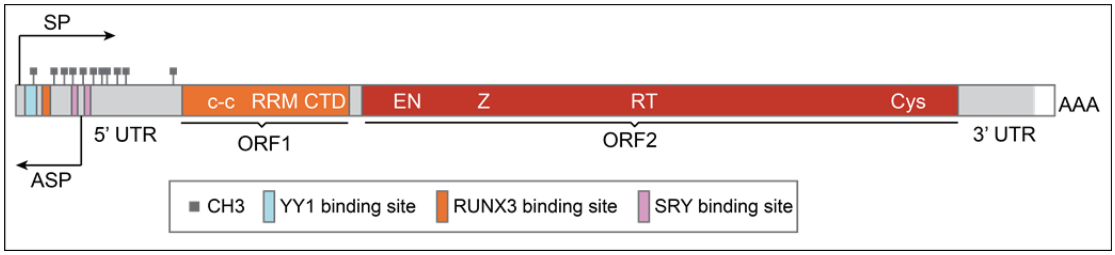


Figure 6. Structure of a LINE-1 element. This schematic representation of LINE-1 elements shows both sense (SP) and antisense (ASP) promoters in the 5' untranslated region (5' UTR) of L1. In this region, the binding sites for Ying Yang 1 protein (YY1), RUNX3 and SRY transcription factors are shown in color. Frequently methylated CpG islands are represented by CH3 grey squares. ORF1 accommodates a coiled-coil domain (c-c), an RNA recognition motif (RRM) and a C-terminal domain (CTD). ORF2 possesses an endonuclease domain (EN), a Z domain that includes a binding sequence for PCNA (Z), a reverse transcriptase domain (RT) and a Cysteine-rich domain (Cys). In its 3' end, L1 includes a 3' UTR and a poly(dA) tail. (Adapted from [74]).

ORF1 codes for an RNA-binding protein, ORF1p. Its translation produces a 40 kDa protein monomer that binds in *cis* to its encoding RNA with high affinity, and forms trimers [75-77]. Structurally, it accommodates an RNA recognition motif (RRM) and a C-terminal domain (CTD). It also contains a coiled coil domain, highly conserved in vertebrates [78,79].

ORF2 encodes a 150 kDa protein with endonuclease (EN) and reverse transcriptase (RT) domains. In its carboxy-terminal part, it also includes a cysteine-rich domain (a C-domain) hypothesized early on to fold into a zinc finger motif, although this has never been demonstrated [80-82]. ORF2p has a very processive reverse transcriptase activity [83-85].

Mouse LINE-1 is overall similar to the human element but also exhibits unique properties. For instance, while it is estimated that 1 in every 100 human births contains a new L1 copy, in mouse this frequency is 1 per every 8 pups [3,4,86-88]. From a structural perspective, mouse

L1s are slightly longer (7kb), but most importantly, their promoter region is completely different from human L1s, as it is formed by an array of ~200bp tandem repeats of variable length [89]. These repeats are used to classify the different mouse L1 families: V, F, A, T_F and G_F [90].

The life cycle of a full-length and replication-competent LINE-1 element starts with one of these copies being transcribed by RNA Polymerase II from its internal sense promoter in the 5' UTR (Figure 7). The promoter activity is located between position 1 to 668 bp of the 5' UTR, the strongest activity being positioned in the first 100 bp [70]. This region can be bound by multiple transcription factors. First, around the nucleotide +12, the Ying Yang 1 protein (YY1) has a binding site on the antisense strand important for proper transcription initiation [91,92]. On the more internal section of the UTR, several binding sites for RUNX3 and SOX (from SRY family) transcription factors have been reported [93,94]. Recently, new transcription factor binding sites were described after analysis of ChIP-seq data from the ENCODE database. CTCF and Myc were identified in several cell lines, sharing some of the sites. Knockdown experiments revealed that Myc might negatively regulate L1 transcription in Human Embryonic Kidney 293 cells (HEK293), and although downregulation of CTCF caused a decrease in the expression of L1 RNA, no direct function was demonstrated [95].
Another important characteristic is the presence of a high density of CpG sites in the 5' UTR, which will be target for methylation and silencing of L1 expression [5].

The transcription start site consists of a GGGGG sequence around the first nucleotide of L1[96]. After transcription is completed, L1 mRNA is translocated from the nucleus to the cytoplasm where it will be translated into ORF1p and ORF2p. ORF1p is more efficiently translated and more abundant than ORF2p, which has never been directly detected so far in an endogenous context [97,98]. L1 replication complexes are formed by several ORF1p trimers that bind in *cis* to their encoding RNA, along with at least one molecule of ORF2p. [82,99]. More precisely, these ribonucleoprotein particle (L1 RNP) contain an estimated ratio of ORF1p:ORF2p between approximately 6:1 and 9:1 [100]. L1 RNPs accumulate in the cytoplasm, forming foci that can colocalize with stress granules [101,102].

It is not clear how L1 RNPs cross the nuclear envelope, but once in the nucleus, they generate insertions through a mechanism known as target primed reverse transcription (TPRT). The details of this process are still not fully understood. Eickbush and collaborators in 1993 introduced the TPRT model studying the site-specific R2

retrotransposon, which helped elucidating some aspects of this process. They observed that the endonuclease encoded by R2 creates a nick in one of the strands of the target DNA, which liberates a 3'-OH group that can be used as a primer by R2 reverse transcriptase. They also showed that the cleavage at the second strand only happens once reverse transcription is terminated [38]. A similar nicking activity was identified in L1 ORF2p, which preferentially cleaves DNA at a consensus sequence 5'-TTTT/AA-3', allowing extension by its RT activity [42,81,85,103-105]. Another relevant study on TPRT was carried out by Boeke's laboratory on human L1, showing that the reverse transcription reaction can be primed by already existing nicked and exposed 3'-OH groups. This suggests that nicking and TPRT do not necessarily happen at the same time [104]. As for L1, it is unknown if and when the cleavage of the second strand happens or how the second strand of L1 cDNA is generated.

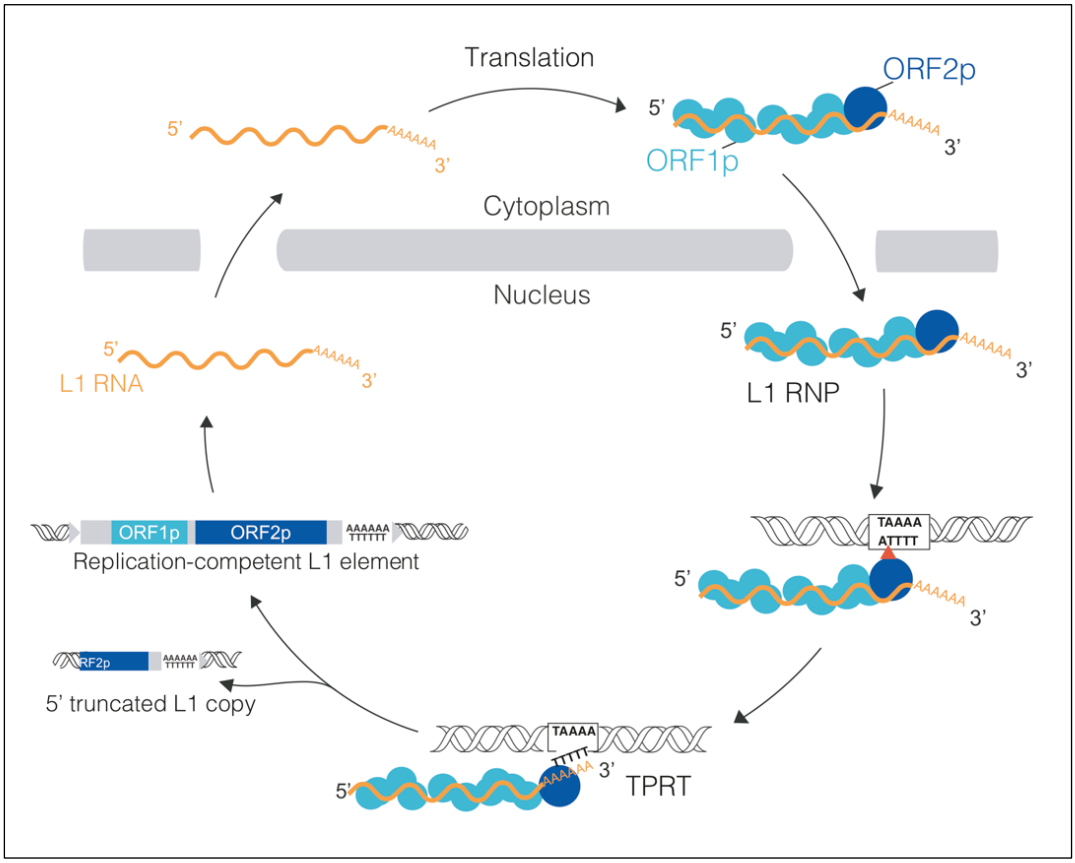


Figure 7. LINE-1 retrotransposition life cycle. The replication of LINE-1 begins after a full-length L1 is transcribed into RNA. In the cytoplasm, L1 RNA (orange) is translated into ORF1p (turquoise) and ORF2p (blue), that will bind in cis to their own RNA, forming the L1 ribonucleoprotein particle (RNP). L1 RNPs translocate back into the nucleus where the endonuclease domain of ORF2p cleaves the genomic DNA at sequences with the consensus 3'-A/TTTT-5'. Then, the reverse transcriptase activity of ORF2p synthesizes L1 DNA using L1 RNA as a template, a mechanism called target-primed reverse transcription (TPRT). Two main outcomes are possible: either a 5' truncated L1 that will not be able to mobilize, or a full-length replication-competent L1 that can initiate new replication cycles, generating new copies and contributing to L1 amplification in the genome. (adapted from [110])

Consistent with these observations, an alternative pathway of L1 insertion was discovered where retrotransposition occurs independently of ORF2p EN activity. In cells defective for the non-homologous end joining (NHEJ) pathway and p53-negative, L1 can insert into pre-existing double-strand breaks or nicks in the genome, a process known as the endonuclease-independent (ENi) retrotransposition [106-109].

1.2.1.1. Genetic consequences of TPRT

Among the genetic consequences of L1 retrotransposition, some can be considered as direct hallmarks of this process at the insertion site [111-115]. First of all, target site duplications (TSDs) are generated at both ends of the insertion site [116]. TSDs are one of the most typical hallmarks of L1 retrotransposition and are normally used in sequencing-based studies to exclude sequencing artefact. Another typical hallmark of retrotransposition is the presence of a poly(dA) tract downstream of the newly inserted L1 and before the TSD [117].

Newly inserted L1s can be either full-length, implying that they will be able to engage into additional rounds of retrotransposition, or can be 5' truncated [113],which actually represents the majority of insertion events. Five-prime truncations can also be considered as a common hallmark of L1-mediated insertions. Among the human-specific L1, only 35% of the insertions are full length, indicating that most of them cannot generate new copies and will not contribute to the expansion of this family [118].

Finally, a process called twin-priming was proposed in order to explain the inversions observed in some L1 insertions (Figure 8). Twin-priming consists on a mechanism where once the first nick of the first strand has primed reverse transcription and L1 cDNA is being synthesized, the EN nicks the second strand and produces a second 3'-OH accessible to RT activity. This second 3'-OH binds L1 RNA internally and cDNA is again synthesized. Then, L1 RNA is removed and polymerization of the remaining DNA is completed [119].

- *Local rearrangements and structural variations*

It is obvious that insertional mutagenesis is the most straight-forward effect of L1 retrotransposition, but the reach of its consequences is broader than the interruption of a coding sequence (Figure 9). One of the most frequent events during L1 mobilization is DNA transductions, where LINE-1 can carry sequences from the flanking regions to their new insertion site [3].

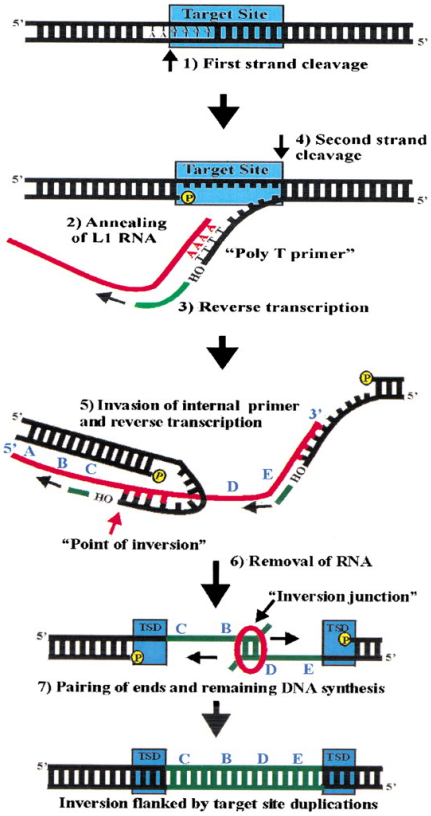


Figure 8. Twin-priming model for LINE-1 retrotransposition resolution. Scheme demonstrating the twin-priming model of L1 insertion. Some LINE-1 insertions have 5' inversions that could not be explained by the initial TPRT model. The twin-priming model proposes that, once a first nick in the first strand has occurred, a second nick is made (4) and can be used to prime reverse transcription again at an internal upstream position within L1 RNA (5), generating an inversion when the RNA is removed and the rest of the DNA is polymerized (6-7). (Adapted from [119])

The 5' transductions originate from the transcription initiated by a promoter upstream of a RC-L1 (Retrotransposition-Competent L1). This will generate an mRNA that contains both the L1 sequence and part of the upstream genomic sequence [120] (Figure 9A).

Transductions can impact the 3' ends of L1 copies as well, since the polyadenylation signal of LINE-1 is rather weak. Thus, RNA polymerase II can use a cryptic polyadenylation signal further downstream in the flanking sequence, omitting the internal L1 signal. As a consequence, the L1 transcript can carry a segment of the flanking sequence, which can be reverse transcribed and integrated, too. The 3' transductions can carry promoters or other cis-acting sequences, as well as exons, and this phenomenon contributes to exon shuffling and to the creation of new genomic sequences [121-124] (Figure 9A).

Other rearrangements are possible due to TPRT, such as target site deletions at the site of the insertion, or even very large deletions (Figure 9B). These events are likely caused by the cleavage of the second strand upstream of the first nick or due to the concomitant action of DNA repair mechanisms [113,115].

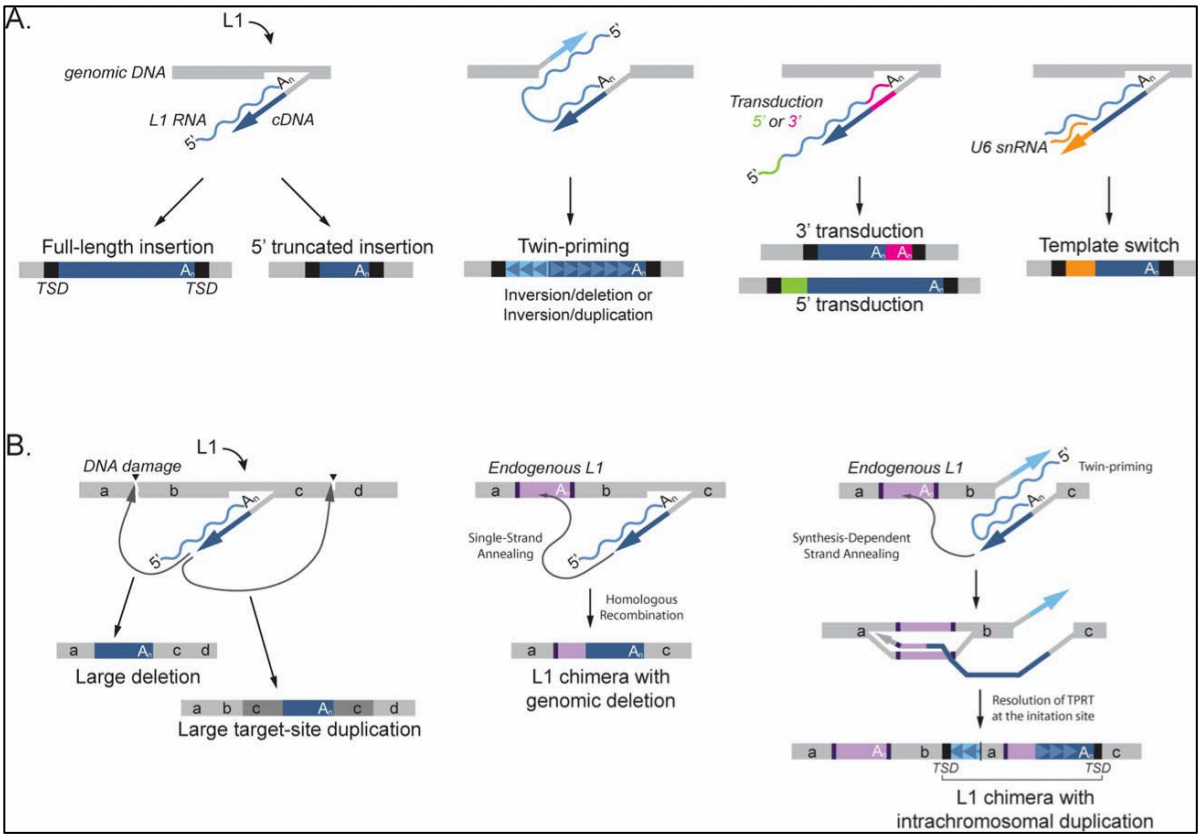


Figure 9. Possible genetic consequences of a LINE-1 insertion. L1 retrotransposition can cause insertional mutagenesis by inserting in a given genomic location and interrupting a gene. However, several other consequences can also result from retrotransposition since TPRT does not take place identically each time. (**A)** An L1 insertion can generate either a full-length L1 or a 5' truncated insertion. Black boxes represent target site duplications (TSD) that are generated after each transposition event and are a hallmark of TPRT. Twin-priming can give rise to inversions by annealing of the upper strand (light blue) with L1 RNA (blue wavy line). When transcription of L1 is driven by a promoter upstream of the L1 sequence, 5' transductions can be generated. 3' transductions are due to the weak polyadenylation signal of L1. Small nuclear RNAs (snRNAs) can be mobilized by LINE-1 when a template switch takes place during synthesis of L1 cDNA. An example is U6 snRNA (wavy orange line). (**B)** "a", "b", "c" and "d" letters illustrate the genomic target sequence where L1 inserts. DNA damage upstream or downstream from the insertion point can give rise to large deletions (loss of segment b) or large target site duplications respectively. During TPRT, when a new L1 (blue box) is inserting near an endogenous L1 (pink box), the new L1 might anneal to the adjacent endogenous one and generate a chimera that finally results in the loss of segment "b". Finally, in a similar situation, but when TPRT resolution results in twin-priming, synthesis-dependent strand annealing can end up on a chimera with target site duplications, an intrachromosomal duplication and an inversion characteristic of twin-priming. (From [132]).

Altogether, these events can lead to disease-causing mutations. For instance, an L1 transduction was reported to be responsible for a case of Duchenne muscular dystrophy [125]. A different case involved a large deletion of 46kb due to L1 retrotransposition in another patient, causing the malfunction of the *Pyruvate Dehydrogenase Complex Component X - PDHX-* gene [126]. In other cases, when more than one L1 or Alu sequences are located close to each other, recombination may occur. A recent study described that non-homologous recombination between L1 and Alu copies deleted part of a gene and caused retinitis pigmentosa [127].

In the last years, more and more studies have described L1-related alterations in cancer [128,129]. The most comprehensive one has analyzed and sequenced 2,954 cancer genomes. Authors found that some L1 insertions caused the deletion of several megabases in some chromosomes or chromosomal translocations, among other structural variations [130].
Lastly, it is important to highlight that ORF2p expression can give rise to double strand breaks (DSB) due to its EN activity, causing DNA damage independently of retrotransposition itself [131].

- *Broader range consequences*

L1 retrotransposition impacts the genome at different scales. As mentioned before, the L1 5' UTR contains numerous binding sites for transcription factors and these can also influence the expression of genes nearby. L1 antisense promoter can drive the expression of adjacent genes, creating chimeric transcripts [71,133,134]. Up to 4% of the human transcriptome could be influenced by L1 antisense promoter [135,136]. The presence of the polyadenylation signal can also cause the premature termination of gene transcription and, in general, L1 insertions can potentially regulate adjacent sequences by modifying their chromatin state [137-141].

1.2.2. LINE-1 activity is regulated by cellular host factors

Strategies that limit the activity of TEs and reduce their impact have been favored by natural selection. Interestingly, mobile DNA and the organisms that carry it are engaged in what is known as an "arms race" [142]. Complementary efforts and approaches have started to unveil the details of L1 interactions with its cellular host. For instance, several interactome analyses based on co-immunoprecipitation of ORF1p or ORF2p have shed light on the possible modulators of L1 activity [100,143-148]. More recently, large genetic screens to identify regulators of L1 retrotransposition by CRISPR-Cas9 or siRNA strategies have been performed [149-151]These efforts have notably identified MORC2 and HUSH complex as regulators of the L1 life cycle, as well as the Fanconi Anemia pathway among other putative candidates [149]. These resources will greatly help forthcoming studies related to L1 regulation. The following section describes the principal mechanisms involved in LINE-1 regulation.

1.2.2.1. Methylation and other epigenetic modifications suppress L1 expression

The transcription of L1 sequences into RNA is the first required step of its life-cycle, and is the target of several cellular pathways that can regulate its mobility. Several epigenetic modifications have been shown to modulate L1 expression.

DNA methylation of 5-methylcytosine (5mC), which typically occurs at cytosine in *CpG* dinucleotides, is the most common strategy of TE repression in higher eukaryotes. Mammalian genomes undergo hypomethylation during two specific periods: the pre-implantation period in embryonic development and gametogenesis [152,153]. Both of these moments imply a window of opportunity for TEs to amplify in the genome, and at the same time a dangerous moment for the genomic instability of the cell [154] (Figure 10). The demethylation of the germline in male mice has been described to activate the expression of TEs, which in consequence triggers the activation of the Piwi-interacting RNA -piRNA- pathway [155].

Focusing on LINE-1, methylation of L1 CpGs in the 5' UTR region is linked with the suppression of its expression [156]. In primordial germ cells, *de novo* methylation of transposable elements requires the DNA methyltransferase 3-like (Dnmt3L) gene, which encodes a co-factor of DNMT-3A and -3B DNA methyltransferases. Mutant mice lacking Dnmt3L present an increase in LINE-1 expression and enter meiotic catastrophe, suggesting that it might be important for heritable restriction of L1 expression in male germ cells [157]. In addition, DNMT3C, a recently discovered methyltransferase, acts in the germline of male mice repressing young TEs through the sequential and coordinated expression of its two isoforms [158]. Furthermore, a de-repression of young LINE-1 elements was described in human embryonic stem cells (hESCs) upon depletion of DNMTs [159]. A similar result was described in human neural progenitor cells (NPCs), where interestingly, L1 elements seem to act as alternative promoters for genes involved in neuronal functions [160]. Lastly, it is important to mention that 5mC methylation is lower in some cancers, coinciding with the expression of TEs [161].

But 5mC methylation is not the only epigenetic mechanism involved in L1 control. Several studies have shown that a family of transcriptional repressors, the Krüppel-associated box (KRAB) zinc-finger (KZNF) protein family and its co-factor KRAB-associated protein-1 (KAP1, also named TRIM28), can silence the expression of TEs in higher vertebrates. KRAB-ZFPs can act as transcriptional repressors by inducing heterochromatinization or DNA methylation in a targeted manner, binding specifically to the TEs, recognized by its Zinc-finger motif [142,162-166]. During murine or human embryogenesis, KRAB-ZFPs expression is regulated according to the transcription pattern of genes containing TEs [35,167-169]. Moreover, when removed, KAP1 can induce the activation of TEs [170-172]. An "arms race" model has been proposed to explain the relationship of LINE-1 and KRAB-ZFP in human ESCs. In this model, expression of new L1s is initially repressed by DNA methylation induced by small RNA mechanisms and eventually KRAB-ZFP that are able to recognize

these new L1s and take over [142,159,173]. A good example of this postulate is the repression of L1PA copies by a member of the KZFN family, KZFN93, and the loss of its binding site in L1Hs copies[142,159]. Jacobs et al. rebuilt the missing ZNF93 binding site on a young L1 copy observing that, upon overexpression of ZNF93, L1 retrotransposition of this element was inhibited [142]. Another model suggests that instead of an arms race, the relationship is rather a domestication where KRAB-ZFPs do not block TE expression in the germline, but they inhibit it in differentiated tissues. A combination of both models is likely to take place [165,174].

Histone modification based on the post-translational methylation of H3K9me3 marks is another important mechanism of LINE-1 repression in ES cells. It has been described that in ES cells, upon acute and almost complete genome demethylation, an epigenetic switch takes over by which histone methylation ensures the regulation of L1 and other TEs [175].
Other epigenetic modifications have been reported to regulate TEs activity. Ten-eleven translocation (TET) enzymes oxidize 5-methylcytosine (5mc) into 5-hydroxymethylcytosine (5hmc). This modification was found to lead to the activation of LINE-1 promoters in pluripotent cells [176-180]. During pre-implantation of the embryo, while general demethylation occurs, an accumulation of 5hmC coincides with expression of L1 [176,181]. In the brain, where L1 is known to express and retrotranspose, 5hmC levels are also high [182]. These findings support the correlation between 5hmC and L1 de-repression.

In summary, these epigenetic mechanisms are coordinated throughout development. 5mC maintains L1 repression until waves of demethylation allow TEs expression, concomitant with the action of TET enzymes, that will oxidize 5mC into 5hmC, leading to L1 activation. From an evolutionary perspective, the expression of young L1 copies is initially restricted by methylation of CpGs through small RNA-based pathways, followed by KRAB-ZNF proteins that take control of these elements after selection. Whether the relationship of L1 with its host genome is an arms race or a domestication, is still debated.

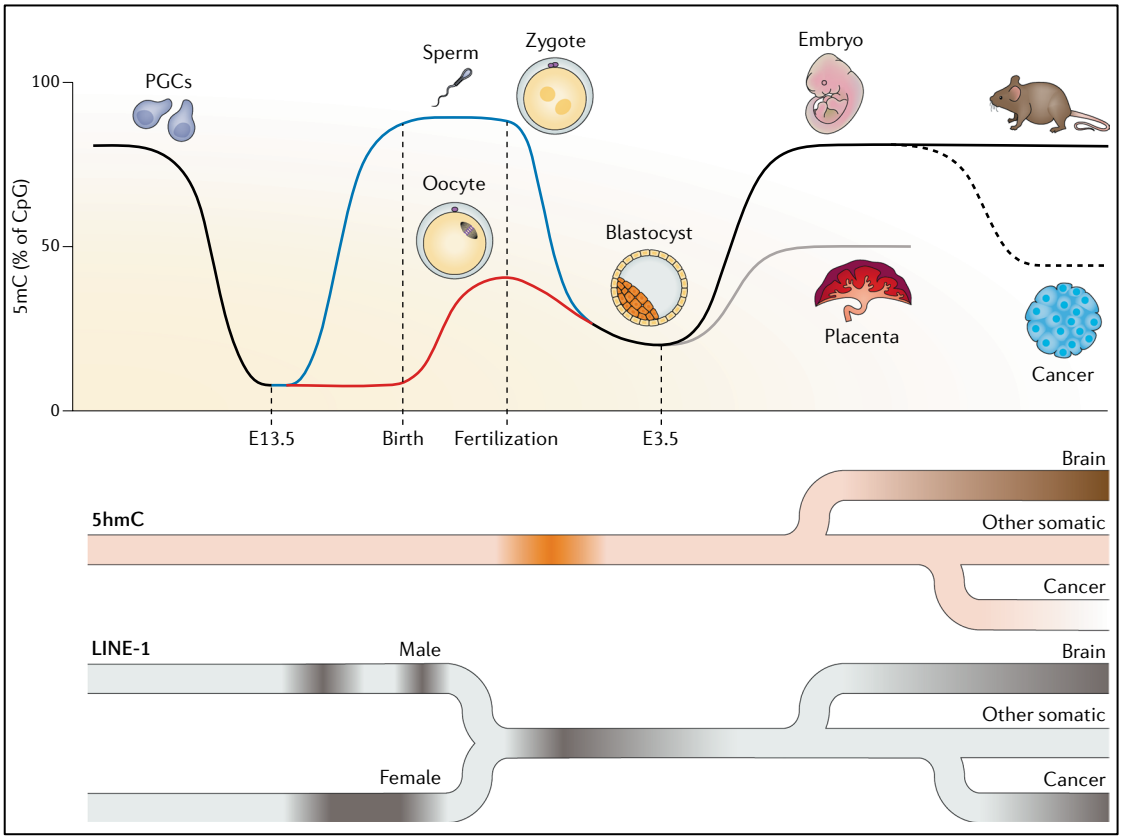


Figure 10. LINE-1 expression, 5mC and 5hmC during mouse development. Two important moments during embryonic development entail epigenetic reprogramming. Loss of 5mC takes places mainly during primordial germ cell (PGCs) migration and after fertilization (upper graph shows levels of 5mC during individual's life). These periods of demethylation are accompanied by an increase of LINE-1 expression, due to loss of 5mC. After fertilization, and before the implantation of the embryo, the levels of 5hmC are elevated at the same time that LINE-1 expression rises. After birth, demethylation can be found in cancer cells, while 5hmC is found in higher levels in the brain. Both coincide with LINE-1 expression. (Adapted from [183]).

1.2.2.2. RNA-based mechanisms participate in L1 regulation

Even if L1 transcription occurs, other mechanisms can still halt the retrotransposition cycle. For instance, a system based on small RNAs (26-31 nt) that guides a series of processing proteins to TE RNA transcripts has been described in several organisms. These proteins are called P-element Induced WImpy testes (PIWI). They are related to the Argonaute family and are important in transposon regulation in the germline of Drosophila [184,185]. The small RNA molecules, or piRNA, are transcribed from specific clusters in the genome. Once they are processed, they lead to the cleavage of transposons' transcripts and generate a second pool of piRNA. This process is called "ping-pong" amplification [186,187]. Similar mechanisms were found in the mammalian germline. In mouse, the murine piwi (MIWI) and miwi-like (MILI) male mutants exhibit meiotic catastrophe and other phenotypical traits of Dnmt3L mutant mice. MIWI and MILI have been proposed to interact with the methylation pathway of TE silencing [188-190]. A study in induced pluripotent stem cells (iPSC) from primates showed an inverse correlation between the expression of PIWIL2 and L1

retrotransposition. Depleting PIWIL2 by shRNA leads to increased retrotransposition, suggesting that this pathway might also be acting during early development of higher primates [191].

The microprocessor complex, formed by an RNase-III (Drosha) and its interactor DGCR8, is responsible for the biogenesis of miRNA. It can also process L1 RNA duplexes presumably formed by sense-antisense transcription in its 5' UTR, leading to L1 RNA degradation and to the repression of L1 and Alu retrotransposition in human cells [192]. This double-stranded RNA (dsRNA) generated from the sense and antisense promoters in the 5'UTR of L1 have also been suggested to be a target of siRNA regulatory mechanisms [193], although these endogenous siRNA (endo-siRNA) more frequently target other TE families such as ERVs, at least in embryonic stem cells [194]. dsRNA can also trigger the activation of RNAse L, that would target and process L1 RNA [195].

1.2.2.3. Other post-transcriptional regulation pathways interfere with L1 RNP stability or with the L1 insertion process

First identified as regulators of retroviral infections, some members of the APOBEC3 (A3) cytidine deaminase family negatively regulate TEs [196-199]. These proteins inhibit L1 and Alu retrotransposition by deamination of L1 DNA, but deamination-independent pathways have also been considered. APOBEC3A and 3B are the more potent L1 inhibitors in the APOBEC3 family [196].

Another interesting example of host factors that regulate L1 are the SAMHD1 and TREX1 proteins which are involved in the Arcadi-Goutières (AGS) syndrome [200]. Both proteins have been described to inhibit L1 when overexpressed. Interestingly, neurons lacking TREX1 expression go through an accumulation of ssDNA derived from LINE-1 activity. This accumulation of dsDNA is toxic and triggers neurotoxic inflammation through type I interferons [201].

L1 RNA accumulation can be reduced by an RNA helicase named Moloney leukemia virus 10 (MOV10). This is an ATP-dependent RNA helicase that was initially described as an antiviral protein in mice [202]. It was later described to associate with L1 RNPs and to impair L1 retrotransposition in cultured cells [147]. Since MOV10 was also shown to co-localize with stress granules, as for L1 RNPs, it is hypothesized that this regulatory pathway may involve the L1 RNA degradation by RNA-based mechanisms [102,147]. Accordingly, reducing MOV10 expression was reported to elevate the levels of L1 RNA and retrotransposition, and both observations were reverted upon MOV10 overexpression [147,203].

Beside negative regulators of LINE-1 activity, several cellular factors are also necessary to help completing L1 replication cycle. This is the case of the polymerase-delta-associated sliding DNA clamp (PCNA). PCNA has an important role in DNA replication, as it serves as a scaffold for different factors implicated in this process, and it ensures processivity of DNA polymerases [145,204]. It interacts with proteins through a motif called the PCNA-interacting peptide (PIP) box, which is present in ORF2p [100]. Indeed, PCNA interacts with ORF2p, and when the PIP motif of ORF2 is mutated, L1 retrotransposition decreases. Moreover, both EN and RT activity are required for the successful interaction of ORF2p with PCNA [100]. This compilation of results implies that PCNA is an important host factor, directly involved in the TPRT process, after initiation of reverse transcription.

1.2.3. LINE-1 activity impacts somatic and embryonic tissues

Despite numerous cellular restriction mechanisms that limit L1 mobility, this retrotransposon does mobilize in certain cell types or at particular developmental stages.

1.2.3.1. L1 retrotransposes in the germline and during early embryonic development

In principle, the propagation of a “selfish DNA” element requires to be inherited, and thus transmitted from one generation to the following ones. Therefore, it is not surprising that ORF1p and the L1 mRNA, which are essential components of the retrotransposition complex, are detected in the germline of both male and female individuals [205-207]. Moreover, primordial germ cells (PGC) from wild type mice bare new and inheritable L1 insertions [88] (Figure 11A). But new retrotransposition events can be transmitted to the progeny even if not taking place in germ cells (Figure 11B). L1 is expressed in human ESCs and these cells contain the set of host factors necessary to enable LINE-1 retrotransposition from an exogenous tagged copy [208-210]. Consistent with the idea that retrotransposition can efficiently occur during embryogenesis before the development of the germline, mosaic individuals can transmit new insertions, including pathogenic ones, to their progeny [211,212]. Human induced pluripotent stem cells (hiPSCs) have also been reported to express L1 RNPs [210] and to accumulate new insertions upon reprogramming [213]. Overall, accumulating new L1 insertion during early development can create somatic and lineage mosaicism, and some can be passed to the next generation, if they occur in cells that will give rise to the future germ line (Figure 11B).

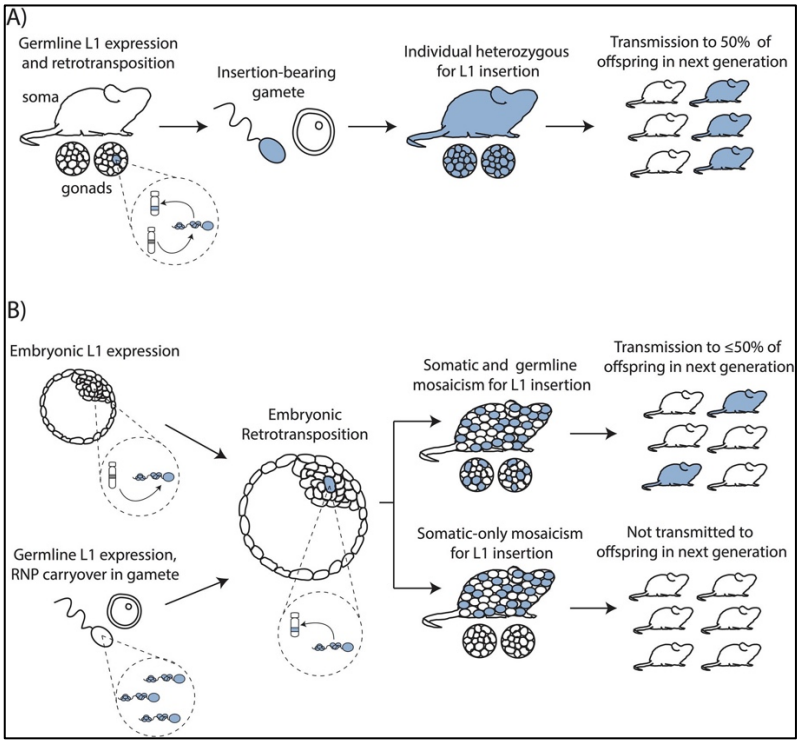


Figure 11. Consequences of L1 retrotransposition in the germline and the embryo. A) LINE-1 retrotransposition in the germline. During the development of the germline, LINE-1 can be expressed and retrotranspose, generating gametes that carry new insertions (blue). If one of these gametes successfully arrives to fertilization, the offspring generated will then carry the insertion in every one of its cells. Since this individual is heterozygous for the L1 insertion, only half of its offspring will have the insertion. **B)** LINE-1 retrotransposition during early embryonic development. L1 expression may occur either during the first stages of embryonic development or in the gametes, pre-fertilization, where L1 RNPs can be carried over into the embryo. Retrotransposition might take place in a pluripotent embryonic cell, generating somatic mosaicism, that will not be inherited by the offspring of the individual. However, if the insertion takes place in an embryonic cell that will contribute to the development of the germline as well, then both somatic and germline mosaic will be generated, and up to 50% of the offspring are likely to inherit the insertion. (From [214]).

1.2.3.2. Somatic L1 retrotransposition is mainly restricted to neural cells in mammals

L1 mobilization is not restricted to embryonic or germ cells. Somatic L1 retrotransposition has been principally described in neuronal cells and neural progenitor cells (NPC). Muotri and colleagues published a seminal study in 2005, showing L1 activity in adult rat NPCs, with an engineered L1 construct, and in the brain of transgenic mice containing a genetically marked L1 element [215]. Following this publication, Coufal et al. described in 2009 that NPCs either isolated from human fetal brain or derived from ESCs are permissive to L1 retrotransposition [216]. In this study, they also developed a quantitative multiplexed PCR assay to assess L1 copy number variations in genomic DNA, that they apply to samples from human adult hippocampus, heart and liver. Interestingly, they found an increase in ORF2 copy number in the brain samples as compared to the other tissues [216]. More recently, a study extended these observations by showing retrotransposition from a plasmid-borne engineered L1 in mature nondividing neurons [217].

Other studies based on next-generation sequencing helped to elucidate the landscape of L1 mosaicism in the brain. Specifically, retrotransposon-capture sequencing (RC-seq) was used and determined that hippocampus from aged patients presented around 7000 somatic L1 insertions, as well as a few thousand of Alu and SVA [218]. Several other studies based on bulk whole tissue sequencing or on single-cell sequencing were performed on brain samples between 2011 and 2016. Although all confirmed the existence of somatic retrotransposition in the mammalian brain, they differed by the number of insertions per cell [182,218-221]. To date, the number of insertions per cell in the brain is still debated, since each method that was utilized has its own caveats or biases [222]. In some of these studies, glial cells were also analyzed to estimate the L1 retrotransposition rate. Upton et al. estimated 6,5 insertions per glial cell, half the rate they recovered in neurons (13,7 insertions per neuron) [182]. Erwin et al., on the other hand, described similar rates for neurons and glial cells, while the rate per cell was 0,58-1 insertions [221]. Therefore, as for neurons, the rate of L1 mobilization in these cells remains unresolved. These results suggest that LINE-1 retrotransposition is an endemic characteristic of neuronal tissue. It has been hypothesized a possible role of L1 as a contributor to neuronal plasticity. Actually, behavioral studies in mice have revealed that early life experience might have an effect in the copy number of L1 in the hippocampus. Mice exposed to lack of maternal care after birth show an accumulation of LINE-1 copies in the hippocampus, which is especially sensitive to environmental stimuli [223]. Likewise, Bundo et al. describe how L1 copy number increases in samples of frontal cortex and in iPSC derived from schizophrenia patients compared to neurotypical individuals [224]. Put together, these studies show that L1 activity in the brain can be a consequence of the interaction of an individual with its environment, conditioning the neuronal phenotype, even acting as modulator of a disease.

Regarding other somatic tissues, fewer studies have been carried out. In 2010, Belancio and colleagues reported low, but detectable, levels of full-length L1 mRNA expression in several human somatic tissues. Stomach, heart, prostate and esophagus seemed to be positive for L1 mRNA, while in cervix, skeletal muscle and spleen, it was not detected [225]. Somatic stem or progenitor cells, purified from tissues or differentiated from hESCs, respectively, were tested for L1 expression and retrotransposition. In contrast to NPCs, levels of retrotransposition and expression of L1 in mesenchymal stem cells, hematopoietic stem cells (HSCs) and progenitor keratinocytes were negligible or below the detection limits [217]. Of note, contrasting results have been obtained by others in HSCs [226].

Transgenic mouse models

Transgenic rodent models have helped to investigate the somatic activity of L1. Kano et al. created transgenic mouse and rat models that contained an L1 sequence driven by its own internal promoter. Surprisingly, they reported that, although high levels of L1 RNA were detected in both the germline and the embryo, the integration of these L1 RNA sequences happened during the embryonic development, making these new copies not inheritable [211]. Another study showed that a tetracycline-controlled mouse codon-optimized L1 would generate mouse with a spotted phenotype, caused by an insertion during melanocyte development that affected its differentiation or migration [227]. Mouse models bearing a GFP-tagged L1 were used to investigate the differences in methylation over embryonic development and its regulation, finding that CpGs in L1 are hypomethylated in the germline and hypermethylated in somatic tissues, and that an intact piRNA pathway is necessary to repress L1 retrotransposition in germ cells [228,229]. Lastly, another study on hematopoietic stem cells (HSCs) used a mouse transgenic L1 model. Authors show that L1 retrotransposition can occur in HSCs. Moreover, they describe that L1 retrotransposition increases upon radiation treatment [226].

In summary, somatic L1 retrotransposition seems to be mostly restricted to brain tissues regarding the evidence available to date. However, no single-cell sequencing study was carried out in cells other than those from the neuronal lineage. Transgenic mice models showed retrotransposition in embryonic or somatic stem cells. However, no study has investigated L1 activity in a specific somatic tissue as thoroughly as it has been done for the brain.

1.2.4. LINE-1 mobility might have an implication on aging and disease

- *LINE-1 in genetic diseases*

LINE-1 activity can be harmful, as exposed before, since its mobility can be followed by several detrimental consequences. The first reported de novo L1 insertions in humans were disease-causing events isolated from two hemophilia A patient, which showed two independent L1 insertion in chromosome X, in the gene encoding for coagulation factor VIII. Interestingly, the progenitor L1 copy of one of them was cloned and used extensively for retrotransposition assays in many studies in the following years [111,141]. Another L1 insertion was described a few years later in the dystrophin gene of a Duchenne muscular dystrophy patient [230]. Nowadays, more than 28 cases of L1-mediated genetic diseases have been described [231,232].

The most common mechanism by which L1 and Alu can cause disease is by insertional mutagenesis; in an exon or in the proximity of an exon, potentially generating splicing variants, as described above [141,230]. However, ectopic recombination between homologous TEs in the same orientation can lead to large deletions. This mechanism involves existing TEs and not de novo insertions. It is frequently observed for Alu insertions. As an example, this was observed for a case of Fanconi anemia: a recombination between Alu copies created a deletion of several exons in the UBE2T (FANCT) gene [233].

Increase LINE-1 activity in the brain has also been observed in both Ataxia telangiectasia (ATX) and Rett syndrome [107,234,235]. A mutation in Ataxia telangiectasia mutated (ATM) protein causes ATX and encompasses accumulation of DNA damage that leads to cell cycle arrest. Knockdown of ATM was shown to increase L1 retrotransposition in NPCs derived from hESCs, thus suggesting the involvement of L1 in the ATX pathology [107]. Rett syndrome is caused by a mutation in the transcriptional repressor methyl CpG binding protein 2 (MeCP2), located in the X chromosome. MeCP2 targets L1 5' UTR, regulating its expression. Muotri et al. described how knocking down MeCP2 increased L1 retrotransposition in the brain of transgenic mice [234], once again proposing the participation of L1 in the disease phenotype.

- *LINE-1 in cancer*

Very strong pieces of evidence have shown the association between L1 expression/retrotransposition, and cancer. Somatic L1 expression is detected in approximately 50% of human tumors [236] and retrotransposition events in 35% of tested tumors [124,237-239 130,240]. It has been shown that both the expression and retrotransposition of L1 is variable depending on the cancer type. For example, lower expression levels than in other malignancies have been found in lung or colon cancer, but the number of retrotransposition events is higher [239]. The contrary happens in ovarian carcinomas [241].

Although the majority of these retrotransposition events do not seem to have a major effect on the development of tumor, at least through insertional mutagenesis, some of them can affect gene function and can generate structural chromosomal changes [124,237]. L1 can generate driver mutations by inactivating tumor-suppressor genes or by increasing the expression of oncogenes. This can be illustrated by the description of a retrotransposition event in the adenomatous polyposis gene (APC) in a colon cancer sample [242]. A recent study, involving several laboratories, analyzed close to 3000 tumor samples to study the insertional pattern of L1 and found that these events can cause major rearrangements in esophageal, lung and head-and-neck carcinomas. Remarkably, they found that L1 activity

can also cause the deletion of megabases in a chromosome or lead to chromosomal translocation [130]. Finally, the impact of L1 in cancer goes beyond insertional mutagenesis. For instance, the expression of LINE-1 can lead to the formation of double-stranded RNA structures (dsRNA), that can result in gene silencing, as reported for the metastasis suppressor tissue factor pathway inhibitor 2 (TFPI2) gene [243] in breast and colon carcinomas. Double stranded RNA (dsRNA) can also trigger the interferon antiviral response in ovarian and skin cancer cells upon DNMT inhibitor treatment [244].

- *LINE-1 in aging*

Many mechanisms contribute to the aging process, affecting cellular homeostasis. An important part of this process at the cellular level, is the alteration of chromatin structure [245]. In fact, a change in the state of heterochromatin was described in several organisms upon aging, like the case of *Drosophila*, where an increase of heterochromatin-based transcripts accompanied by an increase in TE expression was registered in flies as they aged [246]. Similarly, it is also observed *in vitro* in cultured human fibroblasts [247]. This can generate alterations in the transcription pattern of the cell which will eventually alter its functioning [248,249].

L1 activity has been postulated to contribute to the aging phenotype. This concept was based on observations of L1 mobility in cancer or neurodegenerative diseases, the occurrence of which increases with age [54,237,250,251]. As presented above, L1 activity is usually repressed in somatic tissues by several mechanisms. But it appears that, during aging, this repression is less effective, allowing L1 to be expressed and potentially to be mobilized [247,252,253]. Principally, the mechanisms that seem to fail during aging are those related to heterochromatin organization, which are those regulating older L1 elements [201,253,254]. It is possible that other repressive mechanisms, such as RNAi silencing or DNA methylation pathways, are also less efficient, further contributing to the expression of L1, especially of the youngest retrotransposition-competent families [255,256].

Besides L1 mobilization, which has not been demonstrated so far in the context of aging, other L1-related mechanisms have been described. For example, a 2019 study showed that the accumulation of L1 cDNA in the cytoplasm of senescent cells triggers a type-I interferon (INF-I) response in 26 months-old aged mice, generating inflammation in several tissues. The effects of inflammation were reduced upon treatment with a reverse transcriptase inhibitor (RTI) [257]. Another study published the same year, also reported an increase in L1 cDNA that triggered INF-I response in several tissues of SIRT6 knockout mice, that exhibit a progeria phenotype and that could be reverted by treating the animals

with RTI [258]. In this last publication, the authors also describe how the RTI treatment improved other age-related phenotypes in different tissues, like intestine deterioration, improving the animal's lifespan and bodyweight.

These results suggest that L1 activity, either by expression or retrotransposition, might have important effects in aging cells, and its inhibition should be taken into consideration as a possible target to improve or slow down the progression of aging.

1.2.5. Different methods are used to study L1 retrotransposition

The retrotransposition reporter assay and next-generation sequencing-based analyses have been pivotal in our understanding of L1 biology.

1.2.5.1. The retrotransposition assay is based on an engineered and plasmid-borne L1 element.

An L1 retrotransposition reporter assay was first published in 1996 by Moran and colleagues. It consists in transfecting cells with a plasmid containing an RC-L1 element in which a reporter neomycin-resistance cassette was inserted (in the 3' UTR), in opposite direction with respect to L1 [112]. This *neo* reporter cassette, similar to the one published by the group of Thierry Heidmann for the study of retroviruses [259], is interrupted by an intron from the gamma globin gene, which is in the same orientation as L1 (Figure 12). In this configuration, the reporter cassette becomes functional only upon L1 transcription, intron splicing, reverse transcription and integration (or at least long-term maintenance of the extrachromosomal cDNA). G418 is then added to the cell media, and after a couple of weeks the only cells remaining are those in which the reporter cassette was integrated into the genome, rendering them G418 resistant. Each cell colony remaining after the antibiotic selection represents at least one retrotransposition event. As a control, a plasmid containing a mutated version of L1 is often used. A point mutation renders ORF2p RT inactive. In this condition, retrotransposition can only happen if the endogenous ORF2p can reverse transcribe the engineered mutated L1 mRNA in trans [99,260]. This system also provides information about the endogenous L1 machinery of the cell. L1 retrotransposition events can be afterwards recovered by inverse PCR and sequenced, and hallmarks of retrotransposition can be analyzed [114,115,261].

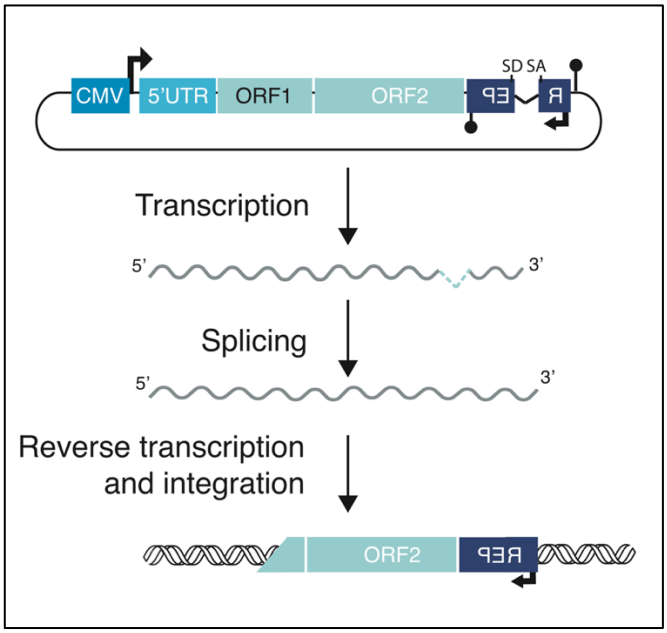


Figure 12. Scheme of the rationale behind the conventional L1 retrotransposition assay. The plasmid scheme shows a LINE-1 sequence, the expression of which is driven by a cytomegalovirus (CMV) promoter, followed by an inverted blasticidin-resistance gene retrotransposition (*mblastI*) cassette. The reporter is interrupted by an intron in the same transcriptional sense of LINE-1. After transcription, splicing of the intron will reconstitute the blasticidin reporter sequence, which after integration into the genome by TPRT will be transcribed by the cell and permit antibiotic resistance. This assay allows a reliable selection of cells containing the insertion of the engineered L1, since only the cells that have undergone retrotransposition and insertion of the BLAST sequence will acquire the resistance.

This approach has been extensively used ever since to study the mechanisms of L1 retrotransposition. It was also crucial in the characterization of active elements in the genome or in understanding the TPRT process. It remains an efficient method to assess how permissive cells are to L1 retrotransposition or the effects of host factors [87,262,263]. Several variants of this assay exist [264], with different antibiotic resistance reporters or even with a GFP cassette, where the readout of the experiment can be recorded by flow cytometry or microscopy [265], or based on Luciferase [266].

1.2.5.2. Sequencing-based and bioinformatic methods allow to record new insertions

Sequencing based-strategies follow two types of approaches (Figure 13):

- Whole-Genome Sequencing: The entire genome is sequenced and specific algorithms are needed to identify L1 insertions, based on split reads and discordant reads. By catching both ends of insertions, it can add another level of validation through the identification of target site duplications (TSDs), flanking the L1 insertion [238,267,268].
- Targeted L1 sequencing: where only L1 junctions with chromosomal DNA elements are first amplified by PCR or captured with probes, and then sequenced [65,238,250,269].

Both approaches require distinguishing between known and unknown insertions and this entails extensive bioinformatic analysis [270,271]. In either approach, the genomic DNA can either be extracted from a single cell or from a population of cells, or tissue [219,220].

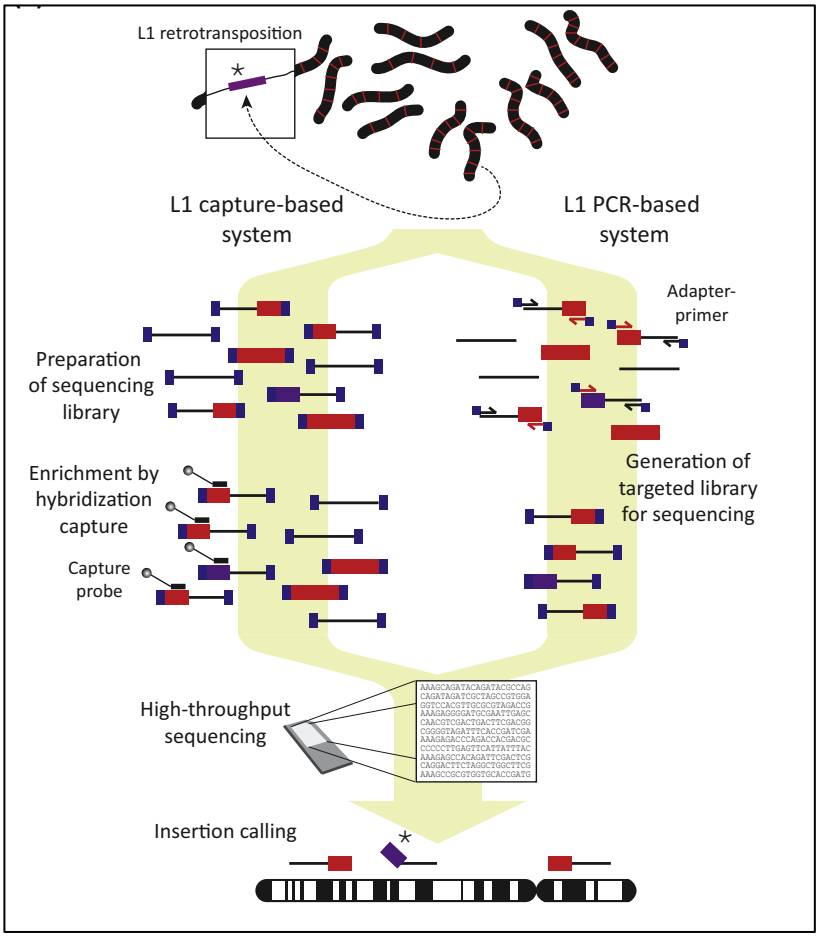


Figure 13. Example of sequencing approaches applied to the mapping of L1 insertions. The scheme illustrates two possible approaches used to map L1 insertions. The purple box represents an unknown L1 insertion, while the red boxes represent reference L1 copies. L1 capture-based system enriches the sample for L1-containing sequences by hybridization and pull-down with specific probes. L1-PCR based systems use primers attached to adapters for amplification of L1-containing sequences. After sequencing and bioinformatic analysis, new insertions can be called. (Adapted from [222]).

It is important to note that many sequencing-based methods are prone to artefacts, especially when applied to repetitive DNA such as L1, or when whole genome amplification is needed as for single-cell sequencing. This has led to major discrepancies in the field. For example, when brain tissue is analyzed, the number of insertions per neuron in mammals or the reality of retrotransposition in Drosophila brain are still a matter of debate [222,272,273].

1.2.6. Evidence of LINE-1 presence in muscle tissue

As we described before, the presence of L1 in somatic tissues other than the brain is still debated and further research is necessary on this topic. However, a limited number of publications have reported the presence of L1 RNA or ORF1p in different types of muscle tissues.

Firstly, Belancio described in 2010 the presence of an ORF2 spliced RNA species by Northern Blot in skeletal muscle human samples, although the signal detected was not strong, and no full-length L1 RNA was observed [225]. De Cecco described the increase of L1 RNA in muscle tissue of aged mice compared to young ones by RT-qPCR on total RNA [274]. Following this study, tissue sections of normal esophagus, including smooth muscles, were reported to be reactive for ORF1p through immunohistochemistry [275]. Lastly, another study by De Cecco in 2019 showed immunofluorescence images where ORF1p is present in 3% of muscle cells from 26 month old mice [257].

Although some of these results suggest that L1 could be active in muscle cells, no actual evidence was provided to show retrotransposition. Therefore, the principal objective of this work has been to determine whether L1 is expressed in muscle cells and tissues, and to investigate if these cells allow the retrotransposition of LINE-1 elements.

1.3. Skeletal muscle

Muscle tissue constitutes the biggest mass of the human body, accounting for 30-50% of its body weight. Of the three principal muscle types (cardiac, smooth and skeletal), skeletal muscle, with approximately 500 different muscles, accounts for most of our whole muscular system. It is in charge of the voluntary movements of the body, making it a very important tissue for optimal physical performance and health [276].

1.3.1. Structure, embryogenesis and regeneration

Structure

Skeletal muscle is composed of myofibers, which are multinucleated contractile cells. Inside myofibers, filaments formed by the proteins responsible for muscle contraction form long cylindric structures called myofibrils. Myofibers form bundles or groups of fibers that are usually surrounded by a layer of connective tissue called perimysium (Figure 14). Each myofiber is formed by single myocytes that differentiate and fuse, forming the multinucleated cell. A single nucleus controls protein synthesis in the area surrounding it, which is called a nuclear domain [278,279]. The plasmatic membrane surrounding a myofiber is called the sarcolemma, which is connected with a complex of proteins that associate to actin filaments inside the myofiber [280](Figure 14). The sarcolemma invaginates into the cell, forming what is known as T-tubules, that are responsible of the transmission of the nerve impulse to the internal parts of the myofiber [281] (Figure 15). The endoplasmic reticulum, called sarcoplasmic reticulum in muscle cells, plays an important role in muscle contraction since it's involved in the maintenance of Ca^{2+} homeostasis [282]. Myofibers are characterized by the spatial organization of their mitochondria, which are abundant and form a network that ensures the proper production of energy and uptake of blood oxygen for muscle contraction [283] (Figure 15).

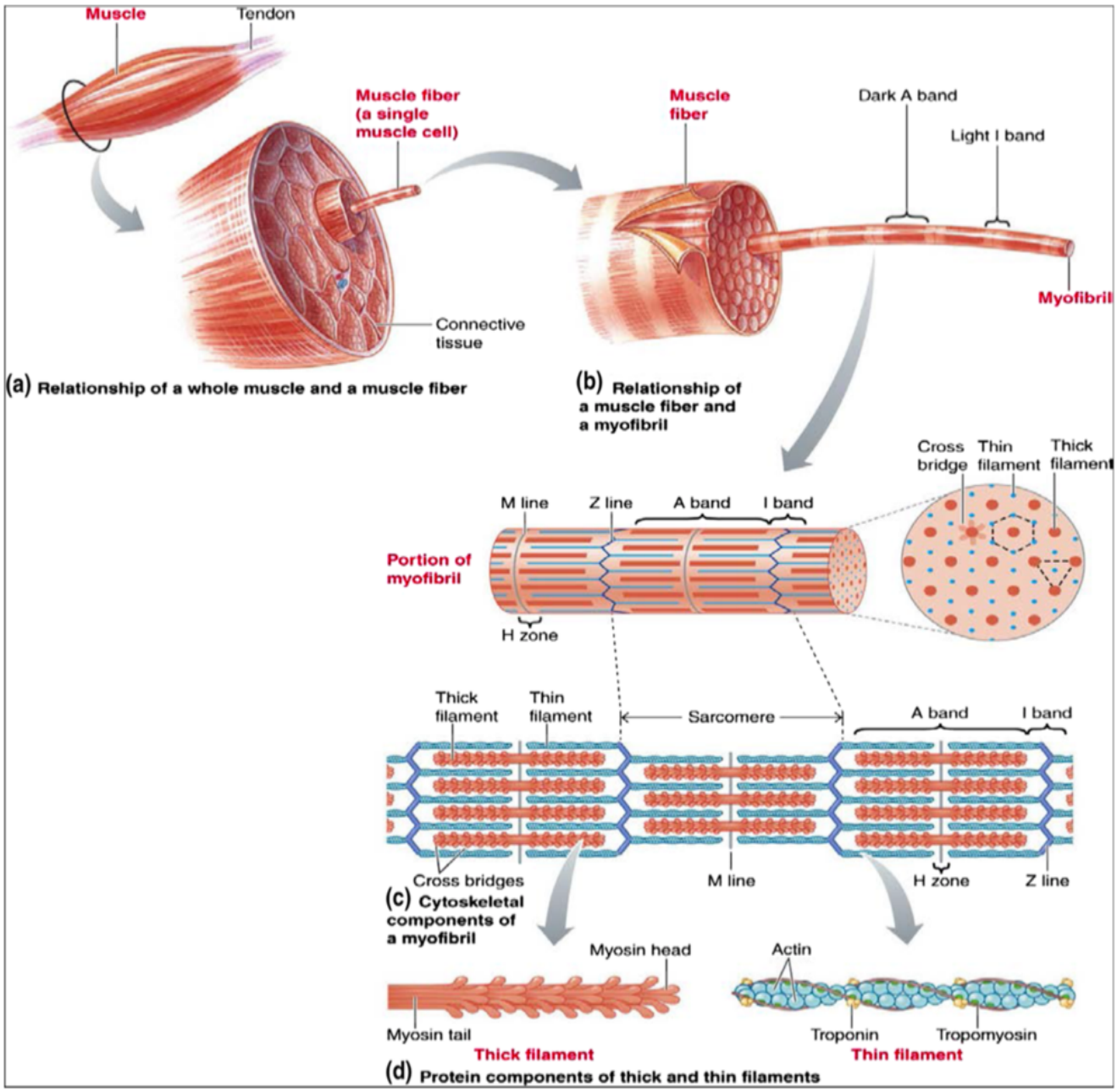


Figure 14. Anatomy of a skeletal muscle. a) The muscle organ is formed by bundles of muscle fibers that are surrounded by connective tissue. The layer of connective tissue surrounding a single muscle is called epimysium. b) Each bundle of fibers is surrounded by the perimysium. Each fiber inside a bundle constitutes one cell or myofiber. Inside the myofiber, protein filaments form the myofibrils. c) Inside each myofibril, we can distinguish several filaments formed mainly by actin (blue) that constitutes the thin filaments, and myosin (red) that constitutes the thick filaments. the bottom part of the figure shows the structure of a sarcomere and how the different filaments are organized. d) Myosin filaments are terminated in a structure called the myosin head, which is involved in the actin-myosin cycling process that allows contraction. On the right, an actin filament presents tropomyosin and troponin (yellow) molecules surrounding it. Both are also involved in the contraction mechanism. (From [277]).

It is important to note that the muscle organ is also constituted by connective tissue, nerve fibers and blood vessels. In fact, a single muscle will have its own nerve, vein and artery that ensures contraction and blood irrigation [285,286].

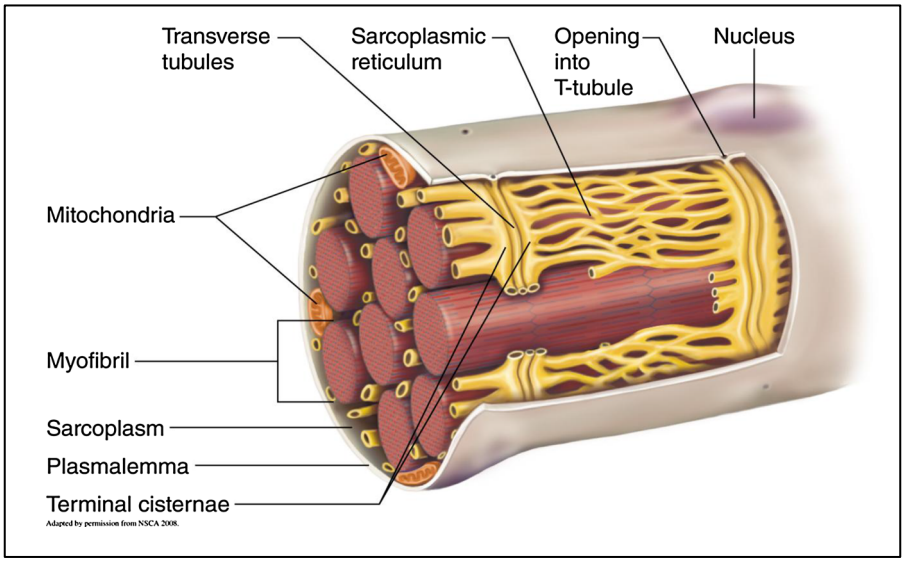


Figure 15. Section of a myofiber. Muscle cells or myofibers are surrounded by a plasmatic membrane called sarcolemma. In the cytoplasm (plasmalemma), myofibrils are surrounded by mitochondria and tubules from the endoplasmic reticulum (sarcoplasmic reticulum) that are crucial for the delivery of Ca^{2+} to the myofibrils for muscle contraction. Transverse tubules (T-tubules) are a system of tubes constituted by an invagination of the plasmatic membrane that ensures the delivery of the Ca^{2+} stimuli to the sarcoplasmic reticulum. The cell nucleus is in the periphery of the cell (upper right). (From [284]).

Function and physiology

Skeletal muscle is an adaptative organ which properties vary depending on mechanical and metabolic functions. Mechanical functions of muscle involve multiple aspects in an individual's life, such as health maintenance or functional independence. They all depend on the ability of muscle to contract [284]. Cellular contraction is executed by a number of proteins present in myofibers that generate mechanical force through chemical interactions. These reactions happen in sarcomeres. Sarcomeres are protein complexes formed by a combination of myosin, actin, troponin, tropomyosin and titin [287-289](Figure 14 c). Contraction of a myofibril starts with its stimulation by a motoneuron, that triggers the release of Ca^{2+} in the cytoplasm [282,290]. Myosin moves down an actin filament and attaches to it, which was called the sliding filament mechanism of muscle contraction, that was discovered in 1954 by Huxley and Hanson [284,291].

The molecular mechanisms behind the interaction of myosin and actin were more recently explained [292,293]. Myosin heads (Figure 14d) contain several hinge segments (S1 region) bend. The S1 region of myosin heads extends, binds, contracts and releases actin filaments upon ATP consumption, in what is named the myosin-actin cycling. In the absence of ATP, myosin remains bound to actin [294].

However, myosin binding sites on actin filaments are covered by tropomyosin. Troponin, a smaller molecule that is bound to tropomyosin, is responsible for the uncovering of the actin binding sites. This process requires the presence of Ca^{2+}, that associates with troponin, inducing the rotation of tropomyosin, and finally exposing actin binding sites [295]. Thus, ATP and Ca^{2+} are the main regulators of muscle contraction, since in absence of one of these molecules, the process would not occur.

Besides its fundamental role in physical movement and posture, skeletal muscle also serves as a regulator of interorgan crosstalk for energy and protein metabolism throughout the body. As such, skeletal muscle acts as a storage of glucose in the form of glycogen and amino acids in the form of proteins. Muscle takes glucose from the blood in order to produce the energy needed for contraction. This uptake is generally stimulated by insulin through translocation of glucose transporter GLUT-4, which increases during exercise [296,297].

Under conditions of starvation, stress, and in some cases disease, muscle is able to hydrolyze proteins and release the amino acids to the blood that will incorporate in the gluconeogenic pathway for energy production [298,299]. They can also be used by other organs for protein synthesis [300]. Additionally, it interacts with other tissues such as liver, adipose tissue or bones through the production of myokines, a type of proteins that have paracrine, autocrine and long-distance endocrine activity [301]. Furthermore, the loss of muscle mass has been shown to alter the response of the organism to stress or disease [300].

It is important to note that muscle is an adaptable tissue, since it is able to respond to more intense effort or a more prolonged one, as well as to different hormonal or metabolic conditions (i.e. hypoxia, starvation, etc) [302]. This capacity to adapt is facilitated by the different types of muscle fibers that compose this tissue. Based on their metabolic performance and the type of myosin heavy chain (MyHC) isoform they express, myofibers can be classified as “fast” (anaerobic metabolism or glycolysis) or “slow” (aerobic metabolism) (Table 1).

Three main types are distinguished:

-Type I (slow-twitch oxidative): These fibers contain a high number of mitochondria, high concentration of myoglobin and a high capillary density. Functionally, they last longer until fatigue.

-Type IIA (fast-twitch glycolytic): Their oxidative capacity is lower due to a lower number of mitochondria; they have less capillary density, but their cross-section area is larger. These fibers are more easily fatigued.

-Type IIX (fast-twitch oxidative glycolytic): these group of fibers have intermediate characteristics, between I and IIA [303].

Fast muscles fatigue sooner than slow fibers, as the conversion of glucose to pyruvate generates less ATP than can be generated by using the rest of central metabolism, ultimately generating CO_2 [303].

Body area	Type of muscle	Example of muscles	Genes	MyHC isoform	Fiber type
Head and neck	Extra occular muscle (EOM)	Rectus superior, rectus lateralis, rectus inferior, rectus medialis, superior oblique or troclearis, inferior oblique	MYH13 MyH14 MyH15	MyHC-EO Slow MyHC MyHC 15	Slow-twitch Slow-twitch Slow-twitch
	Laryngeal	Thyroarytenoid, lateral cricoarytenoid, interarytenoid, posterior cricoarytenoid, cricothyroid	MYH4 MYH1 MYH2	MyHC2B MyHC2X MyHC2A	Fast-twitch Intermediate Fast-twitch
	Middle ear	Stapedius, tensor tympani	MYH2 MYH1	MyHC2A MyHC2X	Fast-twitch Intermediate
	Jaw muscles	Masseter, temporalis, pterygoideus medialis and lateralis, tensor veli palatini, tensor tympani, anterior digastricus, mylohyoideus	MYH6 MYH16 MYH2 MYH1	MyHC M / MyHC alpha (cardiac) MyHC M MyHC2A MyHC2X	Cardiac slow-twitch Fast-twitch Fast-twitch Intermediate
Limb and trunk	Leg	Quadratus lumborum, psoas major, iliacus, tensor fascia latae, adductor longus, adductor magnus, rectus femoris, biceps femoris,gracilistibialis anterior, soleus	MYH4	MyHC2B	Fast-twitch
	Arms	Triceps brachii, biceps brachii, brachioradialis, extensor carpi, abductor pollicis, flexor carpi, flexor digitorium, extensor digitorium	MYH1	MyHC2X	Intermediate
	Back	Rhomboid minor and major, trapezius, deltoid, latissimus dorsi	MYH4	MyHC2B	Fast-twitch
	Chest	Pectorallis minor and major			
	Abdomen	Latissimus dorsi, anterior serratus, extrenal oblique, rectus abdominis	MYH7	Slow MyHC	Slow-twitch
	Thorax	Diaphragm			
	Intercostal	Serratus anterior	MYH2	MyHC2A / MyHC beta (cardiac)	Fast-twitch
	Pelvic floor and prineum	Pubococcygeus, Iliococcygeus, iliacus, levator ani, gluteus maximus, ischiocavernosus, bulbospongiosus			

Table 1. Myosin Heavy Chain isoform expression in different muscle groups. On the left the area of the body, the type of muscles and names of muscles as an example of the diversity of myosin isoforms. Each gene encoding a myosin heavy chain (MyHC) isoform is listed next to the name of the isoform produced. Lastly, on the right, the type of fibers according to their contractile characteristics generated by the expression of each MyHC isoform. The physiological and functional traits of a myofiber are determined by the combination of MyHC isoforms it expresses. Slow-twitch or slow-tonic fibers are characterized by their oxidative metabolism, are used in longer or continuous exercises and have a slower contraction time. Fast-twitch fibers, on the contrary, have a glycolytic metabolism, a faster contraction time and are used in intense exercises. Intermediate fibers have characteristics in between slow and fast fibers, being both oxidative and glycolytic (from [304]).

Embryogenesis

In this section, we will discuss muscle embryogenesis. Many aspects of adult myogenesis resemble or reiterate embryonic morphogenetic episodes. Related signaling mechanisms control the genetic networks that determine cell fate during these two processes. The strong analogy between muscle embryogenesis, differentiation and regeneration is illustrated by the series of transcription factors expressed and the signaling pathways activated [305].

Skeletal muscle has its embryonic origin in the paraxial mesoderm, that flanks the neural tube and the notochord [306]. There, the somites will give place to the myotome, a progenitor of skeletal muscle, where cells already express myosin heavy chain (MyHC), desmin and α-actin typical of myocytes [307]. The cells from the myotome migrate into the limb buds, where muscle is formed after proliferation, fusion and differentiation of myocytes into myofibers [308,309]. These processes are controlled by the expression of Pax3 and other master myogenic regulators such as MyoD, Myf5, MRF4 and myogenin [310-313]. A first wave of primary myofiber formation occurs in the limb buds [314,315]. These cells express MyHC and myosin light chain (MyLC) [316]. A second wave will give rise to secondary fibers that are generated by myocytes fusing to primary fibers [316]. At this stage some cells will start expressing Pax7 to the detriment of Pax3 [317]. $Pax7^+$ cells are proliferative and will constitute the stem cell pool of the individual's skeletal muscle, called satellite cells, that will be later used in adult myogenesis and regeneration [318-320].

Regeneration

It is important to note that muscle tissue holds a strong ability for self-regeneration upon injury, and this capacity extends during the life of an individual. As mentioned before, the stem cell population in the muscle is formed by satellite cells. These cells are located between the basal lamina and the sarcolemma, the plasmatic membrane of myofibers [321]. Fibro-adipogenic progenitors (FAPs), regulatory T cells, macrophages, extracellular matrix and growth factors interact with satellite cells. This environment is called the satellite cell niche, and it is important for the maintenance of stem cell homeostasis and regenerative power [322-326].

Regeneration in muscle is triggered by injury or destruction of myofibers. Satellite cells, the skeletal muscle stem cells, remain in a state of quiescence until their activation is triggered (Figure 16). This will lead to an overexpression of MYOD and Myf5, that will set cells for differentiation into myoblasts [327-329]. Committed myoblasts will downregulate Pax7 and begin the expression of myogenin (MYOG), to later fuse with other myoblasts and form a new multinucleated myotubes [330]. Non-committed cells can downregulate MYOD and Myf5, proliferate and renew the stem cell population for further regeneration [331]. Muscular deterioration and degenerative pathologies have been linked to the loss of satellite cells. This indicates that to maintain a good regeneration potential through the life of an individual, the preservation of the stem cell pool is critical [332,333].

It is interesting to note that muscles with different fiber types present a different stem cell pool size. Muscles with a predominance of slow-twitch fibers show two to three fold higher number of satellite cells than fast-twitch one [334,335].

Thus, skeletal muscle is remarkably plastic adapting continuously to nutrient intake, illness, and physical stress. Changes in adult skeletal muscle also may occur as fiber-type switching, which is influenced by changes in physical activity, loading, nerve stimulation, or hormone and cytokine levels.

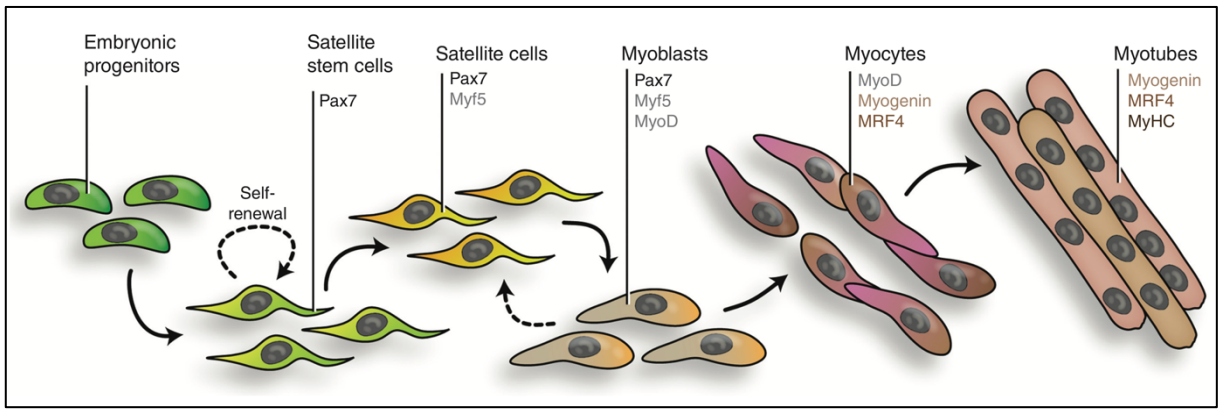


Figure 16. Differentiation of muscle myocytes. Below the basal lamina, quiescent satellite cells express PAX7 and lack MyoD. After injury, satellite cells are activated and initiate the expression of Myf5 and can differentiate into myoblasts, which express Pax7, Myf5 and MyoD. Myoblasts are able to proliferate and after several divisions, they begin to express myogenin and MRF4 in detriment of Pax7 and Myf5. Finally, myocytes fuse, giving rise to myotubes. Myotubes are characterized by the expression of myogenin MRF4 and MyHC (myosin heavy chain) (adapted from [336]).

1.3.2. In vitro culture and isolation

In order to study muscle function and disease, animal models are widely used and have been instrumental in the study of muscular dystrophies [337]. Numerous animal models for muscle research are available, and their importance relies on the interactions between the organ and blood, immune system, connective tissue or neurons, that cannot be replicated *in vitro*. Moreover, the relationship of muscle cells, specifically satellite cells, with the extracellular matrix is important since it is implicated in the process of regeneration [338]. However, *in vivo* models are limited by the difference between human conditions and animal models (systemic features of the *in vivo* environment). Physiological differences and disease phenotypes may vary greatly from mouse to human.

Similarly, most molecular data on myogenesis comes from *in vitro* and *ex vivo* models. Culturing murine or human myocytes *in vitro* is doable and provided the majority of the data about muscle regeneration [339]. Cell culture methods have helped investigating satellite cells, since, due to the limited number of satellite cells in tissue and the possibility of

expanding a population of similar cells in culture, they are convenient for studying genetic and biochemical aspects of muscle processes [338]. Myoblasts are satellite cells that have undergone activation and that continue to proliferate, which allows growing them in culture, even if it is limited to 10-15 passages. These mononucleated muscle cells can be isolated from a freshly extracted tissue biopsy, or derived from induced pluripotent stem cells (iPSC). Nonetheless, as for *in vivo* models, *in vitro* systems are limited as they do not reproduce the interactions with the matrix and the physicochemical properties of the environment. Being aware of these limitations, we chose to use *in vitro* models since we had direct access to human muscle biopsies, from healthy donors as well as myopathic disease patients. The samples were kindly provided by Pr. Sabrina Sacconi, Head of the Peripheral Neurology and Muscle Unit at the CHU in Nice.

Importantly, having access to tissue biopsies is only half of the task, as isolating and culturing primary myoblasts from those samples is very delicate and requires fine tuning. We will focus here on the isolation of myoblast from biopsy sample, as this topic is relevant to this work. This method entails mechanical or enzymatic dissociation of the tissue, until the obtention of mononuclear cells that are then grown in culture. The populations extracted from muscle tissue constitute a mixture of satellite cells, myoblasts, FAP and fibroblasts [326,340]. Fluorescent-activated cell sorting (FACS) is necessary in order to separate muscle cells from other lineages[341]. The most common surface markers used to enrich the samples for myoblast and satellite cells are CD34, α7 integrin, β1 integrin, CD56, CD34, vascular cell adhesion protein 1 (VCAM1), C-X-C chemokine receptor type 4 (CXCR4) and epidermal growth factor receptor (EGFR) [342,343]. It has been described that satellite cells enriched for these markers can be engrafted and successfully improve muscle function in injured muscle of dystrophin-deficient *mdx* mice [344]. In the work presented below, we used CD56-positive cells.

After selection and enrichment, the primary muscle cell population can be grown in culture for several weeks under the appropriate conditions [343]. Culture plate coating is often necessary. Co-culture with fibroblasts was shown to positively influence the differentiation of myoblasts into myotubes, by increasing myotube alignment and adherence [345]. Myoblast culture requires a high concentration of Fetal Bovine Serum (FBS) and the presence of basic fibroblast growth factor (b-FGF) [346,347].

Once in culture, differentiation of myoblasts into myotubes can be induced by starvation and addition of low concentrations of Horse Serum to the media [348]. After several days in differentiation media, myoblasts begin to fuse and to form multinucleated myotubes.

Myotubes will form a myofiber at the end of the differentiation process, at which striations typical of these cells become evident. The differentiation process involves the expression of several transcription factors, which have been described above.

An important aspect to take into consideration relative to skeletal muscle cell culture is that relationship with microenvironment and extracellular matrix (ECM) greatly impacts the functioning of the tissue [349-351]. For instance, ECM stiffness can affect the functioning of satellite cells, such as differentiation, migration and self-renewal, through physical stimuli [352]. Thus, an option for cell culture involves the use of hydrogel polymers (2D) that better mimic the stiffness of *in vivo* conditions. It has been described to accelerate the maturation of myotubes [353]. ECM stiffness can also provide biomechanical stimuli to satellite cells that might alter their regenerative potential during aging [354]. As a consequence of these discoveries, tridimensional cultures and new substrates based on these gel polymers are being implemented in order to imitate the stiffness and spatial *in* vivo conditions[339].

The use of immortalized myoblasts constitutes a powerful tool in the study of muscular disease. Once extracted, myoblasts can be immortalized to avoid senescence and can be kept in culture for a very high number of passages. This is achieved by the expression of human telomerase reverse transcriptase (hTERT) and cyclin-dependent kinase (CDK)-4, which are introduced through a retroviral vector [355,356]. This process permits the establishment of stable cell lines that can be differentiated into myotubes and that express the canonical myogenic pathways [355]. For this reason, we have chosen immortalized myoblasts as one of the models for this study.
C2C12 immortalized myoblasts were generated in 1977 by Yaffe and Saxel [357], and since then they have been used for the study of muscle cells because of their high proliferation rate and their ability to differentiate into myotubes and contract [358-360]. We then included this cell line in our study in order to study L1 activity in mouse myoblasts.

In summary, the proliferative character of myoblasts added to the possibility of myotube generation, makes them a good model for the study of biological processes happening in the muscle. In addition, for muscular diseases such as FSHD, the available mouse models have not been able to properly replicate the disease genetic environment or phenotype [361], since mice lack the DUX4 gene that causes the disorder and in some models, animals were frequently dead at the embryo stage and were not able to achieve adult age [362]. Recently, a new mouse model was proposed, where doxycycline-induced DUX4 expression is specific to the muscle and might be useful for future studies [363]. Nevertheless,

mouse models are greatly time consuming and often more complicated to implement than *in vitro* systems.

1.3.3. Muscular genetic disease

While many muscular diseases exist, it is relevant to the topic of this thesis those that have a genetic cause. Genetic muscular diseases are commonly known as muscular dystrophies. They are characterized by muscle weakness and a loss of muscle mass. Around 50 types of genetic mutations, giving rise to muscular dystrophies, have been described so far. Each disease involves different muscles in the body and has a specific phenotype that seems to highly correlate with a specific genotype [364]. The most common dystrophy in humans is Duchenne muscular dystrophy (DMD), an X-linked recessive disease that is caused by a mutation in the dystrophin gene (Xp21.2–p21.1), which is important to maintain the integrity of the sarcolemma. The result is a disfunction in the dystrophin protein that carries out a deterioration of the muscle tissue and that affects 1 in 3500 males [365-367]. Interestingly, an L1 insertion was described to cause DMD in two individuals. A 600 bp insertion was found in the 3' end of exon 44, which caused said exon to be skipped [230].

1.3.3.1. Facioscapulohumeral dystrophy

Facioscapulohumeral dystrophy (FSHD) is among the 3 most frequent muscle dystrophies. Affecting 1 in 20.000-40.000 people, FSHD consists in a progressive and asymmetrical weakening of the skeletal muscles. It gets its name from the muscles that are affected most often: those of the face (facio-), around the shoulder blades (scapulo-), and in the upper arms (humeral). Although it can eventually affect the legs as well, leaving the patient partially immobilized [368-371]. Onset usually occurs in the teenage years but can begin in childhood or as late as age 50.

FSHD seems to affect type II fibers, since it has been described that the capacity for force-generation is reduced in this type [372]. Moreover, type I fibers predominate in FSHD skeletal muscle samples [373].

Two possible genetic events are responsible for this pathology, giving rise to two types of the disease: FSHD1 and FSHD2. About 95 percent of all cases are FSHD1; the remaining 5 percent are FSHD2. Both types of the disease result from changes in a region of DNA near the end of the chromosome 4 known as *D4Z4*. This array is 10-120 repeats long in healthy individuals.

FSHD1 is caused by a large deletion in the D4Z4 array, located in the 4q35 region, causing a shortening of the array that leaves 10 repeats or less, allowing the expression of the DUX4 homeobox gene, which appears to be toxic for myogenic cells and induces cell death [374,375]. Each subunit of the D4Z4 array contains an ORF for the DUX4 gene, which is normally solely expressed in the germline and in the early embryo [376,377]. When a large portion of the array is deleted, the entire region displays hypomethylation and chromatin relaxation (Figure 17). FSHD2, the least frequent version of the disease, is caused by a mutation in the SMCHD1 gene, a chromatin repressor that is involved in X chromosome inactivation [378]. The depletion of SMCHD1 leads to hypomethylation in the D4Z4 region, ultimately causing the expression of the *DUX4* gene [379]. In both FSHD1 and FSHD2, in order for DUX4 to be expressed, a specific allele must be present. The 4qA haplotype contains the polyadenylation signal necessary to stabilize DUX4 mRNA [380].

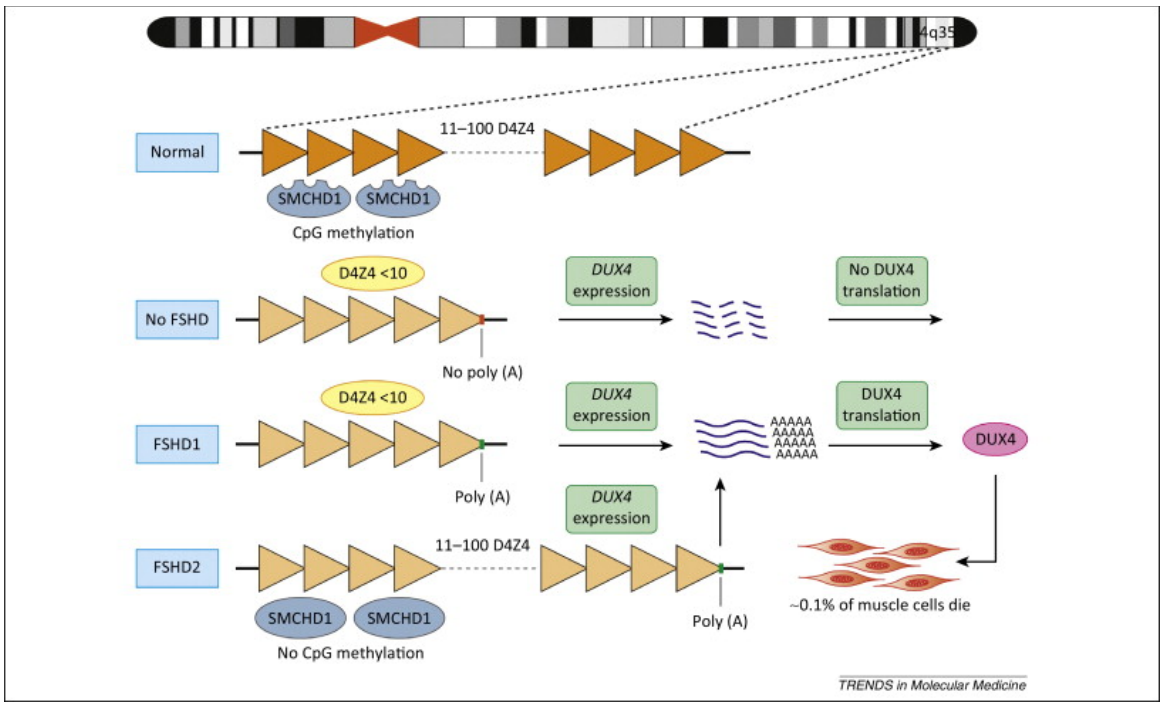


Figure 17. Molecular mechanism of FSHD1 and 2. In a healthy individual, the D4Z4 macroarray located in the 4q35 region of chromosome 4 has between 11 and approximately 100 repeats, and its methylation status depends on the action of SMCHD1 protein. Each repeat of the array contains an ORF for DUX4 protein, that is endogenously expressed only during embryonic development and silenced in the rest of tissues. Only the DUX4 gene located at the last repeat of the array can be expressed in the presence of a specific allele (4qA) that contains the stabilizing polyadenylation signal (PAS). When the D4Z4 array suffers a significant deletion and 10 or less repeats are left, the locus becomes hypomethylated, and DUX4 can be expressed if the contracted array is linked to 4qA allele, leading to FSHD1. If the alternative allele 4qB (red small box – no poly A) accompanies the deletion, no DUX4 expression is possible, and the affected individual is healthy. Alternatively, if the array is longer than 11 or more repeats but a mutation in the SMCHD1 gene impedes the methylation of the region, DUX4 can be expressed in presence of the 4qA allele, causing FSHD2. (From [385])

DUX4 has been described to activate the expression of retrotransposons including HERV-L, MaLR, *Alu* and LINE-1 [381,382]. Recently, both DUX4 and its murine homologue Dux were described to be the principal gene family triggering the zygotic genome activation (ZGA) in the 2-cell embryo of placental mammals [383]. Interestingly, a subsequent study showed that LINE-1 RNA interacts with Nucleolin and Kap1 proteins, repressing Dux and promoting the exit of the 2-cell embryo stage. Additionally, L1 RNA was shown to trigger rRNA synthesis in this context [384]. Taken together, these results suggest that there is a link between LINE-1 and DUX4, making it relevant in the context of this work.

2. Manuscript in preparation

2. Manuscript in preparation

An evaluation of L1 retrotransposon activity in murine and human muscle cells

Paula Peressini[1], Emanuela Repetto[1], Julian Contet[1], Vivien Weber[1], Aurélien Doucet[1], Sabrina Sacconi[1,2], Chloé Féral[1,*], Gael Cristofari[1,*]

[1] University Cote d'Azur, Inserm, CNRS, Institute for Research on Cancer and Aging of Nice (IRCAN), Nice, France

[2] University Cote d'Azur, CHU-Nice, Peripheral Nervous System, Muscle & ALS Department, Pasteur 2 Hospital, Nice, France

* To whom correspondence should be addressed. Email: gael.cristofari@univ-cotedazur.fr. Correspondence may also be addressed to Chloé Féral. Email: chloe.feral@inserm.fr

ABSTRACT

The Long INterspersed Element-1 (LINE-1 or L1) contributes to approximately 20% of human and mouse DNA and is a driver of genome plasticity. L1 activity is tightly regulated at the transcriptional and post-transcriptional levels, limiting potentially damaging new insertions. L1 expression and retrotransposition in normal tissues was found so far mostly restricted to germ cells, early embryonal stages and the brain. It is also detected in nearly half of all epithelial tumors. However, studies of L1 activity in other normal somatic tissues remain limited. Here, we tested L1 expression in skeletal muscle cells and tissues of mouse or human origin. We show that the most abundant L1 protein of the retrotransposition complex, ORF1p, is undetectable under our experimental conditions in muscular cells or tissues, while it is readily detected in cancer cells or testis. Interestingly, we found that L1 is capable of retrotransposition in myogenic cells when expressed from a plasmid or chromosomal copies. In conclusion, while L1 expression is under the limit of detection in skeletal muscle cells and tissues, myoblasts are permissive to retrotransposition, indicating that these cells express all the cellular factors necessary to achieve this process, and do not express significant restriction factors that would block retrotransposition. Thus, L1-mediated genome instability or cellular responses could arise in muscles under environmental or pathological conditions that would unleash L1 expression.

INTRODUCTION

Transposable elements occupy a very large fraction of mammalian genomes. Most of these elements are retrotransposons, which proliferate through an RNA-mediated copy-and-paste mechanism (1). Only a single family of retrotransposons is still able to autonomously mobilize in humans, the Long INterspersed Elements (L1), and more specifically, the L1-HS (L1-Human-Specific) subfamily, which has specifically evolved in hominoids (2). By contrast, more ancient L1 elements shared with apes or even other mammals are immobile. The mouse genome also contains young and retrotransposition-competent L1s, which belong to distinct families, L1-A, L1-Gf and L1-Tf (1, 3). Human and mouse active L1s are related, but differ by their promoters, which likely have different phylogenetic origins (4, 5). Yet, they were extremely successful in expanding and account for approximately 20% of the genome in both species, they rely on the same machinery to achieve retrotransposition and tend to be regulated by the same host defense pathways (6, 7).

Retrotransposition-competent L1s (RC-L1s) are 6 kb DNA sequences with two open reading frames, ORF1 and ORF2. ORF1p is an RNA binding protein(8-10), and ORF2p is an enzyme with endonuclease and reverse transcriptase activities(11, 12). L1 retrotransposition starts by the transcription of a few transcriptionally active L1 loci, which differ among cell types (13-15). The resulting bicistronic L1 mRNA is translated into ORF1p and ORF2p(16, 17), which assemble *in cis* with their RNA to form the L1 ribonucleoprotein particle (L1 RNP), considered as the core of the retrotransposition machinery (8, 18-20). The L1 RNP, or part of it, is imported into the nucleus where it reverse transcribes the L1 RNA directly at the site of integration upon cleavage by its endonuclease activity, a process known as target-primed reverse transcription (TPRT) (21-24). TPRT-mediated L1 insertions have unique hallmarks such as target site duplications (TSDs) and a poly(dA) tract downstream of the newly inserted sequence, resulting from the reverse transcription of L1 mRNA poly(A) tail (25-28). In addition, insertions are often 5' truncated, can carry additional sequences transduced from the 3' genomic flank of the donor element (3' transduction), or can show 5' inversions (29-32). L1 can insert in all regions of the genome, but its integration is restricted by a small motif (consensus 3'-A/TTTT-5') recognized by its endonuclease and reverse transcriptase (33, 34).

Beside direct insertional mutagenesis, L1 activity drives the trans-mobilization of non-autonomous retrotransposons, such as Alu and SVA in humans, or B1 and B2 in mice, and of cellular mRNAs, leading to the formation of processed pseudogenes (35-39). Its activity is also associated with DNA damage, possibly resulting from the combined action of its endonuclease and reverse transcriptase activities and/or from collisions between replication forks and retrotransposition, leading to apoptosis or cellular senescence (40-46). Consistently, primary or cancer cells defective for p53 are more tolerant toward L1 expression (40, 44, 47). Finally,

L1 activity can lead to the synthesis and accumulation of single-stranded cytoplasmic L1 cDNA, by an unknown mechanism, and trigger the activation of interferon signaling pathways, a process observed in cellular senescence (45, 46) or in neuroinflammatory diseases such as the Aicardi-Goutières syndrome (48). Thus, L1 activity is tightly regulated at the transcriptional and post-transcriptional levels, limiting potentially damaging genomic alterations or the accumulation of toxic products.

Several mechanisms can limit L1 RNA accumulation, such as the DNA methylation of L1 promoter or its association with repressive chromatin marks (49-55), or the degradation of L1 RNA by RNAi pathways (56-58). In addition, in recent years, biochemical and genetic screens have dramatically expanded the catalogue of host factors that limit L1 retrotransposition, and those required for this process (59-67). Of note, many of these factors are differentially expressed, suggesting variations in L1 permissiveness across cell types.

Given their mutagenic potential, L1 activity in the germ line can lead to inheritable genetic disease (reviewed in (68, 69)). L1 also drives somatic genome rearrangements during embryogenesis, neural development and tumorigenesis (6). However, current evidence for somatic L1 activity outside of these situations remains limited. Here, we investigate the activity of L1 in skeletal muscle cells and tissues, through L1 ORF1p expression, as well as the ability of myogenic cells to accommodate retrotransposition from engineered L1 elements. We observe that ORF1p is undetectable under our experimental conditions, in mouse and human myoblasts as well as in murine muscle tissues, while readily detected in cancer cells or testis. However, we find that L1 can retrotranspose efficiently in human and mouse myoblasts when expressed from a plasmid or an integrated copy with an inducible promoter. Overall, our results suggest that myogenic cells represent a favorable playground for L1 retrotransposons under environmental or pathological conditions that would trigger their expression.

MATERIALS AND METHODS

Cell lines and mice tissue samples

12U and 12V are myoblasts immortalized with hTERT and CDK4 (70) and were kindly provided by Charles P Emerson Jr. (Univ. of Massachusetts, USA). Human primary myoblasts were obtained from muscle biopsies collected at Nice university hospital. All procedures (biopsies, genetic tests, etc) were performed with the written informed consent of the patients. The biopsy samples were digested by 30 µL of collagenase at 37ºC for 1 h and centrifuged at 1,200 rpm for 10 min. Pellet was collected and resuspended in 10 mL of complete medium (Ham-F10 (Thermofisher 31550031) containing 20% FBS (10270106, Life Technologies), 100 U/mL penicillin and 100 µg/mL streptomycin (15070063, Life Technologies), and 55 ng/mL

dexamethasone (D4902, Sigma)). Suspension was homogenized 10 times with a 5 mL pipet, successively filtered on 100 µm- and 40 µm strainer (FALCON 352360 and 352340), centrifuged for 7 min at 1200 rpm at room temperature (RT). Cell pellet was resuspended in 6 mL of complete medium supplemented with 10 ng/mL recombinant human b-FGF (GF003, Merck Millipore) and seeded in T25 flask at 37°C in 5% CO_2. Medium was changed after 48 h. When 70% confluency was achieved (after a minimum of one week) cells were collected by trypsinization and passed into a T75 flask (P1). Once two T75 flasks were confluent (around 10 millions of cells), cells were collected by trypsinization in a 50 mL centrifugation tube, washed with PBS, incubated with a mouse monoclonal anti-human CD56 antibody (clone B159, BD bioscience, 20 µL/ 1Million cells) for 45 min at RT, centrifuged for 2 min at 1600 x g, cell pellet was washed, resuspended in PBS supplemented with 2% FBS and 0.5 mM EDTA, and sorted by fluorescence-activated cell sorting (FACS) (ARIA III) and further expanded to passage P5.

C2C12 cells are spontaneously immortalized murine myoblasts and were obtained from Sigma. Cultured cells were tested monthly for mycoplasma infection using the MycoAlert Mycoplasma Detection Kit (Lonza, Basel, Switzerland). 12U and 12V were tested positive upon reception and were thus treated by Lookout Mycoplasma Elimination Kit (MP0030-1KT Sigma). Since then, they were tested negative in our routine controls.

MCF-7 breast carcinoma (Sigma 86012803) and HEK-293T embryonic kidney (Sigma 85120602) cell lines were purchased from Sigma. BJ foreskin primary fibroblasts were purchased from ATCC (ATCC-CRL-2522) and immortalized by expression of human telomerase reverse transcriptase (hTERT).

Murine tissue samples

Mouse housing, handling and sacrifice procedure were approved by the Institutional Animal Care and Use Committee at the University of Nice-Sophia Antipolis, Nice, France. Mice C57Bl/6J were purchased from The Jackson Laboratory and arrived at the facility at 8 weeks of age. They were then housed in an enriched environment at 21°C±2°C with food and water available ad libitum. The lights were on between 6:00 and 18:00. Mice were sacrificed by CO_2 inhalation (TEMSega Automate) at 9 weeks and 27 months respectively. Tissue samples were collected from hind limp muscles: soleus, tibialis anterior (TA) and extensor digitorum longus (EDL).

Cell culture

12U and 12V cells were grown in LHCN (lox-hygro-hTERT ("LH"), and Cdk4-neo ("CN")) medium containing Dulbecco's Modified Eagle Medium (DMEM high glucose, GlutaMAX(TM), pyruvate) (31966047, Life Technologies) and 199 Medium mixed at a 4:1 ratio (41150020, Life Technologies), 20% fetal bovine serum (FBS) (10270106, Life Technologies), 50 mM HEPES (15630056, Life Technologies), 0.03 µg/mL zinc sulfate (Z0251-100G, Sigma Aldrich), 14 µg/mL vitamin B12, 55 ng/mL dexamethasone, 2.5 ng/mL hepatocyte growth factor (GF116, Sigma), 10 ng/mL basic fibroblast growth factor (PHG0266, Life Technologies), 100 U/mL penicillin and 100 µg/mL streptomycin. Cells were plated in dishes coated with 0.1% gelatin (G1890, Sigma) and cultured at 37ºC with 5% CO_2. Coating was obtained by adding 0.1% gelatin on plastic plates and incubating for 2h at 37ºC. C2C12 cells were grown in Dulbecco's Modified Eagle Medium (DMEM) (31966047, Life Technologies), supplemented with 10% FBS, 100 U/mL penicillin and 100 µg/mL streptomycin.

Antibodies

For immunoblotting, primary antibodies were directed against human ORF1p (monoclonal mouse antibody, clone 4H1, Merck Millipore, 1:1,000 dilution), mouse ORF1p (rabbit monoclonal antibody, clone EPR21844-108, Abcam, 1:1,000 dilution), α-tubulin (mouse monoclonal antibody, clone T5168, Sigma, 1:10,000 dilution), MyoD (rabbit monoclonal antibody, clone D8G3, Cell Signaling, 1:1000 dilution) or myosin heavy chain (mouse monoclonal antibody directly conjugated to eFluor-660 dye, clone MF20, eBioscience, 1:1,000 dilution). As secondary antibodies for immunoblotting, we used IRDye® Goat anti-Rabbit-680, anti-Rabbit-800, anti-Mouse-680 and anti-Mouse-800 (all from LI-COR Biosciences, 1:10,000).

For immunofluorescence, a primary antibody was directed against human ORF1p (mouse monoclonal antibody, clone 4H1, Merck Millipore, 1:500 dilution). Mouse IgG1k (1447148 Invitrogen) (0.5 mg/mL) was used at the same concentration as primary antibody as a negative control. As secondary antibody, we used a goat anti-mouse Alexa Fluor A488 (A11001 Thermo Fischer, 1:1,000).

Plasmid constructs

JJ101/L1.3: a pCEP4 vector (Invitrogen) containing a full-length human L1 (L1.3) tagged with an antisense *mblastI* retrotransposition indicator cassette (71, 72).

JJ105/L1.3: a plasmid derived from JJ101/L1.3, with a point mutation (D702A) in ORF2 that abolishes its RT activity (71, 72).

pSBtet-Pur (Addgene #60507): a Sleeping Beauty (SB) vector containing a doxycycline-inducible Firefly Luciferase cassette, as well as the reverse tetracycline-controlled transactivator (rtTA) and puromycin resistance genes under the control of a constitutive promoter (73).

pVan924: a derivative of pSBtet-Pur in which the Luc cassette has been replaced by a full-length retrotransposition-competent L1 (LRE3) tagged with an EGFP retrotransposition indicator cassette obtained from 99gfpLRE3 (74). Cassette exchange was obtained by a two step-slice cloning procedure (75): first ORF1 was cloned into the pSBtet-Pur vector digested with SfiI, and subsequently, the ORF2-GFP cassette-polyA was added.

pVan925: a derivative of pVan924 with a dual mutation (R261A and R262A) in ORF1p that abolishes retrotransposition (18). The mutations were introduced by Slice cloning of the mutated ORF1 sequence obtained from plasmid 99gfpLRE3-JM111 (74).

SB100X: plasmid allowing the constitutive expression of a highly active Sleeping Beauty transposase in mammalian cells (Addgene #65487).

pCEP4-TGf21mneoI: a pCEP4 vector (Invitrogen) containing a wild-type mouse L1-Gf element with the *mneoI* retrotransposition indicator cassette (3) and was a kind gift of JL Garcia-Perez (Univ. of Edinburgh, UK).

pWA121: a plasmid based on the pCEP-Puro (puro-marked version of pCEP4) backbone and carrying a synthetic version of mouse L1-spa with codon optimization (also known as mouse ORFeus), and marked with the *mneoI* retrotransposition indicator cassette (76), kindly provided by JD Boeke, NYU, USA.

pVan330: a derivative of pWA121 with a point mutation (D702A) in ORF2 that abolishes its RT activity (23).

Western Blotting

Five millions of cells were lysed in 100 µL of RIPA buffer (10 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.1% SDS, 0.5% NP40, 1mM EDTA pH 8, 0.5% sodium deoxycholate, 1 mM DTT) supplemented with a complete Mini, EDTA-free Protease Inhibitor Cocktail tablet (Roche) at 4ºC for 15 min and centrifuged 15' at 12.000 x g and 4ºC to collect the lysate. Proteins were quantified by the BCA assay (Uptima UP40840A), following manufacturer's protocol. Proteins were separated by gel electrophoresis using precast 4-20% gradient polyacrylamide gels (Biorad) in 1X Tris-Glycine Buffer (1x Tris-Glycine buffer, 0.1% (w/v) SDS), and transferred to a PVDF Immobilon-FL membrane (Merck-Millipore) in transfer buffer (1 X Tris-glycine buffer containing 20% methanol). Membranes were incubated in blocking solution (phosphate-

buffered saline with 0.1% Tween 20 (PBS-T), containing 5% (w/v) fat-free filtered milk) during 1h at room temperature. Subsequently, primary antibody was diluted in blocking solution and membranes were incubated overnight at 4ºC. Membranes were washed 4 times with PBS-T and incubated with a secondary antibody coupled to an infrared fluorochrome diluted in Odyssey blocking buffer (LI-COR Biosciences, Lincoln, NE) for 1h at room temperature. Before detection, the membranes were washed 4 times in PBS-T and a last time in PBS. The detection of the signal was performed with dual-channel Odyssey infrared imaging system (LI-COR Biosciences). If necessary, membranes were stripped with Frogga Bio striping buffer (Frogga Bio ST010) for 15 minutes, washed once with MiliQ water for 5 min and reprobed with the pertinent antibody as described above.

Transfection and retrotransposition assays

Human immortalized myoblasts were electroporated using the Neon Transfection system (Life Technologies) at 1100 V for 1 pulse in a 10 µL Neon tip containing 3 x 10^5 cells and 1 µg of DNA, following manufacturer's instructions. C2C12 were transfected with 4 µg of DNA using 9 µL of Lipofectamine 2000 reagent (Life Technologies) at a density of 1.5 x 10^5 cells/well. For plasmid-based retrotransposition assays (pCEP4-derivatives), selection of retrotransposition events started 3 days after transfection. For SB-based retrotransposition assays, transfected cells with integrated SB vectors were first selected with puromycin-containing media 3 days after transfection. Selective media was changed every 2 days until untransfected control cells were all dead (between 5 and 8 days, depending on the cell type). Then, 5 x 10^4 puromycin-resistant cells were plated in a 10 cm dish and induced with 1 ug/mL of Doxycycline for 3 days. After induction, cells were either selected with blasticidin or GFP-positive cells were counted by flow cytometry, depending on the retrotransposition reporter cassette used. Human immortalized myoblasts were selected with 15 µg/mL blasticidin for 13 days or with 8 µg/mL puromycin for 8 days. C2C12 cells were selected with 1 mg/mL G418 or 1.5 µg/mL puromycin for 7 days. Colonies were fixed and stained with Crystal violet fixing solution [0.5% w/v crystal violet, 10% methanol (v/v), 10% acetic acid (v/v)]. In fluorescence-based assays, GFP expression was analyzed by flow cytometry with dead cells eliminated by gating. Dead cells in the population were marked with 50 µg/mL of propidium iodide (PI) for 5 minutes at room temperature before flow cytometry analysis.

Immunofluorescence

Cells were plated in 12-well plates at a density of 10^5 cells/well on 0.1% gelatin-coated glass coverslips, and fixed the next day with 4% paraformaldehyde (Sigma) for 10 min at

room temperature (RT). Permeabilization was performed in phosphate-buffered saline (PBS) containing 10 mM glycine and 0.5% Triton X-100, for 3 min at RT. Coverslips were blocked for 2h at RT in blocking buffer (PBS supplemented with 1% normal goat serum). Incubation with the primary antibody was performed overnight in a humid chamber at 4ºC. Coverslips were washed 5 times in PBS or PBS containing 10 mM glycine. Secondary antibody was diluted in blocking buffer and incubated for 2 h at 37ºC in a humid chamber. Mounting and DAPI staining was carried out with the use of Vectashield Antifade Mounting Medium (Vector H-1000-NB), according to manufacturer's recommendation. Cells were imaged on a Zeiss LSM confocal microscope, with an objective 40x for Figure 3 and S2 and an objective of 20x for S3.

For shRNA assays, cells were transduced with lentiviral vectors containing shORF1#1 or #2 (13) at a MOI of 10 the day after plating on coverslips. Three days later, cells were selected with 8 µg/mL puromycin for 5 days and fixed. Immunofluorescence was performed as described above.

Quantification with CellProfiler-3.1.8

Twenty images were z-projected by sum slices in Fiji. Then, the intensity of the A488 signal corresponding to ORF1p was quantified with CellProfiler-3.1.8. Briefly, the pipeline included the following steps: identification of primary objects (nuclei) with Otsu thresholding method with the DAPI channel, identification of secondary objects (cells) by the propagation Method with the A488 channel, subtraction of the nuclei from the cell surface in the A488 channel to define cytoplasm, calculation of the median intensity in the A488 channel for each cytoplasm area.

PCR

In retrotransposition assays, genomic DNA was extracted from $3x10^6$ cell pellets 3 days after transfection (plasmid assays) or Dox-induction (SB assays) with the Blood & Cell Culture DNA Mini Kit (Qiagen). DNA was amplified by polymerase chain reaction (PCR) in 25 µL reactions containing 2.5 µL of 10x PCR buffer without $MgCl_2$, 1.5 mM $MgCl_2$, 10 ng of DNA, 10 µM of primers, and 0.5 U Platinum Taq (Invitrogen). For amplification of the *GFP* reporter cassette, forward and reverse primers were 5'-CGTCCATGCCGAGAGTGATC-3' and 5'-GGCAAGCTGACCCTGAAGTTC-3', respectively. For amplification of cytochrome B (loading control), forward and reverse primers were 5'- CCATCCAACATCTCAGCATGATGAAA-3' and 5'-GCCCCTCAGAATGATATTTGTCCTCA-3', respectively. Samples were amplified for 30 cycles ([30 s at 94°C; 30 s at 65°C; 60 s at 72°C] for the GFP reporter cassette and [30 s at

94°C; 30 s at 55°C; 40 s at 72°C] for Cytochrome B). PCR products were separated by 1% agarose gel electrophoresis in 0.5x TBE buffer and stained with 14 mg/µL ethidium bromide.

RESULTS

Endogenous L1 ORF1p expression is undetectable in mouse skeletal muscle cells and tissues

To examine the expression of L1 ORF1p in mouse muscle cells and tissues, we first performed immunoblotting using a rabbit monoclonal antibody (Abcam, EPR21844-108). We validated this antibody by transfecting human HEK-293T cells with a plasmid containing a codon-optimized active mouse L1 (mORFeus). A single protein species compatible with the molecular weight of mouse ORF1p (43 kDa) was detected in as few as 0.25 µg of transfected cell extracts, but not in control cells transfected with the empty vector (Figure 1A, lanes 5-6). Next we examined expression of ORF1p in mouse embryonal carcinoma cells (F9), known to express endogenous L1 (77), as well as in spontaneously immortalized mouse myoblasts (C2C12). While we could easily detect endogenous ORF1p in 0.5 µg of F9 whole cell lysate, no clear signal could be observed in as much as 80 µg of C2C12 whole cell extracts (Figure 1A, compare lanes 1-4 and 7). Thus, ORF1p is undetectable in immortalized murine myoblasts under our experimental conditions, and at least 160 times less expressed than in F9 embryonal cells.

Then, we investigated whether murine skeletal tissue lysates from mice may express ORF1p. Myotubes, which are formed by the differentiation and fusion of myoblasts, represent the predominant cell type in muscle tissues. As previously expected from previous L1 expression or retrotransposition studies (78-82), ORF1p was detected in mouse testis (Figure 1B, lanes 1-4), and barely or not detected in mouse liver (Figure 1B, lane 5), used as positive and negative tissue controls, respectively. In muscle tissues, we repeatedly detected a fuzzy and faint band, but the later migrated slightly faster than ORF1p and overlaps with a very abundant protein specifically observed in the muscle by Ponceau staining of the membrane (star, Figure 1B, lanes 6-11; and Figure S1). We conclude that the antibody slightly cross-reacts with an abundant muscle-specific protein and that ORF1p is undetectable in mouse skeletal muscles, and at least 12-times less expressed than in testis. As recent reports have suggested that L1 expression can be reactivated upon aging (45, 46), we examined whether ORF1p could be upregulated in elderly mice (27 months) as compared to young adult ones (9 weeks), but we observed no difference (Figure 1B, compare lanes 6-8 to 9-11).

Thus, the most abundant L1 protein, ORF1p, is undetectable in whole cell lysates of murine myoblasts or muscle tissues under our experimental conditions.

Endogenous L1 ORF1p expression is undetectable in human immortalized myoblasts

Although human and mouse L1 are relatively conserved in their coding sequences, their promoters are different, and likely have distinct activity profiles (5). Therefore, we examined L1 expression in human immortalized myoblast cell lines using an extensively validated monoclonal antibody directed against human ORF1p (clone 4H1, (83)). MCF7 breast carcinoma and HCT-116 colon cancer cell lines were used as positive and negative controls of ORF1p expression. As previously reported (13), ORF1p was readily detected in whole cell lysate of MCF7 cells, but not in extracts from HCT-116 (Figure 2A, lanes 1-5). However, we could not detect ORF1p expression, neither in 12U nor 12V human immortalized myoblasts (Figure 2A, lanes 6-7). Given the respective quantities of protein extracts loaded, we conclude that ORF1p is expressed in human myoblasts at least 16-times less than in MCF7.

To test the hypothesis that only a minor fraction of cells in the population may express ORF1p and may be undetected in the bulk population by western blot, we assessed ORF1p expression at a single cell level by immunofluorescence. Immortalized human foreskin fibroblasts (BJ-hTERT), which do not express L1 (13), and MCF7 breast carcinoma cells were used as negative and positive controls, respectively. ORF1p was detected at high intensities in most MCF7 cells in the population (Figure 3B and Figure S2). By contrast, BJ-hTERT and most 12V myoblasts were unstained, consistent with the immunoblot results. However, a small fraction of the fibroblasts and myoblasts exhibited a low intensity and punctuated signal (Figure 3B and Figure S2). To test whether this staining corresponds to low-levels of ORF1p expression in a subset of cells or to a low but unspecific signal, we performed shRNA-mediated knock-down of ORF1 using two previously validated shRNAs (13). Consistent with a specific detection of ORF1p in MCF7 cells, the observed staining was strongly decreased in these cells (Figure 3C and S3). This was in sharp contrast to the low signal detected in BJ-hTERT or 12V, which was not lowered by shRNA treatment, suggesting a slight cross-reaction of the antibody with other protein species in these cells.

Altogether, our observations indicate that ORF1p is undetectable in human myoblasts, and at least 16-times less expressed in myoblasts than in MCF7 breast carcinoma cells.

Retrotransposition of an engineered L1 in human immortalized and primary myoblasts

Few untransformed somatic cells can efficiently accommodate retrotransposition from an engineered L1 element (80, 84-87), a process that reflects the balance between host factors that promote L1 retrotransposition, and those that restrict it (59-61, 88). Thus, we first tested

whether human myoblasts are permissive for L1 retrotransposition from plasmid-borne engineered L1 elements(25). In this assay, cells are transfected with a plasmid containing a genetically marked L1 copy containing a retrotransposition reporter cassette, based on the blasticidin- or the neomycin-resistance genes (85). Of note, the reporter is oriented in opposite direction relative to L1, and is interrupted by an intron in the sense orientation (Figure 3A). Consequently, the retrotransposition reporter can only be expressed after L1 transcription, splicing, and reverse transcription, and retrotransposition can be quantified by counting blasticidin- or G418-resistant colonies. Interestingly, although we could not detect endogenous expression in human immortalized myoblasts, we reproducibly observed L1 retrotransposition from an engineered L1 in two distinct immortalized myoblast cell lines with an average rate of ~0.5% of transfected cells (Figure 3B,C), comparable with previous observations in human embryonic stem cells (84). Minimal retrotransposition was observed when a reverse-transcriptase-defective L1 was transfected, in agreement with the notion that trans-complementation by endogenous L1 elements is limited(18) and with our observations that endogenous L1 are undetectable in these cells.

Although skeletal muscle myogenic expression patterns are conserved in immortalized myoblasts (89, 90), it is possible that retrotransposition is facilitated or rendered possible by the immortalization process. Thus, we measured retrotransposition efficiency in human primary myoblasts isolated from muscle biopsies. To do so, CD56-positive myogenic cells were purified from healthy donor muscle biopsies by FACS sorting. We consistently observed retrotransposition events in primary $CD56^+$ myoblasts (Figure 3E). These results suggest that, independently of the immortalization process, myoblasts are permissive to L1 retrotransposition.

Finally, we examined whether mouse myoblasts could sustain L1 retrotransposition. We performed a retrotransposition assay in C2C12 cells using a natural (TGf21) or codon-optimized (mORFeus, pWA121) mouse L1 element marked with a retrotransposition reporter cassette based on the neomycin-resistance gene. As negative control, we used an RT mutant mORFeus (pVan330). Similar to results obtained in human immortalized myoblasts, we observed numerous retrotransposition events. Thus, the permissiveness of myoblasts to L1 retrotransposition is conserved between mouse and human.

High-frequency retrotransposition in human immortalized myoblasts from an inducible and integrated engineered L1 element

The requirement of efficient DNA transfection is a limitation of the plasmid-based retrotransposition assays, especially when studying primary cells, which can be difficult to

transfect. To circumvent this difficulty, we designed a variation of the retrotransposition assay in which an L1 element (LRE3 clone, (91)) with a GFP-based retrotransposition reporter cassette is nested in a *Sleeping Beauty* (SB) transposon vector (Figure 4 A-B). L1 expression is driven by a doxycycline (Dox)-inducible promoter. In addition, the SB vector carries the puromycin-resistance and the reverse tetracycline-controlled transactivator (rtTA) genes, which are constitutively expressed. In this system, the hybrid Sleeping Beauty-L1 (SB-L1) vector is transfected along with a transposase-expressing plasmid. SB transposition leads to the mobilization of the sequence comprised between the two inverted terminal repeats (ITR), and to its insertion into chromosomes. Cells with integrated SB-L1 are selected by puromycin, plated at a defined density, and retrotransposition is induced by Dox for 3 days. Finally, the fraction of GFP-positive cells, corresponding to retrotransposition events, is measured by flow cytometry. As controls, we used an SB-L1 construct with two missense mutations in ORF1p (R261A and R262A) that abolish retrotransposition (18), as well as an SB vector containing a Firefly luciferase gene instead of L1. Interestingly, we detected a very high percentage of GFP-positive cells with the SB-L1 construct upon Dox induction (>30%) in two myoblast cell lines (12V, Figure 4C and D; and 12U, Figure S4). A limited percentage of GFP-positive cells was detected in the absence of induction, suggesting a slight leakage of the inducible promoter under these conditions. By contrast, only background fluorescence was recorded for the mutated SB-L1 construct or the SB-Luc control. To rule out possible false-positives due to dead cell autofluorescence, we stained cells with propidium iodide, which is efficiently excluded from living cells. This analysis confirmed that fluorescent cells were alive and true GFP-positive cells (Figure S5).

To further confirm that the GFP positive cells contain *de novo* insertions, we performed a PCR on genomic DNA to amplify the GFP reporter cassette and distinguish its spliced and unspliced versions (red arrows, Figure 4F and Figure S4). The higher band (unspliced reporter, 1488 bp), corresponds to the integrated SB-L1 vector. It was detected in all L1-containing conditions and is absent in the Luciferase controls. The lower band (spliced reporter, 586 bp), corresponds to the reverse transcribed cassette, and was only detected in cells containing the wild-type SB-L1 transgene upon Dox induction. A faint spliced band was occasionally detected in uninduced samples, consistent with a slight leakage of the inducible promoter. Altogether, these observations indicate that L1 can retrotranspose with high efficiency from integrated hybrid SB-L1 vectors in human myogenic cells.

DISCUSSION

We show here that L1 ORF1p, the most abundant L1 protein in the retrotransposition complex (92), is undetectable in human and mouse cultured immortalized myoblasts, as well as in mouse skeletal muscles, under our experimental conditions. We examined ORF1p expression rather than L1 RNA accumulation since many defective L1 loci unable to retrotranspose and without a full coding potential can be transcribed ((93) in press). In addition, when assessing L1 RNA expression using RT-qPCR or RNA-seq approaches, unit-length L1 transcripts can be confounded with the transcription of the many L1 fragments inserted in genes, or with pervasive transcription ((93) in press). Previous surveys of L1 RNA expression in somatic tissues reported contrasting results. No unit-length L1 RNA could be detected by northern-blot in human skeletal muscle (94), but analysis of RNA-seq data collected across multiple tissues and individuals with a novel algorithm identified muscles among the tissues expressing the highest levels of L1 RNA(95). However, even if this algorithm includes a correction for pervasive transcription, it cannot discriminate unit length transcription from L1-gene co-transcription.

Similarly, ORF1p protein expression was monitored during aging or after physical exercise in human or mouse skeletal muscle, with contrasting results. Examination of mouse skeletal muscles by immunofluorescence against ORF1p suggested that only an extremely low fraction of muscle cells stained positive for ORF1p in mice, even if this fraction increased in old mice as compared to young animals (0.3 vs 3%) (45). Although the precise cell type of stained cells was not defined and slight antibody cross-reactivity cannot be excluded, these observations are consistent with our results showing that ORF1p expression is inexistent or extremely low in bulk extracts of muscle tissues as compared to other tissues (Figure 1). In sharp contrast, another study reported abundant levels of ORF1p in human skeletal muscle irrespective of age or exercise (96). We have not detected ORF1p in human myoblasts (Figure 3). Since the majority of cells in skeletal muscles are terminally differentiated myotubes, rather than their myoblast precursors, we cannot exclude that human L1, in contrast to mouse L1, is upregulated upon differentiation of myogenic cells into myotubes. This possibility will require further investigations in the future. Of note, full length L1 RNA is abundant in the esophagus (94), and ORF1p was detected in epithelial cells, as well as in smooth muscle cells of healthy esophagus by immunohistochemistry(97). In almost every retrotransposition assays performed with an L1 RT mutant, we could detect a few colonies (1 to 3), suggesting a very low - but detectable - level of trans-complementation of this mutant by endogenously expressed L1(18). Overall, our results suggest that L1 elements with ORF1p coding capacity are expressed at extremely low levels in skeletal muscle cells – if expressed at all. L1 expression is locus- and tissue-specific with only a handful of copies being expressed in any

cell type (13). Thus, we cannot exclude that polymorphisms in the internal sequence of L1 among copies and/or individuals alter the epitope recognized by the ORF1p antibodies, while remaining retrotransposition-competent.

L1 retrotransposition can occur in several somatic cells and tissues. Sequencing studies on bulk tissues or single-cells have demonstrated that L1 is mobilized in many epithelial cancers, as well as in the brain (6). Some insertions identified in tumors produced driver mutations (14, 98), and other were identified in a small proportion of cells in the adjacent normal tissue, suggesting that L1 can also retrotranspose at low frequency in normal epithelial cells (99). Tracing *de novo* insertions in small tissue territories or in terminally differentiated cells has been facilitated by single-cell sequencing, but these techniques are also prone to amplification and sequencing artefacts (6). Complementary approaches based on transgenic mice or cultured cells, and on the use of engineered tagged L1 have contributed to refine the somatic tissues or cell types that can accommodate L1 retrotransposition beyond the reproductive system, especially embryonic stem cells, as well as neuronal progenitor cells, and even terminally differentiated neurons(79-81, 84, 87, 100-102). Here, we extended these observations by showing that mouse and human myoblasts can tolerate retrotransposition of engineered L1 elements at levels comparable to what was described in embryonic stem cells (84). These observations imply that myoblasts express all the necessary host factors required for retrotransposition, and no potential L1 restriction factors at levels sufficient to prevent L1 mobilization. We also found that retrotransposition of an engineered L1 can occur in a murine myoblast cell line, indicating that this phenomenon is conserved. Finally, we show that both primary and immortalized human myoblasts are permissive to L1 retrotransposition, suggesting that immortalized myoblasts, which can be more easily obtained and propagated than primary cells, represent a valid model for the study of L1 retrotransposition in the muscle.

Surprisingly, retrotransposition from SB-L1 hybrid transgenes is highly efficient in human myoblasts. High rates of retrotransposition were previously obtained with hybrid adenovirus-L1 vectors, which can infect primary cells, including non-dividing cells (101). Although, our new SB-L1 strategy necessitates transfection, cells with integrated vectors can be selected easily and expanded. An advantage of this experimental system is that the genetically marked L1 is inducible and integrated in the genome. We note that the slight leakage of the inducible promoter, however, could be a limitation for some studies. It will be useful in the future to explore L1 retrotransposition throughout differentiation, from myoblasts to myotubes, or more generally from embryonic stem cells or induced pluripotent stem cells, into any cell type. Thus, the hybrid SB-L1 approach represents an addition to the L1 toolbox complementary to original episomal plasmid-borne L1 constructs and adenovirus-L1 vectors.

Although we did not directly detect L1 expression in muscle cells, retrotransposition of a defective engineered L1 at very low levels suggest that endogenous L1 elements can be expressed in muscle cell under some conditions. Future research will be needed to elucidate whether L1 expression can be upregulated under specific physiological, disease or environmental conditions. As myotubes are non-dividing multinucleated cells formed upon fusion of myoblasts, we note that L1 expression in a single nucleus could in principle compromise the integrity of the entire muscle fiber, a situation particularly relevant to disease states involving progressive muscular degeneration that may reflect the accumulation of L1-mediated alterations or L1 toxic products.

Acknowledgments

We are grateful to J.-L. Garcia-Perez, J. D. Boeke and E. Kowarz for sharing reagents. The authors acknowledge the IRCAN CytoMed, PICMI, GenoMed and Animal core facilities which are supported by grants from the Conseil Général 06, the FEDER, the Ministère de l'Enseignement Supérieur, the Région Provence Alpes-Côte d'Azur, the Canceropole PACA, the foundation ARC and INSERM.

Funding

This work was supported by the Agence Nationale pour la Recherche [Labex SIGNALIFE, ANR-11-LABX-0028-01 to C.F. and G.C.]; by Centre National de la Recherche Scientifique [GDR3546 to G.C.]; and by the Fédération Hospitalo-Universitaire OncoAge [OncoAge to C.F., G.C. and S.S.].

References

1. Huang,C.R.L., Burns,K.H. and Boeke,J.D. (2012) Active transposition in genomes. *Annu. Rev. Genet.*, **46**, 651–675.

2. Mills,R.E., Bennett,E.A., Iskow,R.C. and Devine,S.E. (2007) Which transposable elements are active in the human genome? *Trends in Genetics*, **23**, 183–191.

3. Goodier,J.L. (2001) A Novel Active L1 Retrotransposon Subfamily in the Mouse. *Genome Research*, **11**, 1677–1685.

4. Mears,M.L. and Hutchison,C.A. (2014) The Evolution of Modern Lineages of Mouse L1 Elements. *J Mol Evol*, **52**, 51–62.

5. Sookdeo,A., Hepp,C.M., McClure,M.A. and Boissinot,S. (2013) Revisiting the evolution of mouse LINE-1 in the genomic era. *Mobile DNA*, **4**, 3–15.

6. Faulkner,G.J. and Garcia-Perez,J.L. (2017) L1 Mosaicism in Mammals: Extent, Effects, and Evolution. *Trends in Genetics*, **33**, 802–816.

7. Mandal,P.K. and Kazazian,H.H.,Jr. (2008) SnapShot: Vertebrate Transposons. *Cell*, **135**, 192–192.e1.

8. Hohjoh,H. and Singer,M.F. (1996) Cytoplasmic ribonucleoprotein complexes containing human LINE-1 protein and RNA. *The EMBO journal*, **15**, 630–639.

9. Martin,S.L. and Bushman,F.D. (2001) Nucleic acid chaperone activity of the ORF1 protein from the mouse LINE-1 retrotransposon. *Molecular and Cellular Biology*, **21**, 467–475.

10. Khazina,E., Truffault,V., Büttner,R., Schmidt,S., Coles,M. and Weichenrieder,O. (2011) Trimeric structure and flexibility of the L1ORF1 protein in human L1 retrotransposition. *Nat Struct Mol Biol*, **18**, 1006–1014.

11. Mathias,S.L., Scott,A.F., Kazazian,H.H., Boeke,J.D. and Gabriel,A. (1991) Reverse transcriptase encoded by a human transposable element. *Science*, **254**, 1808–1810.

12. Feng,Q., Moran,J.V., Kazazian,H.H. and Boeke,J.D. (1996) Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell*, **87**, 905–916.

13. Philippe,C., Vargas-Landin,D.B., Doucet,A.J., van Essen,D., Vera-Otarola,J., Kuciak,M., Corbin,A., Nigumann,P., Cristofari,G. and Burns,K. (2016) Activation of individual L1 retrotransposon instances is restricted to cell-type dependent permissive loci. *Elife*, **5**, e13926.

14. Scott,E.C., Gardner,E.J., Masood,A., Chuang,N.T., Vertino,P.M. and Devine,S.E. (2016) A hot L1 retrotransposon evades somatic repression and initiates human colorectal cancer. *Genome Research*, **26**, 745–755.

15. Deininger,P., Morales,M.E., White,T.B., Baddoo,M., Hedges,D.J., Servant,G., Srivastav,S., Smither,M.E., Concha,M., deHaro,D.L., *et al.* (2016) A comprehensive approach to expression of L1 loci. *Nucleic Acids Res.*, **45**, e31–e31.

16. McMillan,J.P. and Singer,M.F. (1993) Translation of the human LINE-1 element, L1Hs. *Proc Natl Acad Sci USA*, **90**, 11533–11537.

17. Alisch,R.S., Garcia-Perez,J.L., Muotri,A.R., Gage,F.H. and Moran,J.V. (2006) Unconventional translation of mammalian LINE-1 retrotransposons. *Genes & Development*, **20**, 210–224.

18. Wei,W., Gilbert,N., Ooi,S.L., Lawler,J.F., Ostertag,E.M., Kazazian,H.H., Boeke,J.D. and Moran,J.V. (2001) Human L1 Retrotransposition: cis Preference versus trans Complementation. *Molecular and Cellular Biology*, **21**, 1429–1439.

19. Kulpa,D.A. and Moran,J.V. (2006) Cis-preferential LINE-1 reverse transcriptase activity in ribonucleoprotein particles. *Nat Struct Mol Biol*, **13**, 655–660.

20. Doucet,A.J., Hulme,A.E., Sahinovic,E., Kulpa,D.A., Moldovan,J.B., Kopera,H.C., Athanikar,J.N., Hasnaoui,M., Bucheton,A., Moran,J.V., *et al.* (2010) Characterization of LINE-1 Ribonucleoprotein Particles. *PLOS Genet*, **6**, e1001150–19.

21. Luan,D.D., Korman,M.H., Jakubczak,J.L. and Eickbush,T.H. (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.

22. Cost,G.J., Feng,Q., Jacquier,A. and Boeke,J.D. (2002) Human L1 element target-primed reverse transcription in vitro. *The EMBO journal*, **21**, 5899–5910.

23. Monot,C., Kuciak,M., Viollet,S., Mir,A.A., Gabus,C., Darlix,J.-L. and Cristofari,G. (2013) The Specificity and Flexibility of L1 Reverse Transcription Priming at Imperfect T-Tracts. *PLOS Genet*, **9**, e1003499–18.

24. Mita,P., Wudzinska,A., Sun,X., Andrade,J., Nayak,S., Kahler,D.J., Badri,S., LaCava,J., Ueberheide,B., Yun,C.Y., *et al.* (2018) LINE-1 protein localization and functional dynamics during the cell cycle. *Elife*, **7**, 210.

25. Moran,J.V., Holmes,S.E., Naas,T.P., DeBerardinis,R.J., Boeke,J.D. and Kazazian,H.H. (1996) High frequency retrotransposition in cultured mammalian cells. *Cell*, **87**, 917–927.

26. Gilbert,N., Lutz-Prigge,S. and Moran,J.V. (2002) Genomic deletions created upon LINE-1 retrotransposition. *Cell*, **110**, 315–325.

27. Symer,D.E., Connelly,C., Szak,S.T., Caputo,E.M., Cost,G.J., Parmigiani,G. and Boeke,J.D. (2002) Human l1 retrotransposition is associated with genetic instability in vivo. *Cell*, **110**, 327–338.

28. Gilbert,N., Lutz,S., Morrish,T.A. and Moran,J.V. (2005) Multiple fates of L1 retrotransposition intermediates in cultured human cells. *Molecular and Cellular Biology*, **25**, 7780–7795.

29. Moran,J.V., DeBerardinis,R.J. and Kazazian,H.H. (1999) Exon shuffling by L1 retrotransposition. *Science*, **283**, 1530–1534.

30. Goodier,J.L., Ostertag,E.M. and Kazazian,H.H. (2000) Transduction of 3'-flanking sequences is common in L1 retrotransposition. *Hum. Mol. Genet.*, **9**, 653–657.

31. Pickeral,O.K., Makałowski,W., Boguski,M.S. and Boeke,J.D. (2000) Frequent human genomic DNA transduction driven by LINE-1 retrotransposition. *Genome Research*, **10**, 411–415.

32. Ostertag,E.M. and Kazazian,H.H. (2001) Twin priming: a proposed mechanism for the creation of inversions in L1 retrotransposition. *Genome Research*, **11**, 2059–2065.

33. Sultana,T., van Essen,D., Siol,O., Bailly-Bechet,M., Philippe,C., Zine El Aabidine,A., Pioger,L., Nigumann,P., Saccani,S., Andrau,J.-C., *et al.* (2019) The Landscape of L1 Retrotransposons in the Human Genome Is Shaped by Pre-insertion Sequence Biases and Post-insertion Selection. *Molecular Cell*, **74**, 555–570.e7.

34. Flasch,D.A., Macia,A., Sanchez,L., Ljungman,M., Heras,S.R., Garcia-Perez,J.L., Wilson,T.E. and Moran,J.V. (2019) Genome-wide de novo L1 Retrotransposition Connects Endonuclease Activity with Replication. *Cell*, **177**, 837–851.e28.

35. Esnault,C., Maestre,J. and Heidmann,T. (2000) Human LINE retrotransposons generate processed pseudogenes. *Nat. Genet.*, **24**, 363–367.

36. Dewannieux,M. and Heidmann,T. (2005) L1-mediated retrotransposition of murine B1 and B2 SINEs recapitulated in cultured cells. *Journal of Molecular Biology*, **349**, 241–247.

37. Dewannieux,M., Esnault,C. and Heidmann,T. (2003) LINE-mediated retrotransposition of marked Alu sequences. *Nat. Genet.*, **35**, 41–48.

38. Hancks,D.C., Goodier,J.L., Mandal,P.K., Cheung,L.E. and Kazazian,H.H. (2011) Retrotransposition of marked SVA elements by human L1s in cultured cells. *Hum. Mol. Genet.*, **20**, 3386–3400.

39. Raiz,J., Damert,A., Chira,S., Held,U., Klawitter,S., Hamdorf,M., Löwer,J., Strätling,W.H., Löwer,R. and Schumann,G.G. (2012) The non-autonomous retrotransposon SVA is trans-mobilized by the human LINE-1 protein machinery. *Nucleic Acids Res.*, **40**, 1666–1683.

40. Haoudi,A., Semmes,O.J., Mason,J.M. and Cannon,R.E. (2004) Retrotransposition-Competent Human LINE-1 Induces Apoptosis in Cancer Cells With Intact p53. *Journal of Biomedicine and Biotechnology*, **2004**, 185–194.

41. Belgnaoui,S.M., Gosden,R.G., Semmes,O.J. and Haoudi,A. (2006) Human LINE-1 retrotransposon induces DNA damage and apoptosis in cancer cells. *Cancer Cell Int.*, **6**, 13–10.

42. Gasior,S.L., Wakeman,T.P., Xu,B. and Deininger,P.L. (2006) The Human LINE-1 Retrotransposon Creates DNA Double-strand Breaks. *Journal of Molecular Biology*, **357**, 1383–1393.

43. Wallace,N.A., Belancio,V.P. and Deininger,P.L. (2008) L1 mobile element expression causes multiple types of toxicity. *Gene*, **419**, 75–81.

44. Ardeljan,D., Steranka,J.P., Liu,C., Li,Z., Taylor,M.S., Payer,L.M., Gorbounov,M., Sarnecki,J.S., Deshpande,V., Hruban,R.H., *et al.* (2020) Cell fitness screens reveal a conflict between LINE-1 retrotransposition and DNA replication. *Nat Struct Mol Biol*, 10.1038/s41594-020-0372-1.

45. De Cecco,M., Ito,T., Petrashen,A.P., Elias,A.E., Skvir,N.J., Criscione,S.W., Caligiana,A., Brocculi,G., Adney,E.M., Boeke,J.D., *et al.* (2019) L1 drives IFN in senescent cells and promotes age-associated inflammation. *Nature*, **566**, 1–33.

46. Simon,M., Van Meter,M., Ablaeva,J., Ke,Z., Gonzalez,R.S., Taguchi,T., De Cecco,M., Leonova,K.I., Kogan,V., Helfand,S.L., *et al.* (2019) LINE1 Derepression in Aged Wild-Type and SIRT6- Deficient Mice Drives Inflammation. *Cell Metabolism*, **29**, 1–38.

47. Wylie,A., Jones,A.E., D'Brot,A., Lu,W.-J., Kurtz,P., Moran,J.V., Rakheja,D., Chen,K.S., Hammer,R.E., Comerford,S.A., *et al.* (2016) p53 genes function to restrain mobile elements. *Genes & Development*, **30**, 64–77.

48. Thomas,C.A., Tejwani,L., Trujillo,C.A., Negraes,P.D., Herai,R.H., Mesci,P., Macia,A., Crow,Y.J. and Muotri,A.R. (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.

49. Deniz,Ö., Frost,J.M. and Branco,M.R. (2019) Regulation of transposable elements by DNA modifications. *Nat Rev Genet*, 10.1038/s41576-019-0106-6.

50. Greenberg,M.V.C. and Bourc'his,D. (2019) The diverse roles of DNA methylation in mammalian development and disease. *Nature Reviews Molecular Cell Biology 2016 17:5*, **20**, 590–607.

51. Sánchez-Luque,F.J., Kempen,M.-J.H.C., Gerdes,P., Vargas-Landin,D.B., Richardson,S.R., Troskie,R.-L., Jesuadian,J.S., Cheetham,S.W., Carreira,P.E., Salvador-Palomeque,C., *et al.* (2019) LINE-1 Evasion of Epigenetic Repression in Humans. *Molecular Cell*, **75**, 590–604.e12.

52. Walter,M., Teissandier,A., Pérez-Palacios,R. and Bourc'his,D. (2016) An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells. *Elife*, **5**, R87.

53. Karimi,M.M., Goyal,P., Maksakova,I.A., Bilenky,M., Leung,D., Tang,J.X., Shinkai,Y., Mager,D.L., Jones,S., Hirst,M., *et al.* (2011) DNA methylation and SETDB1/H3K9me3 regulate predominantly distinct sets of genes, retroelements, and chimeric transcripts in mESCs. *Cell Stem Cell*, **8**, 676–687.

54. Castro-Diaz,N., Ecco,G., Coluccio,A., Kapopoulou,A., Yazdanpanah,B., Friedli,M., Duc,J., Jang,S.M., Turelli,P. and Trono,D. (2014) Evolutionally dynamic L1 regulation in embryonic stem cells. *Genes & Development*, **28**, 1397–1409.

55. Ecco,G., Imbeault,M. and Trono,D. (2017) KRAB zinc finger proteins. *Development*, **144**, 2719–2729.

56. Yang,N. and Kazazian,H.H. (2006) L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells. *Nat Struct Mol Biol*, **13**, 763–771.

57. Berrens,R.V., Andrews,S., Spensberger,D., Santos,F., Dean,W., Gould,P., Sharif,J., Olova,N., Chandra,T., Koseki,H., *et al.* (2017) An endosiRNA-Based Repression Mechanism Counteracts Transposon Activation during Global DNA Demethylation in Embryonic Stem Cells. *Cell Stem Cell*, **21**, 694–703.e7.

58. Heras,S.R., Macias,S., Cáceres,J.F. and Garcia-Perez,J.L. (2014) Control of mammalian retrotransposons by cellular RNA processing activities. *Mobile Genetic Elements*, **4**, e28439–7.

59. Pizarro,J.G. and Cristofari,G. (2016) Post-Transcriptional Control of LINE-1 Retrotransposition by Cellular Host Factors in Somatic Cells. *Front. Cell Dev. Biol.*, **4**, 210–9.

60. Goodier,J.L. (2016) Restricting retrotransposons: a review. *Mobile DNA*, **7**, 1–30.

61. Liu,N., Lee,C.H., Swigut,T., Grow,E., Gu,B., Bassik,M.C. and Wysocka,J. (2018) Selective silencing of euchromatic L1s revealed by genome-wide screens for L1 regulators. *Nature*, **553**, 228–232.

62. Mita,P., Sun,X., Fenyö,D., Kahler,D.J., Li,D., Agmon,N., Wudzinska,A., Keegan,S., Bader,J.S., Yun,C., *et al.* (2020) BRCA1 and S phase DNA repair pathways restrict LINE-1 retrotransposition in human cells. *Nat Struct Mol Biol*, **27**, 179–191.

63. Taylor,M.S., Altukhov,I., Molloy,K.R., Mita,P., Jiang,H., Adney,E.M., Wudzinska,A., Badri,S., Ischenko,D., Eng,G., *et al.* (2018) Dissection of affinity captured LINE-1 macromolecular complexes. *Elife*, **7**, 210.

64. Miyoshi,T., Makino,T. and Moran,J.V. (2019) Poly(ADP-Ribose) Polymerase 2 Recruits Replication Protein A to Sites of LINE-1 Integration to Facilitate Retrotransposition. *Molecular Cell*, **75**, 1286–1298.e12.

65. MacLennan,M., Garcia-Cañadas,M., Reichmann,J., Khazina,E., Wagner,G., Playfoot,C.J., Salvador-Palomeque,C., Mann,A.R., Peressini,P., Sanchez,L., *et al.* (2017) Mobilization of LINE-1 retrotransposons is restricted by Tex19.1 in mouse embryonic stem cells. *Elife*, **6**, 210.

66. Benitez-Guijarro,M., Lopez-Ruiz,C., Tarnauskaitė,Ž., Murina,O., Mian Mohammad,M., Williams,T.C., Fluteau,A., Sanchez,L., Vilar-Astasio,R., Garcia-Cañadas,M., *et al.* (2018) RNase H2, mutated in Aicardi-Goutières syndrome, promotes LINE-1 retrotransposition. *The EMBO journal*, **37**, 46.

67. Warkocki,Z., Krawczyk,P.S., Adamska,D., Bijata,K., Garcia-Perez,J.L. and Dziembowski,A. (2018) Uridylation by TUT4/7 Restricts Retrotransposition of Human LINE-1s. *Cell*, **174**, 1537–1548.e29.

68. Kazazian,H.H.,Jr. and Moran,J.V. (2017) Mobile DNA in Health and Disease. *N Engl J Med*, **377**, 361–370.

69. Payer,L.M. and Burns,K.H. (2019) Transposable elements in human genetic disease. *Nat Rev Genet*, **20**, 760–772.

70. Homma,S., Chen,J.C., Rahimov,F., Lou Beermann,M., Hanger,K., Bibat,G.M., Wagner,K.R., Kunkel,L.M., Emerson,C.P. and Miller,J.B. (2011) A unique library of myogenic cells from facioscapulohumeral muscular dystrophy subjects and unaffected relatives: family, disease and cell function. *European Journal of Human Genetics*, **20**, 404–410.

71. Sassaman,D.M., Dombroski,B.A., Moran,J.V., Kimberland,M.L., Naas,T.P., DeBerardinis,R.J., Gabriel,A., Swergold,G.D. and Kazazian,H.H. (1997) Many human L1 elements are capable of retrotransposition. *Nat. Genet.*, **16**, 37–43.

72. Goodier,J.L., Zhang,L., Vetter,M.R. and Kazazian,H.H. (2007) LINE-1 ORF1 protein localizes in stress granules with other RNA-binding proteins, including components of RNA interference RNA-induced silencing complex. *Molecular and Cellular Biology*, **27**, 6469–6483.

73. Kowarz,E., Löscher,D. and Marschalek,R. (2015) Optimized Sleeping Beauty transposons rapidly generate stable transgenic cell lines. *Biotechnology Journal*, **10**, 647–653.

74. Garcia-Perez,J.L., Morell,M., Scheys,J.O., Kulpa,D.A., Morell,S., Carter,C.C., Hammer,G.D., Collins,K.L., O'Shea,K.S., Menendez,P., *et al.* (2010) Epigenetic silencing of engineered L1 retrotransposition events in human embryonic carcinoma cells. *Nature*, **466**, 769–773.

75. Zhang,Y., Werling,U. and Edelmann,W. (2012) SLiCE: a novel bacterial cell extract-based DNA cloning method. *Nucleic Acids Res.*, **40**, e55–e55.

76. Han,J.S. and Boeke,J.D. (2004) A highly active synthetic mammalian retrotransposon. *Nature*, **429**, 314–318.

77. Martin,S.L. (1991) Ribonucleoprotein particles with LINE-1 RNA in mouse embryonal carcinoma cells. *Molecular and Cellular Biology*, **11**, 4804–4807.

78. Branciforte,D. and Martin,S.L. (1994) Developmental and cell type specificity of LINE-1 expression in mouse testis: implications for transposition. *Molecular and Cellular Biology*, **14**, 2584–2592.

79. Kano,H., Godoy,I., Courtney,C., Vetter,M.R., Gerton,G.L., Ostertag,E.M. and Kazazian,H.H. (2009) L1 retrotransposition occurs mainly in embryogenesis and creates somatic mosaicism. *Genes & Development*, **23**, 1303–1312.

80. Muotri,A.R., Chu,V.T., Marchetto,M.C.N., Deng,W., Moran,J.V. and Gage,F.H. (2005) Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature*, **435**, 903–910.

81. Newkirk,S.J., Lee,S., Grandi,F.C., Gaysinskaya,V., Rosser,J.M., Vanden Berg,N., Hogarth,C.A., Marchetto,M.C.N., Muotri,A.R., Griswold,M.D., *et al.* (2017) Intact piRNA pathway prevents L1 mobilization in male meiosis. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, E5635–E5644.

82. Rosser,J.M. and An,W. (2012) L1 expression and regulation in humans and rodents. *Front Biosci (Elite Ed)*, **4**, 2203–2225.

83. Rodić,N., Sharma,R., Sharma,R., Zampella,J., Dai,L., Taylor,M.S., Hruban,R.H., Iacobuzio-Donahue,C.A., Maitra,A., Torbenson,M.S., *et al.* (2014) Long Interspersed Element-1 Protein Expression Is a Hallmark of Many Human Cancers. *The American Journal of Pathology*, **184**, 1280–1286.

84. Macia,A., Widmann,T.J., Heras,S.R., Ayllon,V., Sanchez,L., Benkaddour-Boumzaouad,M., Muñoz-Lopez,M., Rubio,A., Amador-Cubero,S., Blanco-Jimenez,E., *et al.* (2017) Engineered LINE-1 retrotransposition in nondividing human neurons. *Genome Research*, **27**, 335–348.

85. RANGWALA,S. (2009) The L1 retrotransposition assay: A retrospective and toolkit. *Methods*, **49**, 219–226.

86. Muotri,A.R., Marchetto,M.C.N., Coufal,N.G., Oefner,R., Yeo,G., Nakashima,K. and Gage,F.H. (2010) L1 retrotransposition in neurons is modulated by MeCP2. *Nature*, **468**, 443–446.

87. Coufal,N.G., Garcia-Perez,J.L., Peng,G.E., Yeo,G.W., Mu,Y., Lovci,M.T., Morell,M., O'Shea,K.S., Moran,J.V. and Gage,F.H. (2009) L1 retrotransposition in human neural progenitor cells. *Nature*, **460**, 1127–1131.

88. Sun,X., Wang,X., Tang,Z., Grivainis,M., Kahler,D., Yun,C., Mita,P., Fenyö,D. and Boeke,J.D. (2018) Transcription factor profiling reveals molecular choreography and key regulators of human retrotransposon expression. *Proc. Natl. Acad. Sci. U.S.A.*, **115**, E5526–E5535.

89. Zhu,C.-H., Mouly,V., Cooper,R.N., Mamchaoui,K., Bigot,A., Shay,J.W., Di Santo,J.P., Butler-Browne,G.S. and Wright,W.E. (2007) Cellular senescence in human myoblasts is overcome by human telomerase reverse transcriptase and cyclin-dependent kinase 4: consequences in aging muscle and therapeutic strategies for muscular dystrophies. *Aging Cell*, **6**, 515–523.

90. Thorley,M., Duguez,S., Mazza,E.M.C., Valsoni,S., Bigot,A., Mamchaoui,K., Harmon,B., Voit,T., Mouly,V. and Duddy,W. (2016) Skeletal muscle characteristics are preserved in hTERT/cdk4 human myogenic cell lines. *Skeletal Muscle*, **6**, 43–12.

91. Brouha,B., Meischl,C., Ostertag,E., de Boer,M., Zhang,Y., Neijens,H., Roos,D. and Kazazian,H.H. (2002) Evidence Consistent with Human L1 Retrotransposition in Maternal Meiosis I. *The American Journal of Human Genetics*, **71**, 327–336.

92. Taylor,M.S., LaCava,J., Mita,P., Molloy,K.R., Huang,C.R.L., Li,D., Adney,E.M., Jiang,H., Burns,K.H., Chait,B.T., *et al.* (2013) Affinity Proteomics Reveals Human Host Factors Implicated in Discrete Stages of LINE-1 Retrotransposition. *Cell*, **155**, 1034–1048.

93. Lanciano,S. and Cristofari,G. (2020) Measuring and interpreting transposable element expression. *Nat Rev Genet*.

94. Belancio,V.P., Roy-Engel,A.M., Pochampally,R.R. and Deininger,P. (2010) Somatic expression of LINE-1 elements in human tissues. *Nucleic Acids Res.*, **38**, 3909–3922.

95. Navarro,F.C., Hoops,J., Bellfy,L., Cerveira,E., Zhu,Q., Zhang,C., Lee,C. and Gerstein,M.B. (2019) TeXP: Deconvolving the effects of pervasive and autonomous transcription of transposable elements. *PLoS Comput Biol*, **15**, e1007293.

96. Roberson,P.A., Romero,M.A., Osburn,S.C., Mumford,P.W., Vann,C.G., Fox,C.D., McCullough,D.J., Brown,M.D. and Roberts,M.D. (2019) Skeletal muscle LINE-1 ORF1 mRNA is higher in older humans but decreases with endurance exercise and is negatively associated with higher physical activity. *J. Appl. Physiol.*, **127**, 895–904.

97. Doucet-O'Hare,T.T., Rodić,N., Sharma,R., Darbari,I., Abril,G., Choi,J.A., Young Ahn,J., Cheng,Y., Anders,R.A., Burns,K.H., *et al.* (2015) LINE-1 expression and retrotransposition in Barrett's esophagus and esophageal carcinoma. *Proc Natl Acad Sci USA*, **112**, E4894–E4900.

98. Miki,Y., Nishisho,I., Horii,A., Miyoshi,Y., Utsunomiya,J., Kinzler,K.W., Vogelstein,B. and Nakamura,Y. (1992) Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer. *Cancer Research*, **52**, 643–645.

99. Doucet-O'Hare,T.T., Sharma,R., Rodić,N., Anders,R.A., Burns,K.H. and Kazazian,H.H.,Jr. (2016) Somatically Acquired LINE-1 Insertions in Normal Esophagus Undergo Clonal Expansion in Esophageal Squamous Cell Carcinoma. *Human Mutation*, **37**, 942–954.

100. Garcia-Perez,J.L., Marchetto,M.C.N., Muotri,A.R., Coufal,N.G., Gage,F.H., O'Shea,K.S. and Moran,J.V. (2007) LINE-1 retrotransposition in human embryonic stem cells. *Hum. Mol. Genet.*, **16**, 1569–1577.

101. Kubo,S., Seleme,M.D.C., Soifer,H.S., Perez,J.L.G., Moran,J.V., Kazazian,H.H. and Kasahara,N. (2006) L1 retrotransposition in nondividing and primary human somatic cells. *Proc Natl Acad Sci USA*, **103**, 8036–8041.

102. Barbieri,D., Elvira-Matelot,E., Pelinski,Y., Genève,L., de Laval,B., Yogarajah,G., Pecquet,C., Constantinescu,S.N. and Porteu,F. (2018) Thrombopoietin protects hematopoietic stem cells from retrotransposon-mediated damage by promoting an antiviral response. *J. Exp. Med.*, **215**, 1463–1480.

MAIN FIGURES

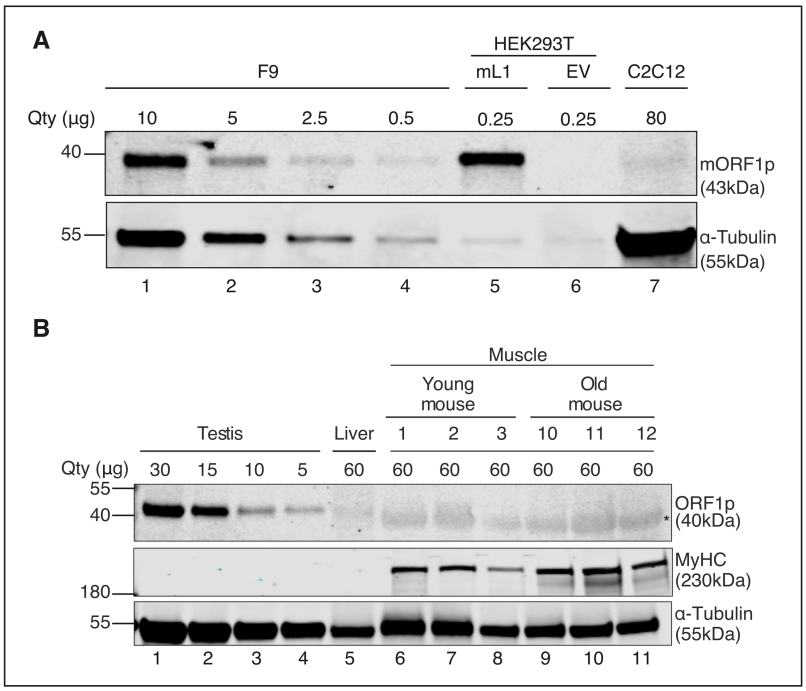


Figure 1. Endogenous L1 ORF1p expression is undetectable in mouse skeletal muscle and tissues.
(**A**) Western blot detection of endogenous ORF1p protein in whole cell lysate (WCL) of C2C12 cells (murine immortalized myoblasts). Assay sensitivity was determined using decreasing quantities of WCL of murine F9 teratocarcinoma cells, known to express L1 (77). Human HEK-293T cells transfected with pWA121, expressing a codon-optimized mouse L1 element (mL1) or an empty vector (EV) served as positive and negative controls, respectively. (**B**) Western Blot detection of endogenous ORF1p protein in total extracts of mouse testis, liver, and muscle. Assay sensitivity was determined using decreasing quantities of testis extracts, and liver extracts were used as negative control. Numbers above muscle lanes reflect the ID of the mouse. Myosin heavy chain (MyHC) is a specific marker of myotubes. Note that the fuzzy band marked by a star corresponds to a non-specific cross-reaction of the antibody with a very abundant protein found in muscle cells, with a slightly lower apparent molecular weight than ORF1p (see also Ponceau staining of the membrane in Figure S1). For both panels, quantities of WCL loaded are indicated at the top of each lane; ORF1p detection was achieved using a rabbit monoclonal antibody against mouse ORF1p (EPR21844-108); and α-tubulin detection was used as a loading control.

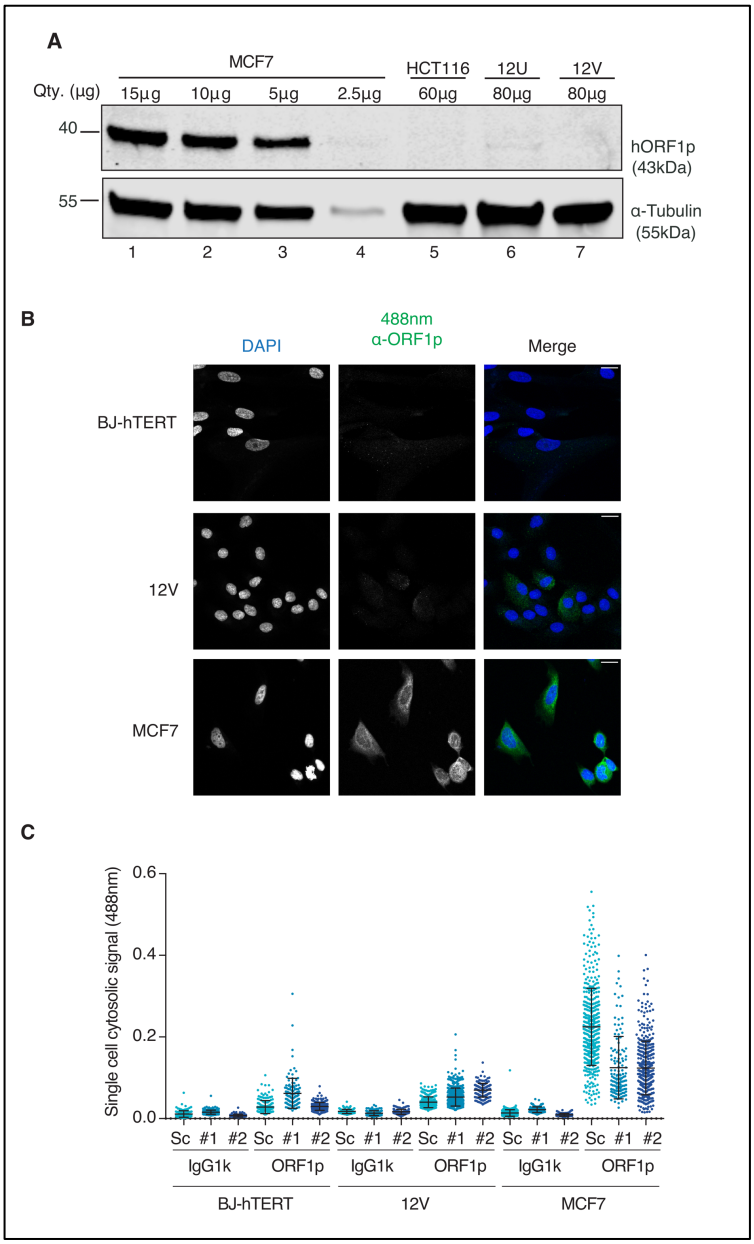


Figure 2. Endogenous L1 ORF1p expression is undetectable in human immortalized myoblasts.
(**A**) Western blot detection of endogenous ORF1p in WCL of 12U and 12V cells (human immortalized myoblasts) using a mouse monoclonal antibody against human ORF1p (4H1). The sensitivity of the antibody was determined using decreasing concentrations of WCL of human MCF7 breast cancer cells, known to express L1 (13). MCF7 and HCT116 cell lines were used as positive and negative controls of ORF1p expression, respectively. α-tubulin detection was used as a loading control, and quantities of WCL loaded is indicated at the top of each lane. (**B**) Immunofluorescence detection of endogenous ORF1p in 12U and 12V cells (human immortalized myoblasts) using a mouse monoclonal antibody against human ORF1p (4H1). Immortalized fibroblasts (BJ-hTERT) and a breast cancer cell line (MCF7) were used as negative and positive controls of ORF1p expression, respectively. Additional control experiments without a primary antibody or with an isotype control antibody are shown in Figure S2. (**C**) shRNA-mediated knock-down of ORF1p confirms that the very low cytoplasmic signal detected in a subset of BJ-hTERT or 12V cells is non-specific. Cells were transduced with lentiviral vectors containing scrambled (sc) or two distinct ORF1 shRNAs (#1 or #2) at high multiplicity of infection (m.o.i.=10) and further examined by immunofluorescence using an anti-ORF1p monoclonal antibody (4H1). Median cytoplasmic staining intensity of single cells was automatically recorded with CellProfiler. Note that the strong signal detected in MCF7 cells is reduced by both ORF1 shRNA. In contrast, the faint signal detected in myoblasts (12V) or fibroblasts (BJ-hTERT) is unaffected by ORF1 shRNAs, suggesting cross-reactivity of the antibody with another protein. Representative images are shown in Figure S3. Scale bar, 20 µm.

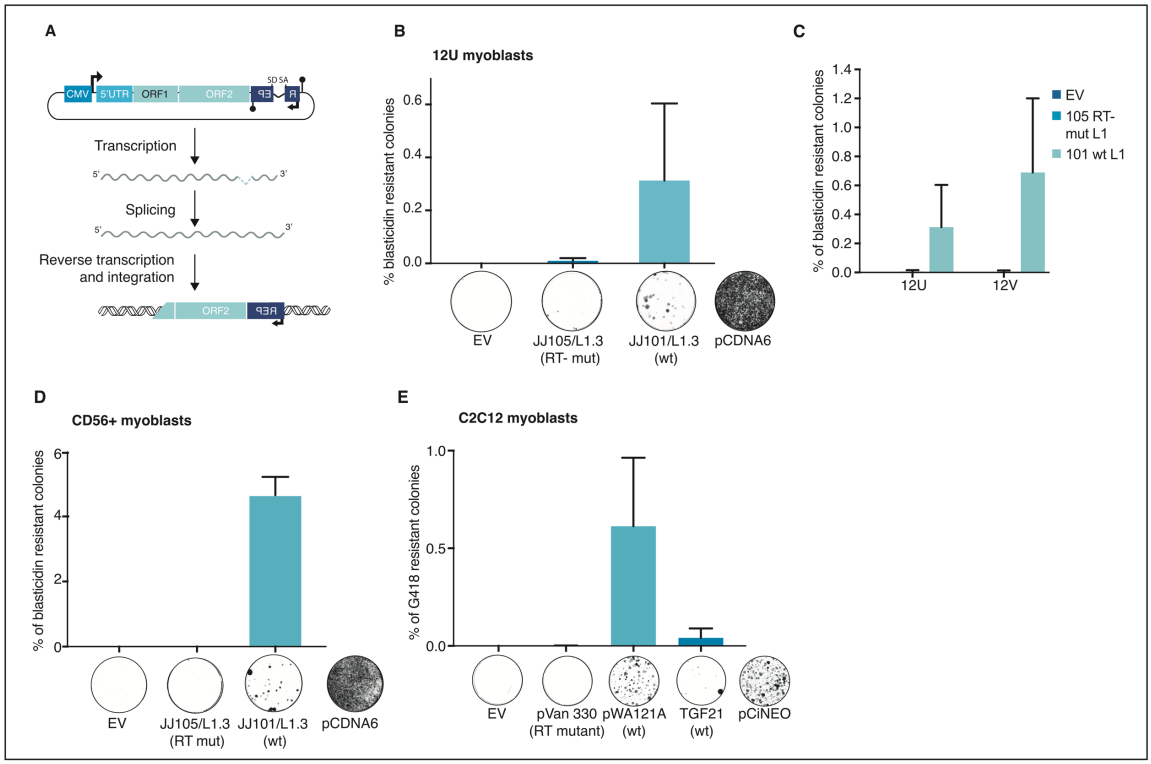


Figure 3. Retrotransposition of an engineered L1 in human and mouse myoblasts.
(**A**) Schematic of the retrotransposition assay. A plasmid-borne RC-L1 element (L1.3) under the control of a strong promoter and tagged with a retrotransposition indicator cassette (REP, blasticidin- or G418-resistance for human and mouse L1 constructs, respectively) is transfected in cultured cells, along with a GFP plasmid to estimate transfection efficiency. (**B**) Retrotransposition frequency in human 12U immortalized myoblasts, calculated as the number of blasticidin-resistant colonies normalized by the number of transfected (GFP-positive) cells. Bars represent the mean ± s.d. (n=3). Under each bar of the graph a picture of a representative plate with stained colonies is displayed for illustrative purposes. A pcDNA6 plasmid, containing the blasticidin S deaminase gene, was used as positive control for selection. (**C**) Retrotransposition frequency in human immortalized myoblast cell lines isolated from three distinct patients (12U, 12V and 12A). Bars represent the mean ± s.d. (n=3). (**D**) Retrotransposition frequency in human primary myoblasts. Cells were isolated from a tissue biopsy obtained from a healthy individual. Bars represent the mean ± s.d. (n=3). Under each bar of the graph a picture of a representative well with stained colonies is displayed for illustrative purposes. As above, a pcDNA6 plasmid was used as positive control for selection. (**E**) Retrotransposition frequency in murine C2C12 immortalized myoblasts, calculated as the number of G418-resistant colonies normalized by the number of transfected (GFP-positive) cells. Bars represent the mean ± s.d. (n=3). Under each bar of the graph a picture of a representative well with stained colonies is displayed for illustrative purposes. mORFeus is codon-optimized version of mouse L1spa, and TGf-21 is a natural and retrotransposition-competent mouse L1. A pCI-neo plasmid, containing the neomycin phosphotransferase gene, was used as positive control for selection.

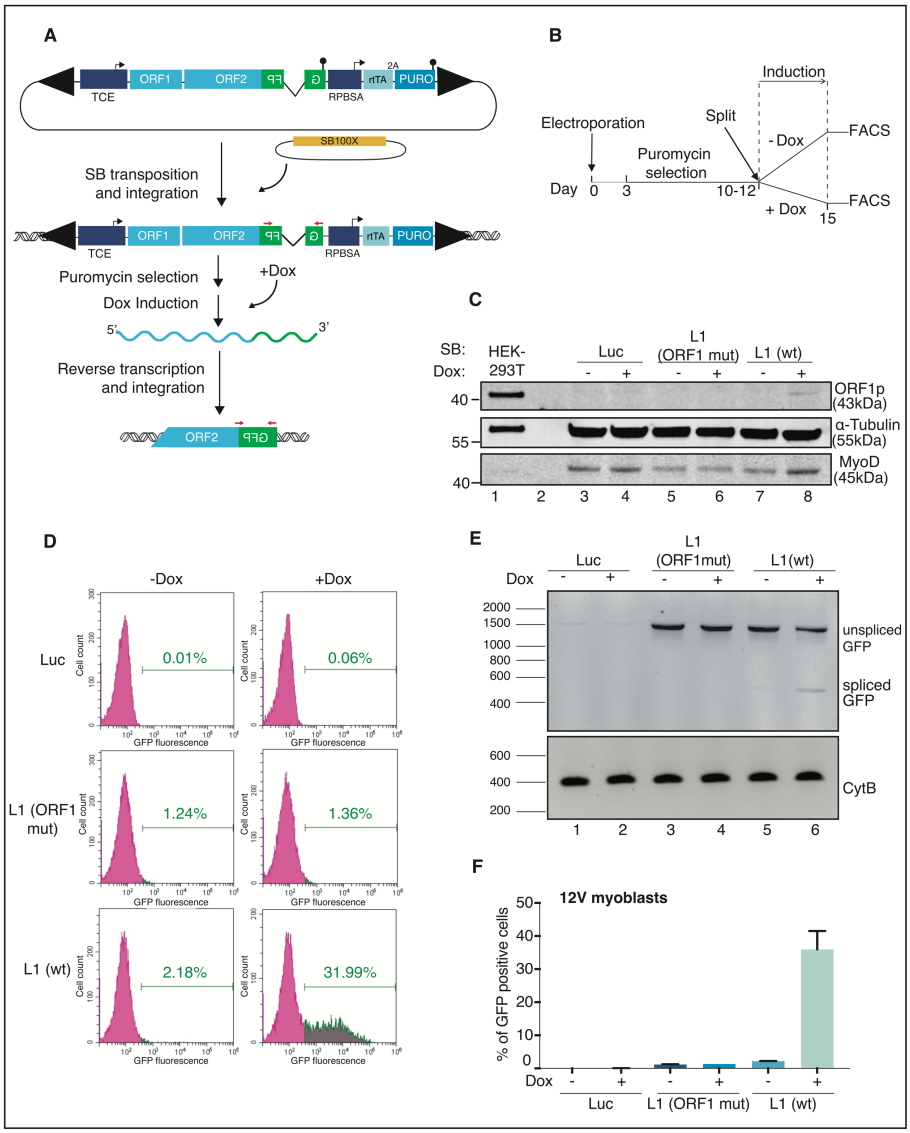


Figure 4. High frequency retrotransposition in human immortalized myoblasts from inducible and integrated engineered L1 elements. (**A**) Schematic of the Sleeping Beauty (SB)-based retrotransposition assay. An RC-L1 element (LRE3, (91) under the control of a Dox-inducible promoter and tagged with a GFP-based retrotransposition indicator cassette is embedded in an SB transposon. The SB backbone contains a puromycin selection cassette and expresses the reverse tetracycline transactivator (rtTA). (**B**) Time course of the SB-based retrotransposition assay. Cells are co-transfected with the SB-L1 construct and the SB transposase expression plasmid (SB100X). After puromycin-selection, L1 expression is induced by doxycycline for 3 days and the fraction of GFP-positive cells is measured by flow-cytometry. (**C**) Western Blot detection of ORF1p in whole cell lysates (60 µg, WCL) of 12V immortalized myoblasts containing SB-Luc, or various SB-L1 transgenes with or without Dox induction. Luc, an SB vector with a Luciferase gene serving as negative control, since L1 is absent; L1mut, an SB-L1 containing the R261A and R262A mutations in ORF1p; L1, a wild-type L1 LRE3. HEK-293T were included as positive control for ORF1p-expressing cells. α-tubulin detection was used as a loading control, and Myo-D as a myoblast marker. **(D)** Representative flow-cytometry profiles of the SB-based retrotransposition assays. EGFP fluorescence is plotted on the x-axis and side scatter on the y-axis. Cells counted as EGFP-positive are shown in green and their percentage is indicated in the upper right corner of each panel. The ORF1mut construct contains a point mutation that abolishes L1 retrotransposition activity (18). Luc, SB construct with a Firefly Luciferase cassette instead of L1-GFP as a negative control. **(E)** Average proportion of GFP positive cells obtained in SB-based retrotransposition assays, with or without doxycyclin (Dox) induction. **(F)** PCR assay on genomic DNA of cells from SB-based retrotransposition assay under each experimental condition showing intron removal of the GFP cassette upon retrotransposition. Top, PCR with primers flanking the GFP intron. The upper band at 1488 bp shows the unspliced version of the GFP cassette contained in the integrated SB-L1 construct. The lower band at 586 bp represents the spliced version of the GFP cassette and thus retrotransposition. Bottom, PCR with primers targeting Cytochrome B (CytB), as loading control.

SUPPLEMENTARY FIGURES

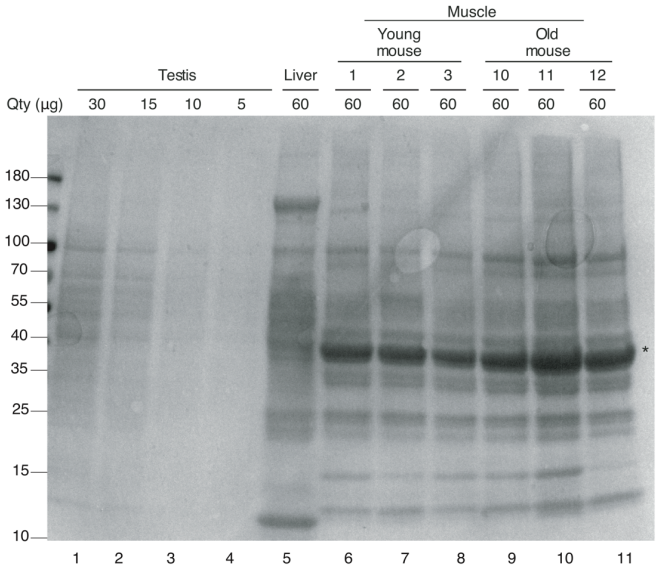


Figure S1. Ponceau staining of the membrane used for western-blot in Figure 1B.

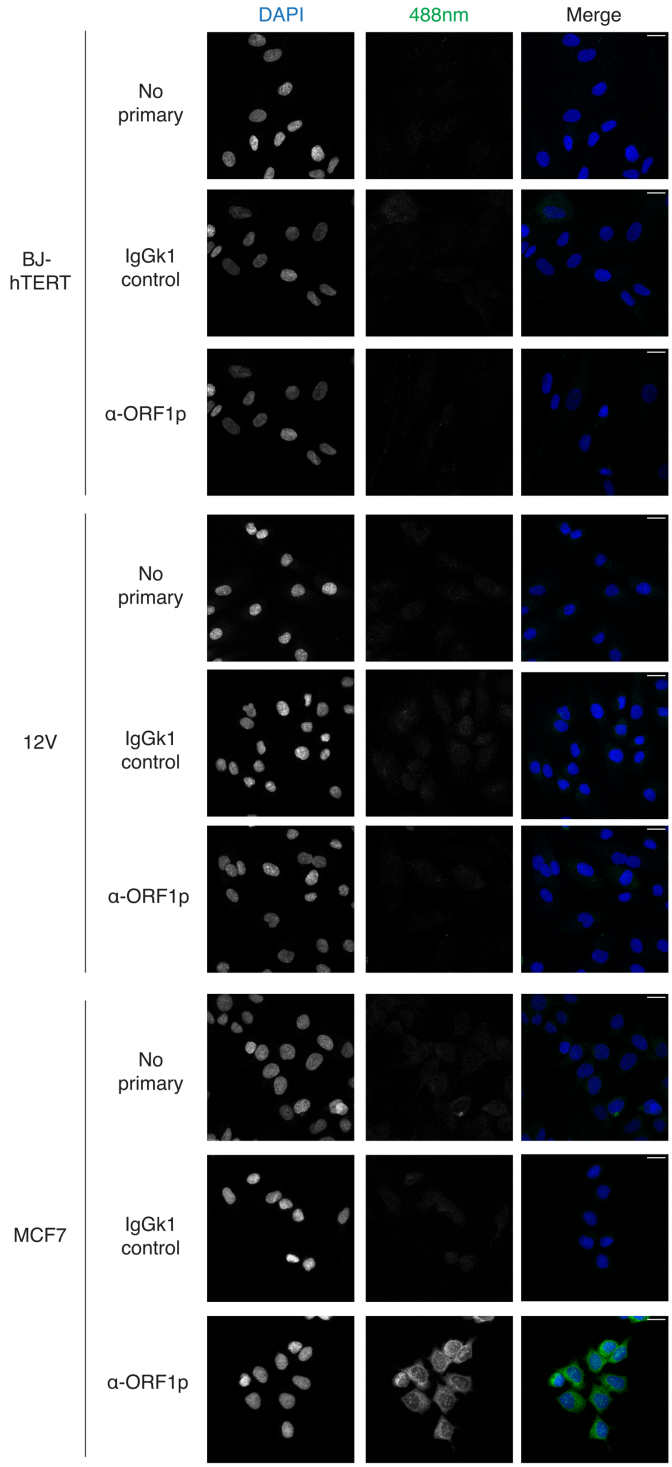


Figure S2. Specific and non-specific immunofluorescence signals obtained under our experimental conditions.
Cells were stained with the 4H1 mouse monoclonal antibody (α-ORF1p), without primary antibody (no primary), or with an isotype IgG1K control antibody, followed in all conditions by an anti-mouse antibody coupled to Alexa-488 (green). Nuclei were stained with DAPI (blue). A very faint signal can be detected with the anti-ORF1p antibody in some BJ-hTERT or 12V cells. However, this signal appears non-specific since similar staining can be obtained with the isotype control antibody and since it is not reduced upon ORF1 knock-down, in contrast to the robust and specific immunofluorescence signal detected in MCF7 cells (see Figure 3 and S3.). Scale bar, 20 µm.

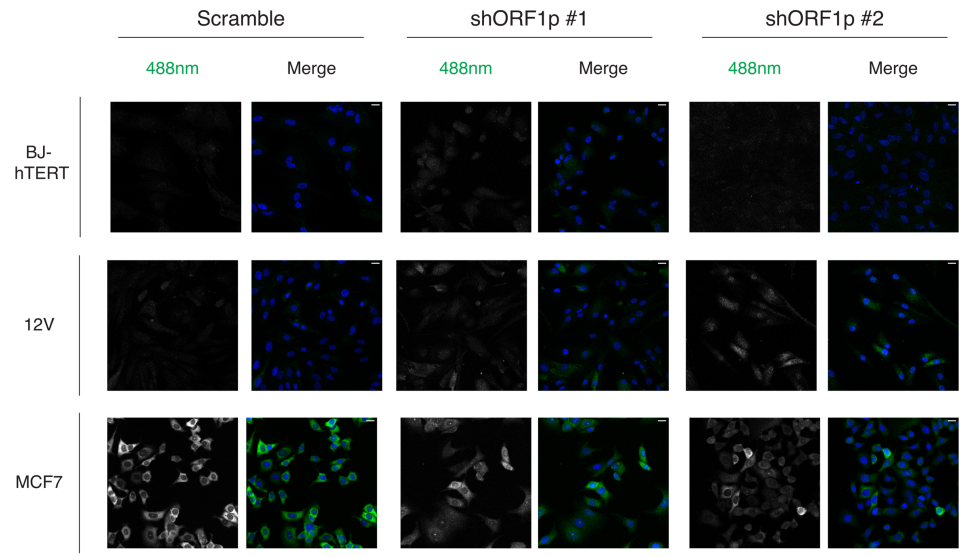


Figure S3. Efficacy of shRNA-mediated ORF1 knock-down.
Conditions are those indicated in the legend of Figure 3, and representative immunofluorescence images, such as those quantified in Figure 3B, are shown. Nuclei were stained with DAPI (blue) and shown in the Merge panels Scale bar, 20 μm.

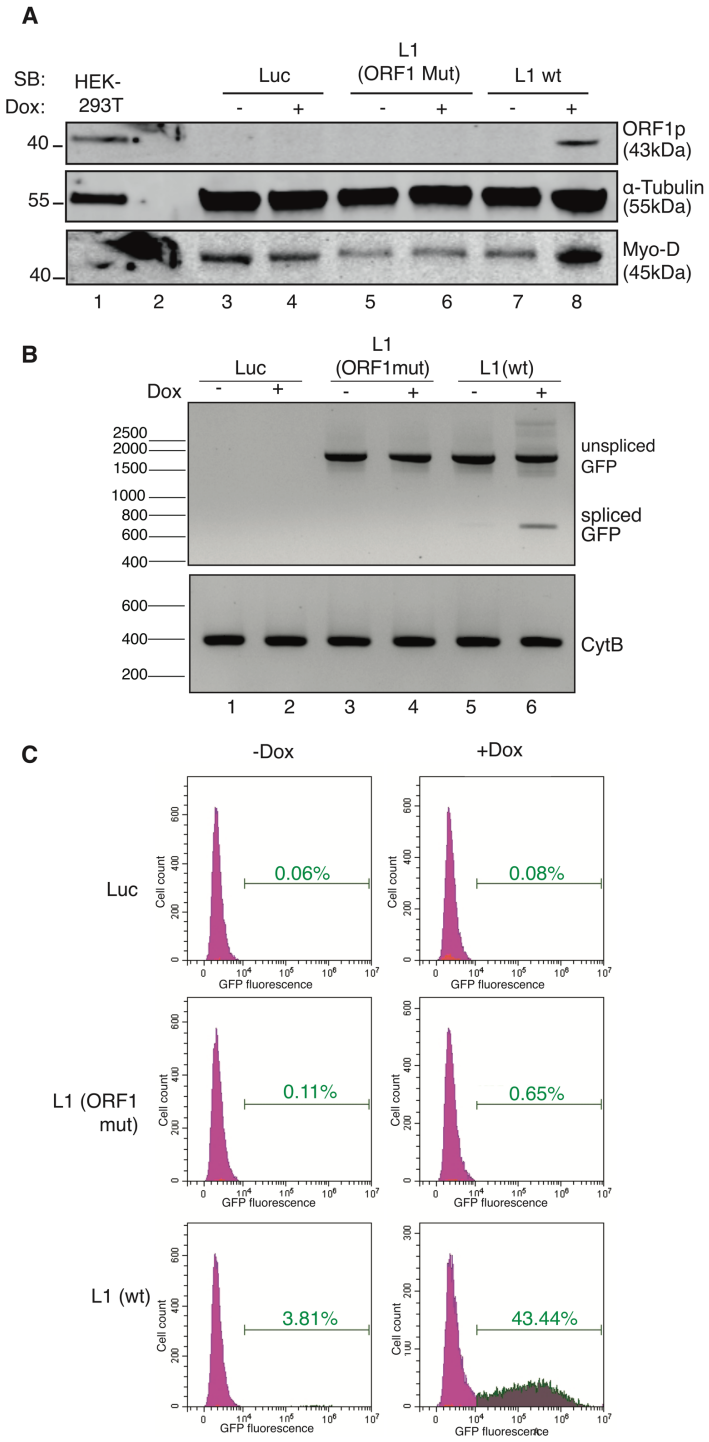


Figure S4. Sleeping Beauty-L1 retrotransposition in 12U cells.
As for Figure 4C, Western Blot detection of ORF1p in whole cell lysates (60 µg, WCL) of 12U immortalized myoblasts containing SB-Luc, or various SB-L1 transgenes with or without Dox induction. Luc, an SB vector with a Luciferase gene serving as negative control, since L1 is absent; L1mut, an SB-L1 containing the R261A and R262A mutations in ORF1p; L1, a wild-type L1 LRE3. HEK-293T were included as positive control for ORF1p-expressing cells. α-tubulin detection was used as a loading control, and MyoD as a myoblast marker.

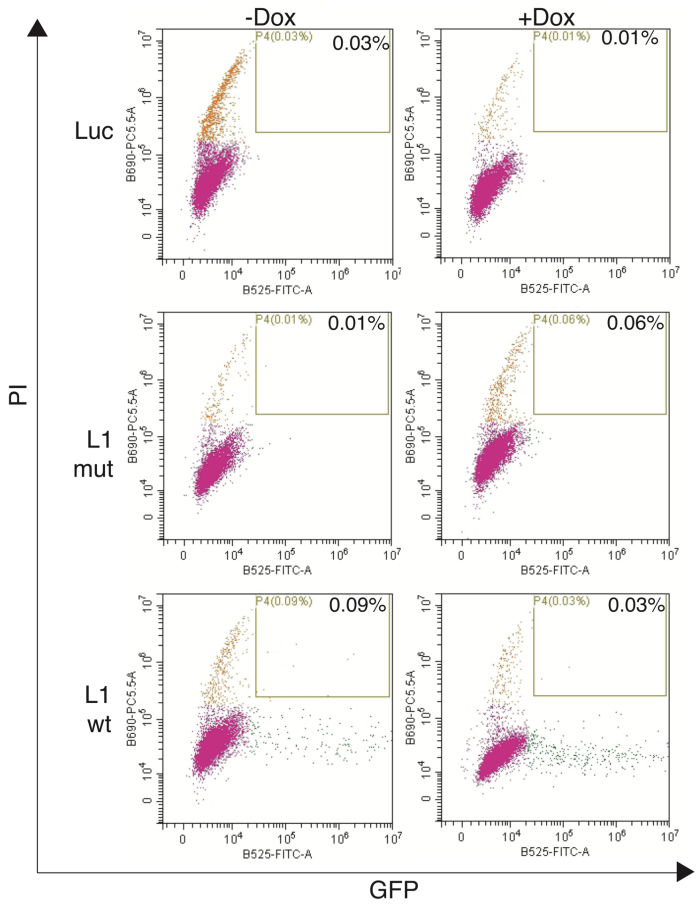


Figure S5. GFP positive cells are not auto fluorescent dead cells on SB GFP assay.
Flow cytometry analysis of human immortalized myoblasts 12V from SB GFP retrotransposition assay stained with Propidium Iodine (PI), to distinguish dead cells from GFP positive ones. The plots display the signal of PI incorporated by dead cells against GFP positive signal for each condition of the experiment. The gating shows double positive cells.

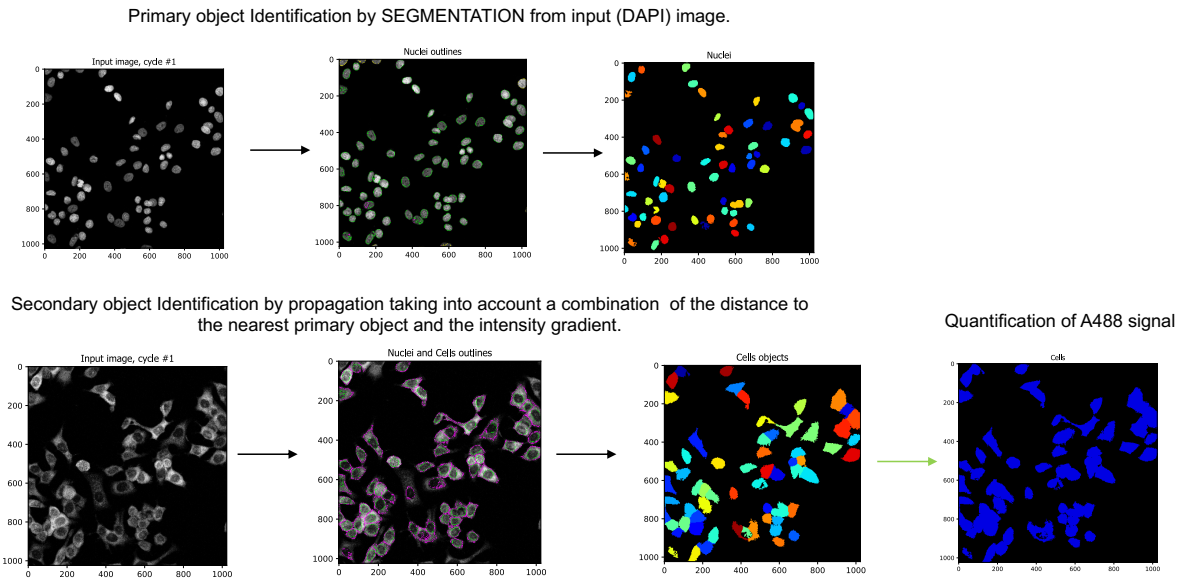


Figure S6. Cell Profiler method. First an identification of primary objects (nuclei) with Otsu thresholding method with the DAPI channel is performed. Then, secondary objects (cells) are identified by the propagation Method with the A488 channel and after, the nuclei are subtracted from the cell surface in the A488 channel to define cytoplasm. Finally, calculation of the median intensity in the A488 channel for each cytoplasm area.

3. Additional results

3. Additional results

3.1 Materials and methods

In this chapter, we will report the methods used to obtain the additional results that are not part of the publication in preparation.

Cell culture

12A cells were grown in LHCN (lox-hygro-hTERT ("LH"), and Cdk4-neo ("CN")) medium containing Dulbecco's Modified Eagle Medium (DMEM) (31966047, Life Technologies) and 199 Medium mixed at a 4:1 ratio (41150020, Life Technologies), 20% fetal bovine serum (FBS) (10270106, Life Technologies), HEPES 50mM (15630056, Life Tecnologies), 0.03 µg/mL Zinc sulfate (Z0251-100G, Sigma Aldrich), 14 µg/mL vitamin B12, 55 ng/mL dexamethasone, 2.5 ng/mL Hepatocyte Growth Factor (GF116, Sigma), 10 ng/mL Basic Fibroblast Growth Factor (PHG0266, Life Techonlogies) and 1% Penicillin-Streptomycin (v/v). Cells were plated in dishes coated with 0.1% gelatin (G1890, Sigma) and cultured at 37ºC with 5% CO_2. Coating was obtained by adding 0.1% gelatin on plastic plates and incubating for 2h at 37ºC.

C2C12, BJ fibroblasts and MCF7 cells were grown in Dulbecco's Modified Eagle Medium (DMEM) (31966047, Life Technologies), supplemented with 10% FBS and 1% Penicillin-Streptomycin (v/v) (Life Technologies).

HCT116 colon carcinoma cells were grown in McCoy's 5a Medium (GIBCO 36600088) 10% FBS, 1% penicillin-streptomycin (v/v) (Life Technologies).

Ribonucleoprotein particle extraction

This experiment was performed following the published protocol by Doucet et al. [386]. Dry pellets of 10 million cells were resuspended in 500 µL of CHAPS lysis buffer filtered at 0.22 µm (10 mM Tris pH 7.5, 0.5 % CHAPS (w/v), 1 mM $MgCl_2$, 1 mM EGTA, 10 % glycerol (v/v), 1x Protease inhibitor cocktail (04693159001 Roche) and 1 mM DTT), incubated on ice for 15 minutes and centrifuged at 20,000 xg at 4ºC for 15 min. The supernatant is separated and diluted in 500 µL of CHAPS lysis buffer, and 500 µL are loaded into an ultra-clear ultracentrifuge tube with sucrose cushion prepared with 17% and 8.5% layers of sucrose solution. Each sucrose solution was prepared from a 47 % sucrose stock solution filtered at 0.22 µm, and containing 20 mM Tris pH 7.5, 80 mM NaCl, 5 mM $MgCl_2$, and 47 % sucrose (w/v) diluted with a sucrose dilution buffer containing 2 mM Tris pH 7.5, 80 mM NaCl, 5 mM $MgCl_2$, 1x protease inhibitor cocktail (04693159001 Roche) and 1 mM DTT. Once loaded, the tubes were centrifuged at 93,000 x g for 2 h at 4ºC. After centrifugation,

the supernatant was carefully removed and the pellets were resuspended in RNAse free water with 1x Protease inhibitor cocktail (04693159001 Roche) overnight (O/N) on an orbital shaker at 4ºC. The next day the samples were quantified and flash-frozen at -80ºC for conservation.

Immunoblotting

The immunoblotting technique used in these experiments has been previously described in the attached publication.

Sleeping Beauty Blasticidin retrotransposition assay

C2C12 were transfected using Lipofectamine 2000 reagent (Life Technologies) at a density of 1.5 x 10^5 cells/well in 6-well plates. Conditions contained a Sleeping Beauty (SB) construct with the wild-type version of L1, an RT mutant version or a luciferase gene as a control. In every case, cells were co-transfected with the transposase plasmid or with an empty vector as negative control. Cells were selected into puromycin-containing media (1.5 µg/mL), starting 72h after transfection. Fresh media was added every 48 h. After puromycin selection, cells were plated at equal densities in every condition and induced with 1 µg/mL of doxycycline for 72 h. After induction, cells were selected with 2 µg/mL blasticidin for 7 days. Colonies were fixed and stained with Crystal violet fixing solution [0.5% (w/v) crystal violet, 10% methanol (v/v), 10% acetic acid(v/v)].

Plasmids

- pVan847: a derivative of pSBtet-Pur in which the Luc cassette has been replaced by a full-length retrotransposition-competent L1 (L1.3) tagged with a blasticidin antisense retrotransposition indicator cassette. pSBtet-Pur backbone contains an inducible SfiI cloning site for GOI (firefly luciferase replaced by L1) and constitutive expression of rtTA and puromycin resistance gene (Addgene #60507).

-pVan849: a derivative of pVan9847 with a point mutation in ORF1p that abolishes retrotransposition.

-SB LUC: Sleeping Beauty construct with luciferase gene. pSBtet-Pur backbone with inducible SfiI cloning site for GOI (firefly luciferase) and constitutive expression of rtTA and puromycin resistance gene (Addgene #60507).

SB100X: plasmid allowing the constitutive expression of a highly active Sleeping Beauty transposase in mammalian cells (Addgene #65487).

SINE retrotransposition assay

C2C12 were transfected using 9µL of Lipofectamine 2000 reagent (Life Technologies) and 4µg of DNA at a density of 1.5 x 10^5 cells/well in 6-well plates. Cells were transfected with a B1 element, a B1 element with a mutation that enhances its transposition efficiency and a B2 element. For each condition a co-transfection with an ORF2p containing plasmid or a control empty vector was made. Additionally, transfection efficiency was assessed by co-transfection of each condition with a GFP plasmid and posterior flow cytometry analysis.
Seventy two hours after transfection, cells were transferred to growth media containing G418 at 1 mg/mL (Life Technologies). Selection was carried out for 7 days. Once the cells in the control wells were dead, plates were fixed and stained with Crystal violet fixing solution (0.5% w/v crystal violet, 10% methanol (v/v), 10% acetic acid(v/v)).

The SINE plasmids used in this experiment were kindly provided by Marie Dewannieux and Thierry Heidmann [387].

Plasmids:

-B1-NeoTet WT (clone B1-2): B1 expression vector with the NeoTet retrotransposition reporter (NEO + Tetrahymena self-splicing group I intron)

-B1-NeoTet mutT24G (clone B1-2): B1 mutant version expression vector with the NeoTet retrotransposition reporter (NEO + Tetrahymena self-splicing group I intron)

-B2-NeoTet: B2 expression vector with the NeoTet retrotransposition reporter (NEO + Tetrahymena self-splicing group I intron)

-pAD001: expression vector containing a human L1 ORF2p sequence driven by a Cytomegalovirus (CMV) promoter.

Blasticidin cassette retrotransposition assay on 12A FSHD1 cells

12A cells were provided by Charles P Emerson Jr. from University of Massachusetts Medical School. They were immortalized by integration of hTERT and CDK4 constructs. In this case, these cells were extracted from a biopsy taken from the biceps of an FSHD1 24-year-old patient.
This experiment was carried out exactly as described for 12U and 12V cells in the methods of the attached publication.

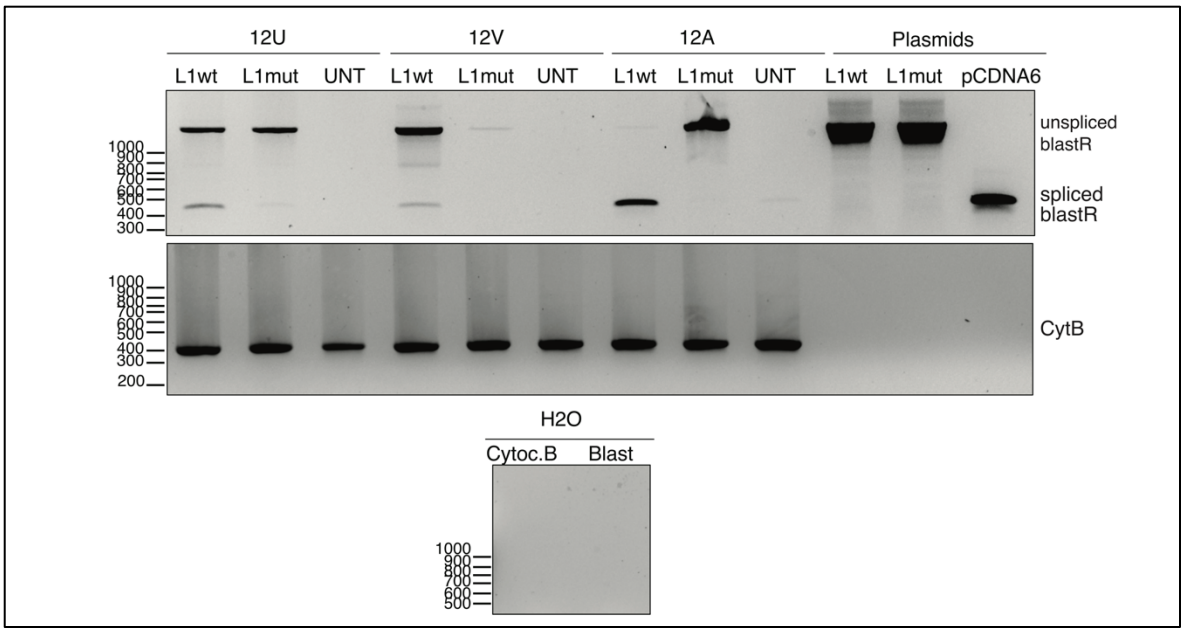
3.2. Results

The results shown in this chapter were not included in the previously shown publication, but they complement our work and are important for future perspectives.

Intron removal on insertions in human immortalized myoblasts

After performing a conventional retrotransposition assay in human immortalized myoblasts, we aimed to validate the retrotransposition events by analyzing the insertions and confirming that the intron in the blasticidin-resistance cassette was spliced out. Consequently, I performed a PCR on genomic DNA obtained from cells used for the retrotransposition assay. Primers against the blasticidin-resistance gene were located internally in the cassette, giving a product of 1488 bp in its unspliced version (Fig.18). This product appears in every condition containing the L1 wt and L1 mutant plasmid although the intensity of each bands is variable. This might indicate that some cells may still contain the plasmid, while in other conditions it was diluted or lost upon cell passages. Once the intron is removed, the expected band at 586 bp is detected. In all three cell lines, the lower band is detected for wt L1 plasmid, while in 12U a fainter band can also be appreciated in for the mutant.

These results confirm that human immortalized myoblast from either healthy and FSHD muscle can sustain L1 retrotransposition from an engineered plasmid.



(Legend in next page)

Figure 18. **Intron removal upon retrotransposition in human immortalized myoblasts with the plasmid-borne retrotransposition assay.** PCR assay on genomic DNA of cells from retrotransposition assay of each experimental condition showing intron removal of the Blasticidin cassette intron upon retrotransposition. Top, PCR with primers flanking the Blasticidin intron. The upper band at 1488 bp shows the unspliced version of the blasticidin-resistance cassette contained in the engineered L1 construct. The lower band at 586 bp represents the spliced version of the cassette and thus retrotransposition. 12V and 12U are healthy human myoblasts, while 12A was derived from an FSHD patient. L1wt represents JJ101 plasmid containing a WT version of the L1.3 sequence, while L1mut represents JJ105, an RT mutant that is unable to retrotranspose. pcDNA6 plasmid containing a blasticidin-resistant gene was used as a positive control for the spliced version of the Blasticidin cassette. UNT refers to untransfected cells. Bottom, PCR with primers targeting Cytochrome B, as loading control.

Endogenous ORF1p is not detected in 12U myoblasts ribonucleoprotein particle preparation

Since we were unable to detect ORF1p in WCL from human myoblasts, we decided to sediment cellular RNPs, a procedure that allows to enrich L1 RNP [82] and could raise our detection limit using the same antibody as used in previous immunoblots (Figure 1 of publication). While we were able to detect ORF1p in RNP preparations obtained from MCF7 cells, RNP prepared from 12U myoblasts, HCT116 colon carcinoma, MRC5 and BJ fibroblasts cell lines were all negative for ORF1p (Figure 19). All the samples were positive for S6, a ribosomal protein from 40S ribosome complex that served as an RNP preparation and loading control. Altogether, our results indicate that ORF1p is not present in the human myoblast tested.

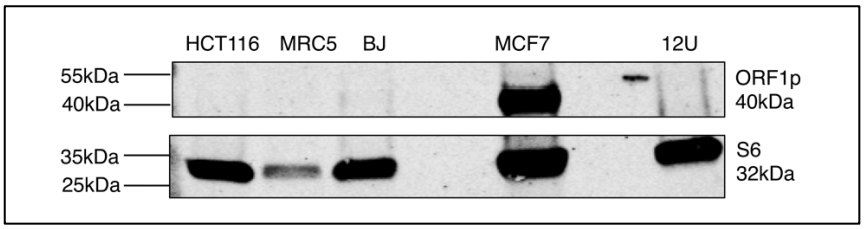


Figure 19. Endogenous ORF1p is not detected in 12U myoblasts ribonucleoprotein particle preparation. Western Blot detection of endogenous ORF1p in a ribonucleoprotein particle (RNP) preparation of 12U human myoblasts. Membranes were probed with Millipore 4H1 monoclonal antibody against ORF1p. HCT116 colon carcinoma, MRC5 and BJ primary fibroblasts, known for expressing low or undetectable levels of ORF1p were selected as negative control [65]. MCF7 breast carcinoma cells were used as positive control since their high level of expression for ORF1p has been described as well [65]. S6 ribosomal protein was used as an RNP preparation and loading control. 60 µg of RNP preparation were loaded in each lane.

Sleeping beauty retrotransposition assay on C2C12 mouse immortalized myoblasts.

We developed another variant of the SB retrotransposition assay, in which the L1 is a human L1.3 copy carrying a blasticidin-resistance retrotransposition reporter cassette, different from the SB GFP construct previously used in the attached publication that bared an LRE3 L1 sequence and an EGFP reporter. This assay was carried out identically to the

SB-L1-GFP one, with the exception that after 72h of Dox induction, cells were selected with blasticidin for 7 days, and colonies were fixed once control cells containing the SB-Luc vector died. We tested this assay in C2C12 cells, and found that similar to the neomycin-resistance-based mouse L1 assay, these cells are also able to mobilize a human L1 from an integrated copy at a maximum rate of 0.25% (Figure 20). In the previously shown conventional mL1 assay (Figure 3 in article), the retrotransposition rate was 0.6 and 1% for the synthetic codon-optimized mouse L1 (pWA121), We can conclude that, C2C12 mouse myoblasts can also mobilize a human L1 sequence.

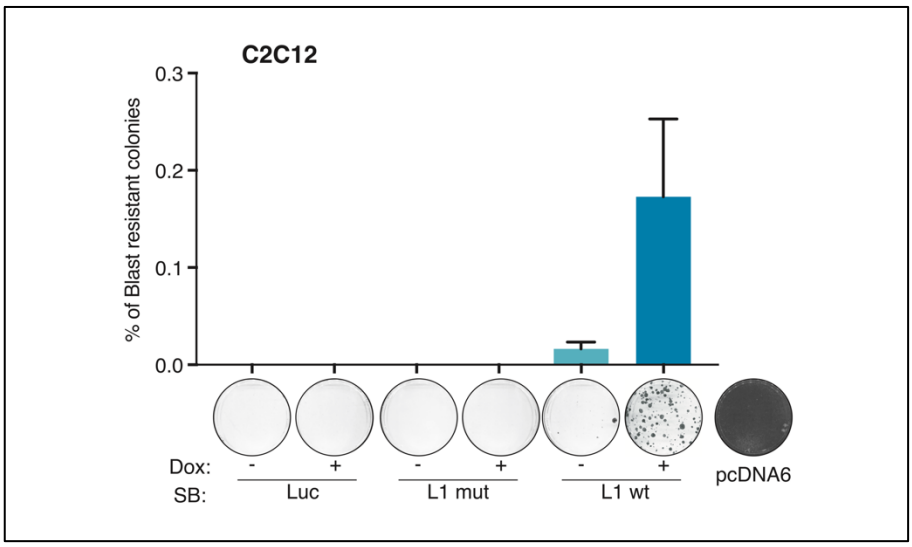


Figure 20. Sleeping Beauty retrotransposition assay in C2C12 mouse immortalized myoblasts. Retrotransposition frequency of the SB-based retrotransposition assay on C2C12 cells. The RT mutant construct contains a point mutation that abolishes L1 retrotransposition activity. Luc, SB construct with a Firefly Luciferase cassette instead of L1 as a negative control. Bars represent the average proportion of blasticidin-resistant cells obtained in SB-based retrotransposition assays, with or without doxycyclin (Dox) induction [mean ± s.d. (n=3)]. A pcDNA6 empty vector plasmid containing a blasticidin-resistance gene was used as positive control for blasticidin selection.

SINE retrotransposition assay in C2C12 mouse myoblasts

SINE elements can only retrotranspose in the presence of L1 ORF2p, since they do not encode any enzymatic activities. Similar to the L1 retrotransposition assay, a SINE assay indicates if cells are permissive for SINE mobilization, but they also inform of the presence of endogenous L1 ORF2p proteins. To test if the endogenous LINE-1 machinery of mouse myoblast is able to transpose SINE elements, we performed a retrotransposition assay on C2C12 myoblasts that consists on transfecting cells with a plasmid containing the SINE sequence followed by a reporter cassette with a neomycin-resistance gene interrupted by a self-splicing intron (which does not require Pol II transcription), similar to the conventional L1 retrotransposition assay. SINE plasmids are co-transfected with a plasmid expressing

L1 ORF2p or with an empty vector. Three mouse SINE sequences were tested: a B1, a B2 and a B1 sequence that carries a mutation that improves its capacity for retrotransposition.

If the tested cells are able to produce ORF2p endogenously, expressing only the SINE element is sufficient to obtain G418 resistance. On the contrary, if the cells lack ORF2p expression, but can accommodate SINE retrotransposition, resistant colonies will be obtained when cells are co-transfected with the marked SINE and the ORF2p-expressing plasmid. This is precisely what we observed in C2C12 cells for the B1 mutant version and for B2 (Figure 21). Our results suggest that, very likely, C2C12 cells do not express ORF2p endogenously, but are permissive to SINE retrotransposition when the L1 machinery is present.

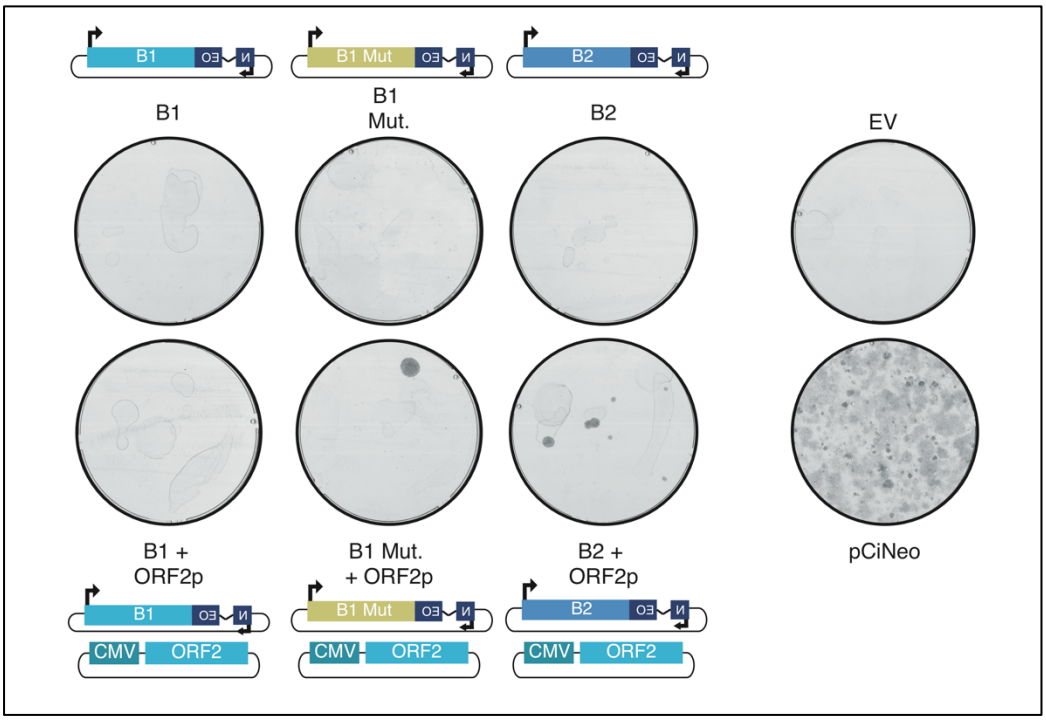


Figure 21. SINE retrotransposition assay in C2C12 mouse myoblasts. Transposition assay showing neomycin-resistant colonies in C2C12 cells co-transfected with either SINE sequences with or without an ORF2p-expressing plasmid. A pCiNeo empty vector plasmid containing a neomycin-resistant gene was used as positive control. B1 Mut is a mutated version of B1 that improves its retrotransposition efficiency. Schemes of the plasmids are shown for illustration purposes.

Blasticidin-based retrotransposition assay in 12A FSHD human immortalized myoblasts

FSHD is a muscular dystrophy characterized by an abnormal expression of DUX4 homeobox gene that ends up inducing apoptosis and eventually causes muscular weakness [375,388]. DUX4 is normally expressed in the early embryo where it triggers Zygotic

genome activation (ZGA) [383]. Two relevant findings about DUX4 have been published recently. Firstly, DUX4 was shown to bind and activate to TEs promoter, including LINE-1 [382]. Second, DUX4 expression during 2-cell embryo state was reported to be reduced by L1 RNA bound by Nucleolin [384]. These two discoveries made us consider the possibility of an implication of L1 in the FSHD phenotype. Thus, we decided to include 12A cells from an FSHD patient in our experiments. We performed the conventional blasticidin-based retrotransposition assay in these cells, identically as we did for cells from healthy individuals, which has been reported in the attached publication of this thesis. Surprisingly, we found that although FSHD cells are permissive for L1 retrotransposition, they show a lower retrotransposition rate (0.13-0.2%) compared myoblasts isolated from healthy individuals (0.6-1.3%) in the same experiment (Figure 22). Additionally, blasticidin-resistant colonies were smaller than the ones obtained for 12U and 12V myoblasts, and cell morphology significantly changed, cells being irregular after the blasticidin selection. This suggests that although they can sustain retrotransposition, 12A cells could be more sensitive to L1 expression or retrotransposition, affecting cell proliferation or survival, as compared to 12U and 12V healthy cells.

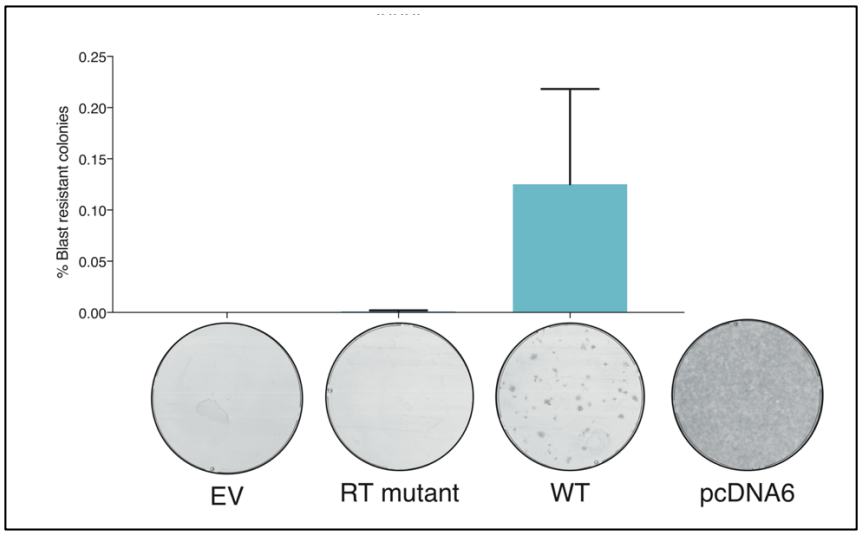


Figure 22. Blasticidin-based retrotransposition assay in 12A FSHD human immortalized myoblasts. Retrotransposition frequency in 12A immortalized myoblasts, calculated as the number of blasticidin-resistant colonies normalized by the number of transfected (GFP-positive) cells. Bars represent the mean ± s.d. (n=3). Under each bar of the graph a picture of a representative well with stained colonies is displayed for illustrative purposes. A pcDNA6 empty vector plasmid containing a blasticidin-resistance gene was used as positive control for blasticidin resistance.

4. Discussion

4. Discussion

Transposable elements are a component of the genome of almost every organism. They are involved in gene regulation and can be drivers of evolution. Specifically, LINE-1, is of great relevance in the study of human genetics because it is currently the only active autonomous element in the human genome. Consequently, we consider that a significant but not fully explored niche in this field is the mobility of LINE-1 elements in somatic tissues.

L1 insertions, regulatory modifications or L1 derived mutations are transmitted to the following generations when they take place in the germline, affecting the genome of the future gametes [88] [205-207]. During the formation of the germline, early embryo cells can acquire new L1 insertions, which would then be passed to the individual's progeny potentially affecting the fitness of the carrier.

However, inherited insertions are not the only ones that can modify the fitness of an organism. Embryonic development constitutes a window of opportunity for L1 to spread in the genome. Since these cells will become different tissues, this insertion could ultimately generate somatic mosaicism [208-211,226-229].
Once embryonic stem cells differentiate into their final tissue, a pool of stem cells usually remains present and since they retain the ability to proliferate, these cells are responsible for tissue regeneration when needed. The scope of consequences of L1 mobility in somatic stem cells is evidently not comparable to those of ESCs, however, if retrotransposition happens in a somatic stem cell, every cell derived from it would carry the new L1 copy and its potential effect on cellular function [216,226].

A similar case, but with a different reach, is likely to take place in differentiated cells. The mosaic derived from retrotransposition in differentiated cells would entail individual and non-inheritable variations. The scope of this complex mosaicism in somatic tissues is greatly understudied. No clear conclusion can be drawn regarding the expression or retrotransposition of LINE-1 in somatic tissues. More specifically, L1 activity is unclear in differentiated cells and in most somatic stem cells outside the brain [182,215,221]. Moreover, somatic cells' permissiveness to L1 retrotransposition in most of the organs has not yet been demonstrated.

Based on the observations that ORF1p might be present in muscle of human esophagus and of aged mice [257,275], we decided to focus our work on skeletal muscle. We analyzed

healthy human and murine myoblasts, and one cell line derived from an FSHD muscular dystrophy patient. In this work, we focused our efforts in two important aspects in the life cycle of LINE-1: expression and retrotransposition.

ORF1p expression in muscle samples

Initially, we investigated the expression of ORF1p in human and mouse immortalized myoblasts. Since no ORF2p specific antibody able to detect endogenous ORF2p expression is available to date and its expression has been described to be extremely low, only ORF1p analysis was performed. ORF1p expression has been widely utilized for the assessment of L1 activity [98,386]. Here, immunoblotting did not show any specific signal for ORF1p although the monoclonal antibodies used for both mouse and human cells adequately recognizes ORF1p in control samples. The most direct explanation for these results would be that ORF1p is not being expressed in these cells, or that the expression is below the detection limit of our techniques. However, it was described for the L1HS-Ta family that only a small number of copies contribute to the general pool of L1 transcribed sequences [65]. This is accompanied by a cell-type-dependent activation. It is possible that the few L1 copies being expressed in muscle cells are not recognized by the antibodies used, due to individual polymorphisms that might affect its epitope.

In order to avoid the possible technical caveats that might have impaired the detection of ORF1p through immunoblotting, we performed immunofluorescence on human immortalized myoblasts using the same monoclonal antibody. However, consistent with our previous results, we could not clearly detect ORF1p with this technique either. The signal obtained in human immortalized myoblasts was very close to the one detected with control IgG or in BJ-hTERT fibroblasts (known to not express ORF1p), and significantly inferior to the one in MCF7 cells, that are described to show a very high expression of ORF1p [65]. To test the specificity of the weak signal detected, we knocked down ORF1 expression by shRNA. Again, we obtained no changes in the level of ORF1p signal in both myoblasts and fibroblasts (or even a slight increase), in contrast to MCF7 controls where the signal was clearly decreased by shRNA directed against ORF1. These results support the conclusion that ORF1p is not present in these cells, or only at extremely low levels, although as discussed previously, individual locus polymorphisms could also be an explanation in this case.

Belancio et al. described the presence of L1 RNA in several somatic tissues. Through northern-blots, they showed the presence of either full-length (FL) L1 or spliced ORF2 (spORF2) RNA. FL RNA expression was clear in the heart, prostate, esophagus and

stomach. However, other tissues showed a lower and less evident signal. Among them, the skeletal muscle only showed only a weak signal for spORF2[225]. Healthy esophagus muscle is reactive to ORF1p antibody by immunohistochemistry, and skeletal muscle of aged mice seem to express ORF1p as well [258,275]. ORF1p has been detected recently in mouse skeletal muscle from 26 months-old mice [257]. Moreover, recent studies on SIRT6 mutant mice with a progeria phenotype show that treatment with RT inhibitors are able to revert the muscle mass loss and muscle fiber thickness in these animals, suggesting that L1 expression might contribute to muscle affectation [258]. In these publications, the authors report that a cytoplasmic accumulation of L1 cDNA triggers the interferon (IFN) response through the activation of the cyclic GMP-AMP Synthase (cGAS) pathway, which was also previously described in TREX1 deficient neural cells [201].

These findings prompted us to explore whether ORF1p is expressed in muscle from aged animals. Thus, we compared the expression of this protein in both 9 weeks and 27 months-old mice by immunoblotting. Unfortunately, no specific signal was detected in samples from 3 different mice from each age range. The discrepancies between the published results by De Cecco and our findings might be explained by a technical issue inherent to the detection method or by the age of our animals, which might not be old enough to express ORF1p [257,274]. ORF1p detection in the mentioned publication was carried out through immunofluorescence, and performed directly on tissue, while, in our work, a protein extract from the muscle sample was prepared. We then hypothesized that the treatment applied to prepare the extract from these samples might reduce the detection capacity of the antibody or impair the collection of an already not highly concentrated ORF1p. Therefore, we believe that direct analysis of ORF1p expression in tissue sections through immunodetection might be an appropriate technique for future L1 research on this tissue.

Other studies have chosen a different approach involving RNA analysis, such as quantitative RT-PCR (qRT-PCR). De Cecco described in 2013 an increase of L1 RNA in muscle of aged mice compared to younger ones using this technique on total RNA extracted from tissue samples [274]. However, we believe that this approach can be misleading since numerous copies of L1 are inserted in genomic sequences and thus can be co-expressed with genes, masking active L1 RNA molecules. A more accurate approach might be the application of qRT-PCR to polyA+ samples where mRNA is enriched, reducing the amount of nuclear RNA and the possibility of false positive L1 signal. Nonetheless, some L1 insertions may still be transcribed if they are fused to a gene. Previous estimates indicate that as much as 99% of L1-containing transcripts are co-transcripts and not full-length L1 RNA [389].

Recently, a couple of studies have described the increase of L1 RNA in skeletal muscle of rat and human aged individuals by RT-PCR [390,391]. Additionally, one of these studies showed a marked presence of ORF1p in human skeletal muscle by immunoblotting. Interestingly, the band shown appears at 34kDa, when the expected size of ORF1p is 42 kDa, and contrasts with the very low expression found in the work of De Cecco [391].
A limitation in our study is the lack of analysis of L1 RNA expression. Northern-blot can be helpful in order to elucidate the presence of full length L1 RNA. Additionally, the comparison of RNA expression in young and aged muscle through radioactive probes has not been performed to our knowledge, and we consider that this approach could provide interesting information on this matter, although the sensitivity of northern-blots is limited.
L1 RNP purification has been used previously to demonstrate the presence of L1. We have performed this technique in our samples, and similarly to the whole cell lysate preparations, no detection of ORF1p was detected by immunoblotting.

Several investigations to date have described the global demethylation of DNA during muscle differentiation. More precisely, *Tet2* expression is necessary for differentiation of C2C12 myoblasts and is upregulated during this process, while 5hmC levels augment in adult muscle fibers compared to myoblasts or myotubes [392,393]. Consistently, an analysis of the methylation profiles of myoblast and myotubes compared to mature skeletal muscle fibers show a loss of 90% of hypermethylated sites [393]. As we previously discussed, 5hmC has been related to active L1 promoters, suggesting that myoblast differentiation might constitute an opportunity for L1 expression. Therefore, we hypothesize that L1 is being repressed by mechanisms not related to DNA methylation in mature skeletal muscle cells. This could be confirmed by analyzing bisulfite-whole genome sequencing data obtained in this tissue. If true, an interesting line of work would be to elucidate which repression factors block L1 expression in myoblasts or myofibers. Screening approaches have been used successfully to identify possible regulators and we believe they will contribute to future understanding of L1 regulation [149].

Overall, these results suggest that ORF1p expression in myoblasts is likely to be rather marginal even in aged mice, since the percentage of positive cells described in old mice skeletal muscle was fairly low (3%) [257]. We consider that it is possible that the aging phenotype might activate the expression of L1, but further investigation is necessary to accurately establish the presence or absence of L1 RNA or protein in aged skeletal muscle.

LINE-1 retrotransposition in muscle cells

Besides the insertions described in NPCs and glial cells [182,221], only a few studies based on bulk sequencing describe discrete insertions in healthy liver and gastrointestinal cells [262,269], and none of these studies investigated the capacity of somatic cells to sustain L1 retrotransposition *in vitro.*

In this work, we demonstrate that both mice and human muscle cells allow retrotransposition of an engineered L1 construct. Moreover, we show that both immortalized and primary myoblasts permit L1 mobilization. To our knowledge, this is the first study on permissiveness of muscle cells to retrotransposition.

The conventional retrotransposition assay has been extremely useful in the study of L1 biology. For instance, this assay was crucial in the discovery L1 regulation, the recovery of L1 insertions, the characterization of active L1s and the retrotransposition of SINEs, SVA, other non-autonomous elements and cellular mRNA [113,115,208,234,394]. It has therefore become an essential tool in the field of L1 research. We used different variants of the retrotransposition assay to finally show that in every case, muscle cells permit L1 mobilization. Remarkably, C2C12 myoblasts were able to form colonies with both mouse L1 plasmids and our Sleeping Beauty (SB) based assay carrying a human L1 sequence. Human immortalized myoblast, 12U and 12V, were able to form colonies with both the conventional plasmid-based assay and with the SB-L1-GFP assay. Additionally, we were able to confirm *de novo* insertions in both assays through PCR amplification of the spliced version of the retrotransposition reporter cassettes, which can only be generated upon reverse transcription.

a) Sleeping Beauty GFP retrotransposition assay

It is somewhat surprising that we were able to detect a very high percentage of GFP positive cells with SB-L1-GFP hybrid vectors (39,6%). Previous publications utilizing the conventional retrotransposition assay with an EGFP reporter in NPCs have recorded lower retrotransposition rates. For example, Muotri et al. showed that 0,75% of rat adult hippocampus-derived neural progenitor (AHNP) cells and rat hippocampus neural stem (HCN) cells were positive for GFP expression in a retrotransposition assay with the same L1 LRE3 sequence, and the EGFP gene driven by the CMV promoter [215]. Another study on NPCs derived from human embryonic stem cells shows a rate of retrotransposition of 5.6% by flow cytometry analysis of GFP-positive cells [217]. However, these studies also reported that EGFP retrotransposition assays can give rise to false negative insertions due to either an early 5' truncation at the beginning of TPRT, or to epigenetic silencing of the CMV promoter driving the transcription of the EGFP reporter which would ultimately impair the

expression of the fluorescent protein [139,215,217]. We thus speculate that such an epigenetic regulation of the EGFP promoter does not take place in human myoblasts, and therefore we were able to register a high percentage of GFP-positive cells. To strengthen these results, we demonstrated that the GFP signal detected was not caused by cell death autofluorescence since staining by propidium iodide did not reveal double stained cells in the main cell population gated for this experiment.

It is important to highlight that in our assay, L1 expression was induced by doxycycline added to the culture media, which might force the production of L1 RNA template molecules, improving retrotransposition success. Nevertheless, immunoblotting of myoblasts from the SB-L1-GFP retrotransposition assay did not yield to a strong signal for ORF1p. This finding was unexpected and does not correlate with the high level of retrotransposition found in human immortalized myoblast in this assay, making us consider once again a potential L1 repression at the protein or RNA level. We therefore planned an analysis (ongoing) of the insertions carrying the EGFP cassette through inverse PCR and long-read sequencing to confirm the presence of TPRT hallmarks and rule out possible false positives.

b) Conventional retrotransposition assay

The conventional retrotransposition assays carried out in human immortalized myoblasts yielded a retrotransposition rate closer to the ones previously described for hESC, ~0.5%. Similar rates were obtained in C2C12 mouse immortalized myoblasts for the codon optimized version of mouse L1 (pWA121) and slightly lower for the natural L1 TGf-21 element. These cells were also able to sustain retrotransposition from a SB vector containing a human L1.3 sequence, at a similar rate to what was observed for the TGf21 plasmid, in agreement with previous observations showing that transgenic mice and rats are able to retrotranspose engineered human L1 constructs at a similar rate as compared to mouse L1 constructs [211,215].

Interestingly, human primary myoblasts were also permissive for L1 retrotransposition, showing a retrotransposition rate of around 5%. These results point out that myoblasts from different genetic backgrounds allow retrotransposition of LINE-1 sequences, which suggests that no regulation mechanism seems to halt the mobilization of L1 and the necessary host factors that permit TPRT completion are present in these cells.

c) SINE retrotransposition assay

To complement these results, we performed a SINE retrotransposition assay in C2C12 cells that showed that, although the endogenous mouse L1 machinery was not able to mobilize B1 or B2 sequences, the addition of a plasmid containing a human ORF2p sequence was enough to (modestly) mobilize both B1 and B2 elements. This might be due to the absence of endogenous ORF2p in C2C12 cells. This is in agreement with the results that we have presented in the attached publication, by immunoblotting for the presence of ORF1p in these cells. Again, these results confirm the ones obtained in human cells, where the host cell factors that enable retrotransposition seem to be present, but the endogenous expression of L1, appear to be repressed.

We propose then, the use of sequencing techniques in which muscle tissue can be compared to another tissue sample of the same individual, in order to distinguish *de novo* insertions that took place in the somatic lineage (myoblasts or myotubes) or during embryonic development. New techniques like Oxford Nanopore might be useful in the detection of TPRT hallmarks and the identification of insertion sites due to the improved length of the reads.

Retrotransposition in FSHD cells

Lastly, we have decided to explore the potential involvement of L1 in the FSHD muscular dystrophy. Caused by the expression of DUX4, FSHD causes the weakening of several muscle group, leading to muscle atrophy. Due to the involvement of DUX4 in the stimulation of TE expression, we investigated its possible relationship with L1 retrotransposition.

To do so, we added to our samples a third human myoblast cell line that was obtained from muscle tissue of an FSHD patient, 12A. These cells were included in the conventional plasmid-based retrotransposition assays, giving positive results for L1 retrotransposition. However, we observed that 12A cells present a lower percentage (0.13-0.25%) of blasticidin-resistant colonies and a smaller size of the individual colonies compared to 12U and 12V (0.5-1.3%) Additionally, the morphology of 12A cells that underwent retrotransposition changed, showing irregularity and a smaller colony size, which denotes problems in proliferation, compared to the untreated cells. FSHD myoblasts are known to be susceptible to oxidative stress [395,396], which can ultimately cause DNA damage [397,398]. It is possible that ORF2p expression in 12A cells generates DSB that contribute to a preexisting sensitive environment, where cellular defense mechanisms are not able to perform correctly.

The expression of DUX4 homeobox gene on muscle cells is characteristic of FSHD. It causes apoptosis partially through the activation of p53 [399]. DUX4 was recently demonstrated to bind and activate the transcription of several TEs, including L1 [382,400]. Another scenario that could explain the lower retrotransposition rate, smaller colonies and change in morphology, would be a preexisting DUX4-induced L1 expression in these cells that, upon transfection and transcription of an engineered L1 copy, elevates the overall L1 activity, causing a high number of DSB that triggers p53 response in some cells, which inevitably die before being able to express the Blasticidin resistant gene [150,401]. Thus, we propose to perform an analysis of L1, DUX4 and p53 levels in these cells before and after L1 transfection and retrotransposition to help elucidate if this might be the case.

FSHD is an example of the potential involvement of L1 in genetic disease in the muscle. Cancer is another well-known context in which L1 and disease are intertwined, and recently studies have been highlighting its possible role in mental disorders [223,224]. We believe that a more thorough understanding of somatic L1 biology is important for further understanding its impact in pathological situations.

Retrotransposition in differentiated myotubes

Myoblasts are proliferative activated satellite cells that, once committed, can fuse and differentiate into myotubes that will mature into myofibers. If L1 mobilization takes place in a myoblast, the new insertions or L1 activity derived effects will be carried out into several nuclei of a single or multiple myotubes, potentially affecting the functioning of the myofiber. For this reason, the study of L1 activity in these cells is important. It is relevant to highlight that the multinucleated character of myotubes presents the possibility of diluting the potential deleterious mutations happening in a specific nuclear domain, and this aspect should be taken into consideration in further investigations since they may render myotubes more tolerant toward retrotransposition [279].

As mentioned above, muscle cells go through DNA demethylation during the differentiation process, and adult myofibers seem to have lost most of their hypermethylated sites. This could potentially allow L1 to mobilize if no other downstream regulation mechanism was involved. Furthermore, 5hmC overall levels were higher in mature myofibers than in myoblasts, suggesting that L1 expression could be permitted in differentiated myotubes. However, our results obtained in immunoblotting of young and aged mice muscle extract show no detectable expression of ORF1p. We consider that, given the recently published data on somatic L1 activation upon aging [257,258], further research is necessary on this topic.

Relative to this subject, we are currently investigating the retrotransposition of LINE-1 in myotubes, taking advantage of the SB-L1-GFP assay that was developed for this project. The inducible promoter allows us to trigger L1 expression before and after differentiation of human myoblasts. These experiments are ongoing and will be added to the attached publication.

Lastly, we would like to highlight that skeletal muscle is not a tissue prone to develop cancer, beside rhabdomyosarcoma, an infrequent extra muscular sarcoma, the only known malignancy in this organ [402]. We speculate that one of the reasons for this might be a potential tight regulation of L1 expression that blocks all L1 transcription and translation, even if the host factors for L1 mobilization are present.

Conclusions

-The expression of ORF1p protein on human or mouse myoblasts could not be detected on whole cell lysate, RNP preparation or cell samples by immunoblotting or immunofluorescence.

-The expression of ORF1p in mice tissue samples was under the detection limit and the age of the mice (at least 27 months-old) does not seem to raise the intensity of the signal.

-Human immortalized and primary myoblasts are permissive to LINE-1 retrotransposition from an engineered construct, presenting a high retrotransposition rate with an integrated L1 and a Dox-induced system.

-Mouse immortalized myoblasts are permissive for the retrotransposition of mice and human L1 engineered elements and allow the mobilization of SINE constructs in the presence of exogenous ORF2p.

-FSHD human immortalized myoblasts allow retrotransposition at a lower rate compared to healthy myoblasts, forming smaller colonies.

Final remarks

In a model in which L1 mobility in the germline or in the early embryo has the most potential to have an impact on the individual's fitness, somatic retrotransposition in individual cells does not seem to impose a great danger unless L1 expression or retrotransposition is generalized on a tissue or organ. So far, the studies on somatic retrotransposition have pointed in the direction of a tight regulation of L1 expression. However, a context like aging or disease, where L1 is eventually expressed all over the organ can constitute a situation that would lead to genomic instability or inflammation. It is therefore important to elucidate which tissues express the host factors that allow L1 mobilization once expression takes place. Indeed, we believe that FSHD and aging, in which L1 expression might be activated,

constitute two scenarios in which L1 activity can potentially have repercussions for the muscle. Therefore, we expect that further research will be carried out to elucidate the possible link between L1 and FSHD. Similarly, the effect of L1 on the aging muscle is an aspect we encourage for future studies.

Although a great part of the research on L1 activity has been carried out in the brain, the studies showing the interaction of L1 with external stimuli and its involvement in mental disorders, in addition to its possible involvement in cancer, aging and disease in other tissues, highlight the necessity of understanding L1 biology in its entirety. It is for this reason that we reinforce the idea that a more integrative study of somatic mosaicism is crucial, and we expect that over the next years the rapidly developing tools in this field will be applied to obtain a wider and more complete vision of the somatic landscape in L1 biology, with straight-forward and accurate detection of *de novo* somatic insertions.

5. References

5. References

1. McClintock, B. Controlling Elements and the Gene. *Cold Spring Harb Symp Quant Biol* 197–216 (1956).
2. Venter, C. The Sequence of the Human Genome. *Science* 1–51 (2001). doi:10.1126/science.1058040
3. Lander, E. Initial sequencing and analysis of the human genome. *Nature* 1–62 (2001).
4. Waterson, R. H. Initial sequencing and comparative analysis of the mouse genome. *Nature* 1–43 (2002).
5. Rollins, R. A. Large-scale structure of genomic methylation patterns. *Genome Research* **16,** 157–163 (2005).
6. Shapiro, J. A. Molecular model for the transposition and replication of bacteriophage Mu and other transposable elements. *Proc Natl Acad Sci USA* 1–5 (1979).
7. Pritham, E. J. & Feschotte, C. Massive amplification of rolling-circle transposons in the lineage of the bat Myotis lucifugus. *Proc Natl Acad Sci USA* 1–6 (2007).
8. Fedoroff, N. V., Wessler, S. & Shure, M. Isolation of the Transposable Maize Controlling Elements Ac and Ds. *Cell* 1–8 (1983).
9. Handler, A. M., McCombs, S. D., Fraser, M. M. & Saul, S. H. The lepidopteran transposon vector, piggyBac, mediates germ-line transformation in the Mediterranean fruit fly. *Proc Natl Acad Sci USA* 1–6 (1998).
10. Mahillon, J., Leonard, C. & Chandler, M. IS elements as constituents of bacterial genomes. *Research in Microbiology* 1–13 (1999).
11. Rio, D. C., Laski, F. A. & Rubin, G. M. Identification and Immunochemical Analysis of Biologically Active Drosophila P Element Transposase. *Cell* 1–12 (1986).
12. Kidwell, M., Kidwell, J. F. & Nei, M. A case of high rate of spontaneous mutation affecting viability in Drosophila Melanogaster. *Genetics* 1–21 (1973).
13. Ivics, Z., Hackett, P. B., Plasterk, R. H. & Izsvák, Z. Molecular Reconstruction of Sleeping Beauty, a Tc1-like Transposon from Fish, and Its Transposition in Human Cells. *Cell* 1–10 (1997).
14. Kowarz, E., Löscher, D. & Marschalek, R. Optimized Sleeping Beauty transposons rapidly generate stable transgenic cell lines. *Biotechnology Journal* **10,** 647–653 (2015).
15. Boeke, J. D., Garfinkel, D., Styles, C. & Fink, G. Ty Elements Transpose through an RNA Intermediate. *Cell* 1–10 (1985).
16. Eickbush, T. H. & Jamburuthugoda, V. K. The diversity of retrotransposons and the properties of their reverse transcriptases. *Virus Res.* **134,** 221–234 (2008).
17. Garfinkel, D. J., Boeke, J. D. & Fink, G. R. Ty element transposition: Reverse transcriptase and virus-like particles. *Cell* **42,** 507–517 (1985).
18. Cristofari, G., Ficheux, D. & Darlix, J.-L. 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 (2000).

19. Llorens, C., Muñoz-Pomer, A., Bernad, L., Botella, H. & Moya, A. Network dynamics of eukaryotic LTR retroelements beyond phylogenetic trees. *Biol. Direct* **4,** 41–31 (2009).
20. Frame, I. G., Cutfield, J. F. & Poulter, R. T. New BEL-like LTR-retrotransposons in Fugu rubripes, Caenorhabditis elegans, and Drosophila melanogaster. *Gene* **263,** 219–230 (2001).
21. la Chaux, de, N. & Wagner, A. BEL/Pao retrotransposons in metazoan genomes . *BMC Evolutionary Biology 2007 7:1* **11,** 1 (2011).
22. Kim, J. M., Vanguri, S., Boeke, J. D., Gabriel, A. & Voytas, D. F. Transposable elements and genome organization: a comprehensive survey of retrotransposons revealed by the complete Saccharomyces cerevisiae genome sequence. *Genome Research* **8,** 464–478 (1998).
23. Wicker, T. *et al.* A unified classification system for eukaryotic transposable elements. *Nat Rev Genet* **8,** 973–982 (2007).
24. Gdula, D. A., Gerasimova, T. I. & Corces, V. G. Genetic and molecular analysis of the gypsy chromatin insulator of Drosophila. *Proc Natl Acad Sci USA* **93,** 9378–9383 (1996).
25. Sandmeyer, S. B. & Menees, T. M. Morphogenesis at the retrotransposon-retrovirus interface: gypsy and copia families in yeast and Drosophila. *Curr. Top. Microbiol. Immunol.* **214,** 261–296 (1996).
26. Malik, H. S. & Eickbush, T. H. Phylogenetic analysis of ribonuclease H domains suggests a late, chimeric origin of LTR retrotransposable elements and retroviruses. *Genome Research* **11,** 1187–1197 (2001).
27. Terzian, C., Ferraz, C., Demaille, J. & Bucheton, A. Evolution of the Gypsy endogenous retrovirus in the Drosophila melanogaster subgroup. *Molecular Biology and Evolution* **17,** 908–914 (2000).
28. Kassiotis, G. & Stoye, J. P. Immune responses to endogenous retroelements: taking the bad with the good. *Nature Reviews Immunology* **16,** 207–219 (2016).
29. Tugnet, N., Rylance, P., Roden, D., Trela, M. & Nelson, P. Human Endogenous Retroviruses (HERVs) and Autoimmune Rheumatic Disease: Is There a Link? *Open Rheumatol J* **7,** 13–21 (2013).
30. Tatkiewicz, W. *et al.* Characterising a human endogenous retrovirus (HERV)-derived tumour-associated antigen: enriched RNA-Seq analysis of HERV-K(HML-2) in mantle cell lymphoma cell lines. *Mobile DNA* **11,** 9–15 (2020).
31. Garson, J. A. *et al.* Quantitative analysis of human endogenous retrovirus-K transcripts in postmortem premotor cortex fails to confirm elevated expression of HERV-K RNA in amyotrophic lateral sclerosis. *Acta Neuropathol Commun* **7,** 45–9 (2019).
32. Grow, E. J. *et al.* Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. *Nature* **522,** 221–225 (2015).
33. Nilsson, B. O., Jin, M., Andersson, A. C., Sundström, P. & Larsson, E. Expression of envelope proteins of endogeneous C-type retrovirus on the surface of mouse and human oocytes at fertilization. *Virus Genes* **18,** 115–120 (1999).
34. Kigami, D., Minami, N., Takayama, H. & Imai, H. MuERV-L is one of the earliest transcribed genes in mouse one-cell embryos. *Biology of Reproduction* **68,** 651–654 (2003).

35. Macfarlan, T. S. *et al.* Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* **487,** 57–63 (2012).
36. Svoboda, P. *et al.* RNAi and expression of retrotransposons MuERV-L and IAP in preimplantation mouse embryos. *Developmental Biology* **269,** 276–285 (2004).
37. Zhang, W. *et al.* Zscan4c activates endogenous retrovirus MERVL and cleavage embryo genes. *Nucleic Acids Res.* **47,** 8485–8501 (2019).
38. Luan, D. D., Korman, M. H., Jakubczak, J. L. & Eickbush, T. H. 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 (1993).
39. Malik, H. S., Burke, W. D. & Eickbush, T. H. The age and evolution of non-LTR retrotransposable elements. *Molecular Biology and Evolution* **16,** 793–805 (1999).
40. Kapitonov, V. V., Tempel, S. & Jurka, J. Simple and fast classification of non-LTR retrotransposons based on phylogeny of their RT domain protein sequences. *Gene* **448,** 207–213 (2009).
41. Yang, J., Malik, H. S. & Eickbush, T. H. Identification of the endonuclease domain encoded by R2 and other site-specific, non-long terminal repeat retrotransposable elements. *Proc Natl Acad Sci USA* **96,** 7847–7852 (1999).
42. Feng, Q., Moran, J. V., Kazazian, H. H. & Boeke, J. D. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell* **87,** 905–916 (1996).
43. Kojima, K. K. & Fujiwara, H. Evolution of target specificity in R1 clade non-LTR retrotransposons. *Molecular Biology and Evolution* **20,** 351–361 (2003).
44. Eickbush, D. G. & Eickbush, T. H. R2 retrotransposons encode a self-cleaving ribozyme for processing from an rRNA cotranscript. *Molecular and Cellular Biology* **30,** 3142–3150 (2010).
45. Kojima, K. K., Kuma, K.-I., Toh, H. & Fujiwara, H. Identification of rDNA-specific non-LTR retrotransposons in Cnidaria. *Molecular Biology and Evolution* **23,** 1984–1993 (2006).
46. Kolosha, V. O. & Martin, S. L. In vitro properties of the first ORF protein from mouse LINE-1 support its role in ribonucleoprotein particle formation during retrotransposition. *Proc Natl Acad Sci USA* 1–6 (1997).
47. Ivancevic, A. M., Kortschak, R. D., Bertozzi, T. & Adelson, D. L. LINEs between Species: Evolutionary Dynamics of LINE-1 Retrotransposons across the Eukaryotic Tree of Life. *Genome Biol Evol* **8,** 3301–3322 (2016).
48. Bennett, E. A. *et al.* Active Alu retrotransposons in the human genome. *Genome Research* **18,** 1875–1883 (2008).
49. Cordaux, R. & Batzer, M. A. The impact of retrotransposons on human genome evolution. *Nature Publishing Group* **10,** 691–703 (2009).
50. Ahl, V., Keller, H., Schmidt, S. & Weichenrieder, O. Retrotransposition and Crystal Structure of an Alu RNP in the Ribosome-Stalling Conformation. *Molecular Cell* **60,** 715–727 (2015).
51. Rubin, C. M., Houck, C. M., Deininger, P. L., Friedmann, T. & Schmid, C. W. Partial nucleotide sequence of the 300-nucleotide interspersed repeated human DNA sequences. *Nature* **284,** 372–374 (1980).

52. Chu, W. M., Liu, W. M. & Schmid, C. W. RNA polymerase III promoter and terminator elements affect Alu RNA expression. *Nucleic Acids Res.* **23,** 1750–1757 (1995).
53. Chesnokov, I. & Schmid, C. W. Flanking sequences of an Alu source stimulate transcription in vitro by interacting with sequence-specific transcription factors. *J Mol Evol* **42,** 30–36 (1996).
54. Hancks, D. C. & Kazazian, H. H. Active human retrotransposons: variation and disease. *Curr. Opin. Genet. Dev.* **22,** 191–203 (2012).
55. Song, X. *et al.* Predicting human genes susceptible to genomic instability associated with Alu/Alu-mediated rearrangements. *Genome Research* **28,** 1228–1242 (2018).
56. Ichiyanagi, K. Epigenetic regulation of transcription and possible functions of mammalian short interspersed elements, SINEs. *Genes Genet. Syst.* **88,** 19–29 (2013).
57. Ostertag, E. M., Goodier, J. L., Zhang, Y. & Kazazian, H. H., Jr. SVA Elements Are Nonautonomous Retrotransposons that Cause Disease in Humans. *The American Journal of Human Genetics* **73,** 1444–1451 (2003).
58. Damert, A. *et al.* 5'-Transducing SVA retrotransposon groups spread efficiently throughout the human genome. *Genome Research* **19,** 1992–2008 (2009).
59. Hancks, D. C., Ewing, A. D., Chen, J. E., Tokunaga, K. & Kazazian, H. H. Exon-trapping mediated by the human retrotransposon SVA. *Genome Research* **19,** 1983–1991 (2009).
60. Krayev, A. S. *et al.* The nucleotide sequence of the ubiquitous repetitive DNA sequence B1 complementary to the most abundant class of mouse fold-back RNA. *Nucleic Acids Res.* **8,** 1201–1215 (1980).
61. Quentin, Y. A master sequence related to a free left Alu monomer (FLAM) at the origin of the B1 family in rodent genomes. *Nucleic Acids Res.* **22,** 2222–2227 (1994).
62. Daniels, G. R. & Deininger, P. L. Repeat sequence families derived from mammalian tRNA genes. *Nature* **317,** 819–822 (1985).
63. Smit, A. F., Tóth, G., Riggs, A. D. & Jurka, J. Ancestral, mammalian-wide subfamilies of LINE-1 repetitive sequences. *Journal of Molecular Biology* **246,** 401–417 (1995).
64. Brouha, B. & Moran, J. V. Hot L1s account for the bulk of retrotransposition in the human population. *Proc Natl Acad Sci USA* 1–6 (2003).
65. Philippe, C. *et al.* Activation of individual L1 retrotransposon instances is restricted to cell-type dependent permissive loci. *Elife* **5,** e13926 (2016).
66. Lutz, S. M., Vincent, B. J., Kazazian, H. H., Jr., Batzer, M. A. & Moran, J. V. Allelic Heterogeneity in LINE-1 Retrotransposition Activity. *The American Journal of Human Genetics* **73,** 1431–1437 (2003).
67. Seleme, M. D. C. *et al.* Extensive individual variation in L1 retrotransposition capability contributes to human genetic diversity. *Proc Natl Acad Sci USA* **103,** 6611–6616 (2006).
68. Khan, H., Smit, A. & Boissinot, S. Molecular evolution and tempo of amplification of human LINE-1 retrotransposons since the origin of primates. *Genome Research* **16,** 78–87 (2006).
69. Yang, L., Brunsfeld, J., Scott, L. & Wichman, H. Reviving the dead: history and reactivation of an extinct I1. *PLOS Genet* **10,** e1004395 (2014).

70. Swergold, G. D. Identification, characterization, and cell specificity of a human LINE-1 promoter. *Molecular and Cellular Biology* **10,** 6718–6729 (1990).
71. Denli, A. M. *et al.* Primate-Specific ORF0 Contributes to Retrotransposon-Mediated Diversity. *Cell* **163,** 583–593 (2015).
72. Howell, R. & Usdin, K. The ability to form intrastrand tetraplexes is an evolutionarily conserved feature of the 3' end of L1 retrotransposons. *Molecular Biology and Evolution* **14,** 144–155 (1997).
73. Sahakyan, A. B., Murat, P., Mayer, C. & Balasubramanian, S. G-quadruplex structures within the 3' UTR of LINE-1 elements stimulate retrotransposition. *Nat Struct Mol Biol* **24,** 243–247 (2017).
74. Ade, C. M., Servant, G., Morales, M. E. & Roy-Engel, A. M. in *Human Retrotransposons in Health and Disease* **20,** 157–194 (Springer International Publishing, 2017).
75. Martin, S. L. *et al.* LINE-1 retrotransposition requires the nucleic acid chaperone activity of the ORF1 protein. *Journal of Molecular Biology* **348,** 549–561 (2005).
76. Martin, S. L. The ORF1 protein encoded by LINE-1: structure and function during L1 retrotransposition. *Journal of Biomedicine and Biotechnology* **2006,** 45621–6 (2006).
77. Khazina, E. *et al.* Trimeric structure and flexibility of the L1ORF1 protein in human L1 retrotransposition. *Nat Struct Mol Biol* **18,** 1006–1014 (2011).
78. Muthukumaran, R., Sangeetha, B. & Amutha, R. Conformational analysis on the wild type and mutated forms of human ORF1p: a molecular dynamics study. *Mol Biosyst* **11,** 1987–1999 (2015).
79. Januszyk, K. *et al.* Identification and solution structure of a highly conserved C-terminal domain within ORF1p required for retrotransposition of long interspersed nuclear element-1. *J. Biol. Chem.* **282,** 24893–24904 (2007).
80. Ostertag, E. M. & Kazazian, H. H. Biology of mammalian L1 retrotransposons. *Annu. Rev. Genet.* **35,** 501–538 (2001).
81. Cost, G. J. & Boeke, J. D. Targeting of Human Retrotransposon Integration Is Directed by the Specificity of the L1 Endonuclease for Regions of Unusual DNA Structure †. *Biochemistry* **37,** 18081–18093 (1998).
82. Kulpa, D. A. & Moran, J. V. Cis-preferential LINE-1 reverse transcriptase activity in ribonucleoprotein particles. *Nat Struct Mol Biol* **13,** 655–660 (2006).
83. Piskareva, O. & Schmatchenko, V. DNA polymerization by the reverse transcriptase of the human L1 retrotransposon on its own template in vitro. *FEBS Lett.* **580,** 661–668 (2006).
84. Piskareva, O., Ernst, C., Higgins, N. & Schmatchenko, V. The carboxy-terminal segment of the human LINE-1 ORF2 protein is involved in RNA binding. *Elsevier* **3,** 433–437 (2013).
85. Monot, C. *et al.* The Specificity and Flexibility of L1 Reverse Transcription Priming at Imperfect T-Tracts. *PLOS Genet* **9,** e1003499–18 (2013).
86. Naas, T. P. *et al.* An actively retrotransposing, novel subfamily of mouse L1 elements. *The EMBO journal* **17,** 590–597 (1998).
87. Goodier, J. L. A Novel Active L1 Retrotransposon Subfamily in the Mouse. *Genome Research* **11,** 1677–1685 (2001).
88. Richardson, S. R. *et al.* Heritable L1 retrotransposition in the mouse primordial germline and early embryo. *Genome Research* **27,** 1395–1405 (2017).

89. DeBerardinis, R. J. & Kazazian, H. H. Analysis of the promoter from an expanding mouse retrotransposon subfamily. *Genomics* **56,** 317–323 (1999).
90. Loeb, D. D. *et al.* The sequence of a large L1Md element reveals a tandemly repeated 5' end and several features found in retrotransposons. *Molecular and Cellular Biology* **6,** 168–182 (1986).
91. Becker, K. G., Swergold, G. D., Ozato, K. & Thayer, R. E. Binding of the ubiquitous nuclear transcription factor YY1 to a cis regulatory sequence in the human LINE-1 transposable element. *Hum. Mol. Genet.* **2,** 1697–1702 (1993).
92. Athanikar, J. N., Badge, R. M. & Moran, J. V. A YY1-binding site is required for accurate human LINE-1 transcription initiation. *Nucleic Acids Res.* **32,** 3846–3855 (2004).
93. Tchénio, T., Casella, J. F. & Heidmann, T. Members of the SRY family regulate the human LINE retrotransposons. *Nucleic Acids Res.* **28,** 411–415 (2000).
94. Yang, N. *et al.* An important role for RUNX3 in human L1 transcription and retrotransposition. *Nucleic Acids Res.* **31,** 4929–4940 (2003).
95. Sun, X. *et al.* Transcription factor profiling reveals molecular choreography and key regulators of human retrotransposon expression. *Proc. Natl. Acad. Sci. U.S.A.* **115,** E5526–E5535 (2018).
96. Lavie, L., Maldener, E., Brouha, B., Meese, E. U. & Mayer, J. The human L1 promoter: variable transcription initiation sites and a major impact of upstream flanking sequence on promoter activity. *Genome Research* **14,** 2253–2260 (2004).
97. Naufer, M. N. *et al.* L1 retrotransposition requires rapid ORF1p oligomerization, a novel coiled coil-dependent property conserved despite extensive remodeling. *Nucleic Acids Res.* **44,** 281–293 (2016).
98. Ardeljan, D. *et al.* LINE-1 ORF2p expression is nearly imperceptible in human cancers. *Mobile DNA* **11,** 1–19 (2020).
99. Wei, W. *et al.* Human L1 Retrotransposition: cis Preference versus trans Complementation. *Molecular and Cellular Biology* **21,** 1429–1439 (2001).
100. Taylor, M. S. *et al.* Affinity Proteomics Reveals Human Host Factors Implicated in Discrete Stages of LINE-1 Retrotransposition. *Cell* **155,** 1034–1048 (2013).
101. Goodier, J. L., Zhang, L., Vetter, M. R. & Kazazian, H. H. LINE-1 ORF1 protein localizes in stress granules with other RNA-binding proteins, including components of RNA interference RNA-induced silencing complex. *Molecular and Cellular Biology* **27,** 6469–6483 (2007).
102. Doucet, A. J. *et al.* Characterization of LINE-1 Ribonucleoprotein Particles. *PLOS Genet* **6,** e1001150–19 (2010).
103. Sultana, T., Zamborlini, A., Cristofari, G. & Lesage, P. Integration site selection by retroviruses and transposable elements in eukaryotes. *Nature Publishing Group* **18,** 292–308 (2017).
104. Cost, G. J., Feng, Q., Jacquier, A. & Boeke, J. D. Human L1 element target-primed reverse transcription in vitro. *The EMBO journal* **21,** 5899–5910 (2002).
105. Flasch, D. A. *et al.* Genome-wide de novo L1 Retrotransposition Connects Endonuclease Activity with Replication. *Cell* **177,** 837–851.e28 (2019).
106. Morrish, T. A. *et al.* Endonuclease-independent LINE-1 retrotransposition at mammalian telomeres. *Nature* **446,** 208–212 (2007).

107. Coufal, N. G. *et al.* 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 (2011).
108. Sen, S. K., Huang, C. T., Han, K. & Batzer, M. A. Endonuclease-independent insertion provides an alternative pathway for L1 retrotransposition in the human genome. *Nucleic Acids Res.* **35,** 3741–3751 (2007).
109. Morrish, T. A. *et al.* DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. *Nat. Genet.* **31,** 159–165 (2002).
110. Viollet, S., Monot, C. & Cristofari, G. L1 retrotransposition. *Mobile Genetic Elements* **4,** e28907–7 (2014).
111. Dombroski, B. A., Mathias, S. L., Nanthakumar, E., Scott, A. F. & Kazazian, H. H. Isolation of an active human transposable element. *Science* **254,** 1805–1808 (1991).
112. Moran, J. V. *et al.* High frequency retrotransposition in cultured mammalian cells. *Cell* **87,** 917–927 (1996).
113. Gilbert, N., Lutz-Prigge, S. & Moran, J. V. Genomic deletions created upon LINE-1 retrotransposition. *Cell* **110,** 315–325 (2002).
114. Symer, D. E. *et al.* Human l1 retrotransposition is associated with genetic instability in vivo. *Cell* **110,** 327–338 (2002).
115. Gilbert, N., Lutz, S., Morrish, T. A. & Moran, J. V. Multiple fates of L1 retrotransposition intermediates in cultured human cells. *Molecular and Cellular Biology* **25,** 7780–7795 (2005).
116. Jurka, J. Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons. *Proc Natl Acad Sci USA* **94,** 1872–1877 (1997).
117. Doucet, A. J., Wilusz, J. E., Miyoshi, T., Liu, Y. & Moran, J. V. A 3′ Poly(A) Tract Is Required for LINE-1 Retrotransposition. *Molecular Cell* **60,** 728–741 (2015).
118. Boissinot, S., Chevret, P. & Furano, A. V. L1 (LINE-1) retrotransposon evolution and amplification in recent human history. *Molecular Biology and Evolution* **17,** 915–928 (2000).
119. Ostertag, E. M. & Kazazian, H. H. Twin priming: a proposed mechanism for the creation of inversions in L1 retrotransposition. *Genome Research* **11,** 2059–2065 (2001).
120. Chen, J., Rattner, A. & Nathans, J. Effects of L1 retrotransposon insertion on transcript processing, localization and accumulation: lessons from the retinal degeneration 7 mouse and implications for the genomic ecology of L1 elements. *Hum. Mol. Genet.* **15,** 2146–2156 (2006).
121. Goodier, J. L., Ostertag, E. M. & Kazazian, H. H. Transduction of 3'-flanking sequences is common in L1 retrotransposition. *Hum. Mol. Genet.* **9,** 653–657 (2000).
122. Pickeral, O. K., Makałowski, W., Boguski, M. S. & Boeke, J. D. Frequent human genomic DNA transduction driven by LINE-1 retrotransposition. *Genome Research* **10,** 411–415 (2000).
123. Holmes, S. E., Dombroski, B. A., Krebs, C. M., Boehm, C. D. & Kazazian, H. H. A new retrotransposable human L1 element from the LRE2 locus on chromosome 1q produces a chimaeric insertion. *Nat. Genet.* **7,** 143–148 (1994).
124. Tubio, J. M. C. *et al.* Extensive transduction of nonrepetitive DNA mediated by L1 retrotransposition in cancer genomes. *Science* **345,** 1251343–1251343 (2014).

125. Solyom, S. *et al.* Pathogenic orphan transduction created by a nonreference LINE-1 retrotransposon. *Human Mutation* **33,** 369–371 (2012).
126. Miné, M. *et al.* A large genomic deletion in the PDHX gene caused by the retrotranspositional insertion of a full-length LINE-1 element. *Human Mutation* **28,** 137–142 (2007).
127. Jonsson, F., Burstedt, M., Kellgren, T. G. & Golovleva, I. Non-homologous recombination between Alu and LINE-1 repeats results in a 91 kb deletion in MERTK causing severe retinitis pigmentosa. *Mol. Vis.* **24,** 667–678 (2018).
128. Xia, Z. *et al.* LINE-1 retrotransposon-mediated DNA transductions in endometriosis associated ovarian cancers. *Gynecol. Oncol.* **147,** 642–647 (2017).
129. Pradhan, B. *et al.* Detection of subclonal L1 transductions in colorectal cancer by long-distance inverse-PCR and Nanopore sequencing. *Sci Rep* **7,** 14521–12 (2017).
130. Rodriguez-Martin, B. *et al.* Pan-cancer analysis of whole genomes identifies driver rearrangements promoted by LINE-1 retrotransposition. *Nature Publishing Group* **52,** 306–319 (2020).
131. Gasior, S. L., Wakeman, T. P., Xu, B. & Deininger, P. L. The Human LINE-1 Retrotransposon Creates DNA Double-strand Breaks. *Journal of Molecular Biology* **357,** 1383–1393 (2006).
132. Richardson, S. R. *et al.* The Influence of LINE-1 and SINE Retrotransposons on Mammalian Genomes. *Microbiol Spectr* **3,** MDNA3–0061–2014 (2015).
133. Speek, M. Antisense Promoter of Human L1 Retrotransposon Drives Transcription of Adjacent Cellular Genes. *Molecular and Cellular Biology* **21,** 1973–1985 (2001).
134. Nigumann, P., Redik, K., Mätlik, K. & Speek, M. Many Human Genes Are Transcribed from the Antisense Promoter of L1 Retrotransposon. *Genomics* **79,** 628–634 (2002).
135. Criscione, S. W. *et al.* Genome-wide characterization of human L1 antisense promoter-driven transcripts. *BMC Genomics* **17,** 1–15 (2016).
136. Mätlik, K., Redik, K. & Speek, M. L1 Antisense Promoter Drives Tissue-Specific Transcription of Human Genes. *Journal of Biomedicine and Biotechnology* **2006,** 1–16 (2006).
137. Wheelan, S. J., Aizawa, Y., Han, J. S. & Boeke, J. D. Gene-breaking: a new paradigm for human retrotransposon-mediated gene evolution. *Genome Research* **15,** 1073–1078 (2005).
138. Perepelitsa-Belancio, V. & Deininger, P. RNA truncation by premature polyadenylation attenuates human mobile element activity. *Nat. Genet.* **35,** 363–366 (2003).
139. Garcia-Perez, J. L. *et al.* Epigenetic silencing of engineered L1 retrotransposition events in human embryonic carcinoma cells. *Nature* **466,** 769–773 (2010).
140. Han, J. S., Szak, S. T. & Boeke, J. D. Transcriptional disruption by the L1 retrotransposon and implications for mammalian transcriptomes. *Nature* **429,** 268–274 (2004).
141. Kazazian, H. H. *et al.* Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. *Nature* **332,** 164–166 (1988).

142. Jacobs, F. M. J. *et al.* An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. *Nature* **516,** 242–245 (2014).
143. Goodier, J. L., Cheung, L. E. & Kazazian, H. H., Jr. Mapping the LINE1 ORF1 protein interactome reveals associated inhibitors of human retrotransposition. *Nucleic Acids Res.* **41,** 7401–7419 (2013).
144. Taylor, D. W. *et al.* Structures of the CRISPR-Cmr complex reveal mode of RNA target positioning. *Science* **348,** 581–585 (2015).
145. Moldovan, G.-L., Pfander, B. & Jentsch, S. PCNA, the Maestro of the Replication Fork. *Cell* **129,** 665–679 (2007).
146. Pizarro, J. G. & Cristofari, G. Post-Transcriptional Control of LINE-1 Retrotransposition by Cellular Host Factors in Somatic Cells. *Front. Cell Dev. Biol.* **4,** 210–9 (2016).
147. Goodier, J. L., Cheung, L. E. & Kazazian, H. H. MOV10 RNA helicase is a potent inhibitor of retrotransposition in cells. *PLOS Genet* **8,** e1002941 (2012).
148. Miyoshi, T., Makino, T. & Moran, J. V. Poly(ADP-Ribose) Polymerase 2 Recruits Replication Protein A to Sites of LINE-1 Integration to Facilitate Retrotransposition. *Molecular Cell* **75,** 1286–1298.e12 (2019).
149. Liu, N. *et al.* Selective silencing of euchromatic L1s revealed by genome-wide screens for L1 regulators. *Nature* **553,** 228–232 (2018).
150. Ardeljan, D. *et al.* Cell fitness screens reveal a conflict between LINE-1 retrotransposition and DNA replication. *Nat Struct Mol Biol* 1–28 (2020). doi:10.1038/s41594-020-0372-1
151. Mita, P. *et al.* BRCA1 and S phase DNA repair pathways restrict LINE-1 retrotransposition in human cells. *Nat Struct Mol Biol* **27,** 179–191 (2020).
152. Fadloun, A. *et al.* Chromatin signatures and retrotransposon profiling in mouse embryos reveal regulation of LINE-1 by RNA. *Nat Struct Mol Biol* **20,** 332–338 (2013).
153. Molaro, A. *et al.* Two waves of de novo methylation during mouse germ cell development. *Genes & Development* **28,** 1544–1549 (2014).
154. Muñoz-Lopez, M., Macia, A., Garcia-Cañadas, M., Badge, R. M. & Garcia-Perez, J. L. An epi [c] genetic battle. *Mobile Genetic Elements* **1,** 122–127 (2014).
155. Aravin, A. A. *et al.* A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. *Molecular Cell* **31,** 785–799 (2008).
156. Yoder, J. A., Walsh, C. P. & Bestor, T. H. Cytosine methylation and the ecology of intragenomic parasites. *Trends in Genetics* **13,** 335–340 (1997).
157. Bourc'his, D. & Bestor, T. H. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature* **431,** 96–99 (2004).
158. Barau, J. *et al.* The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. *Science* **354,** 909–912 (2016).
159. Castro-Diaz, N. *et al.* Evolutionally dynamic L1 regulation in embryonic stem cells. *Genes & Development* **28,** 1397–1409 (2014).
160. Jönsson, M. E. *et al.* Activation of neuronal genes via LINE-1 elements upon global DNA demethylation in human neural progenitors. *Nature Communications* **10,** 3182–11 (2019).
161. Babaian, A. & Mager, D. L. Endogenous retroviral promoter exaptation in human cancer. *Mobile DNA* **7,** 24–21 (2016).
162. Wolf, D. & Goff, S. P. Embryonic stem cells use ZFP809 to silence retroviral DNAs. *Nature* **458,** 1201–1204 (2009).

163. Quenneville, S. *et al.* The KRAB-ZFP/KAP1 system contributes to the early embryonic establishment of site-specific DNA methylation patterns maintained during development. *CellReports* **2,** 766–773 (2012).
164. Rowe, H. M. *et al.* TRIM28 repression of retrotransposon-based enhancers is necessary to preserve transcriptional dynamics in embryonic stem cells. *Genome Research* **23,** 452–461 (2013).
165. Ecco, G. *et al.* Transposable Elements and Their KRAB-ZFP Controllers Regulate Gene Expression in Adult Tissues. *Developmental Cell* **36,** 611–623 (2016).
166. Schmitges, F. W. *et al.* Multiparameter functional diversity of human C2H2 zinc finger proteins. *Genome Research* **26,** 1742–1752 (2016).
167. Theunissen, T. W. *et al.* Molecular Criteria for Defining the Naive Human Pluripotent State. *Cell Stem Cell* **19,** 502–515 (2016).
168. Corsinotti, A. *et al.* Global and stage specific patterns of Krüppel-associated-box zinc finger protein gene expression in murine early embryonic cells. *PLoS ONE* **8,** e56721 (2013).
169. Kunarso, G. *et al.* Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nat. Genet.* **42,** 631–634 (2010).
170. Matsui, T. *et al.* Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. *Nature* **464,** 927–931 (2010).
171. Rowe, H. M. *et al.* KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature* **463,** 237–240 (2010).
172. Turelli, P. *et al.* Interplay of TRIM28 and DNA methylation in controlling human endogenous retroelements. *Genome Research* **24,** 1260–1270 (2014).
173. Thomas, J. H. & Schneider, S. Coevolution of retroelements and tandem zinc finger genes. *Genome Research* **21,** 1800–1812 (2011).
174. Imbeault, M., Helleboid, P.-Y. & Trono, D. KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. *Nature* **543,** 550–554 (2017).
175. Walter, M., Teissandier, A., Pérez-Palacios, R. & Bourc'his, D. An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells. *Elife* **5,** R87 (2016).
176. Ficz, G. *et al.* Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature* **473,** 398–402 (2011).
177. Kang, J. *et al.* Simultaneous deletion of the methylcytosine oxidases Tet1 and Tet3 increases transcriptome variability in early embryogenesis. *Proc. Natl. Acad. Sci. U.S.A.* **112,** E4236–45 (2015).
178. la Rica, de, L. *et al.* TET-dependent regulation of retrotransposable elements in mouse embryonic stem cells. *Genome Biol.* **17,** 1–14 (2016).
179. Zhang, P. *et al.* L1 retrotransposition is activated by Ten-eleven-translocation protein 1 and repressed by methyl-CpG binding proteins. *Nucleus* **8,** 548–562 (2017).
180. Deniz, Ö., la Rica, de, L., Cheng, K. C. L., Spensberger, D. & Branco, M. R. SETDB1 prevents TET2-dependent activation of IAP retroelements in naïve embryonic stem cells. *Genome Biol.* **19,** 6–11 (2018).
181. Shen, L. *et al.* Tet3 and DNA replication mediate demethylation of both the maternal and paternal genomes in mouse zygotes. *Cell Stem Cell* **15,** 459–471 (2014).

182. Upton, K. R. *et al.* Ubiquitous L1 Mosaicism in Hippocampal Neurons. *Cell* **161,** 228–239 (2015).
183. Deniz, Ö., Frost, J. M. & Branco, M. R. Regulation of transposable elements by DNA modifications. *Nat Rev Genet* 1–15 (2019). doi:10.1038/s41576-019-0106-6
184. Siomi, M. C., Sato, K., Pezic, D. & Aravin, A. A. PIWI-interacting small RNAs: the vanguard of genome defence. *Nature Reviews Molecular Cell Biology 2016 17:5* **12,** 246–258 (2011).
185. Théron, E. *et al.* The interplay between the Argonaute proteins Piwi and Aub within Drosophila germarium is critical for oogenesis, piRNA biogenesis and TE silencing. *Nucleic Acids Res.* **46,** 10052–10065 (2018).
186. Brennecke, J. *et al.* Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. *Cell* **128,** 1089–1103 (2007).
187. Gunawardane, L. S. *et al.* A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in Drosophila. *Science* **315,** 1587–1590 (2007).
188. Carmell, M. A. *et al.* MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. *Developmental Cell* **12,** 503–514 (2007).
189. Kuramochi-Miyagawa, S. *et al.* DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. *Genes & Development* **22,** 908–917 (2008).
190. Kabayama, Y. *et al.* Roles of MIWI, MILI and PLD6 in small RNA regulation in mouse growing oocytes. *Nucleic Acids Res.* **45,** 5387–5398 (2017).
191. Marchetto, M. C. N. *et al.* Differential L1 regulation in pluripotent stem cells of humans and apes. *Nature* **503,** 525–529 (2013).
192. Heras, S. R. *et al.* The Microprocessor controls the activity of mammalian retrotransposons. *Nature Publishing Group* **20,** 1173–1181 (2013).
193. Yang, N. & Kazazian, H. H. L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells. *Nat Struct Mol Biol* **13,** 763–771 (2006).
194. Berrens, R. V. *et al.* An endosiRNA-Based Repression Mechanism Counteracts Transposon Activation during Global DNA Demethylation in Embryonic Stem Cells. *Cell Stem Cell* **21,** 694–703.e7 (2017).
195. Zhang, A. *et al.* RNase L restricts the mobility of engineered retrotransposons in cultured human cells. *Nucleic Acids Res.* **42,** 3803–3820 (2013).
196. Bogerd, H. P. *et al.* Cellular inhibitors of long interspersed element 1 and Alu retrotransposition. *Proc Natl Acad Sci USA* **103,** 8780–8785 (2006).
197. Chen, H. *et al.* APOBEC3A is a potent inhibitor of adeno-associated virus and retrotransposons. *Curr. Biol.* **16,** 480–485 (2006).
198. Hulme, A. E., Bogerd, H. P., Cullen, B. R. & Moran, J. V. Selective inhibition of Alu retrotransposition by APOBEC3G. *Gene* **390,** 199–205 (2007).
199. Wissing, S., Montano, M., Garcia-Perez, J. L., Moran, J. V. & Greene, W. C. Endogenous APOBEC3B restricts LINE-1 retrotransposition in transformed cells and human embryonic stem cells. *J. Biol. Chem.* **286,** 36427–36437 (2011).
200. Rice, G. I. *et al.* Mutations in ADAR1 cause Aicardi-Goutières syndrome associated with a type I interferon signature. *Nature Publishing Group* **44,** 1243–1248 (2012).

201. Thomas, C. A. *et al.* 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 (2017).
202. Mooslehner, K., Müller, U., Karls, U., Hamann, L. & Harbers, K. Structure and expression of a gene encoding a putative GTP-binding protein identified by provirus integration in a transgenic mouse strain. *Molecular and Cellular Biology* **11,** 886–893 (1991).
203. Li, X. *et al.* The MOV10 helicase inhibits LINE-1 mobility. *J. Biol. Chem.* **288,** 21148–21160 (2013).
204. Xu, H., Zhang, P., Liu, L. & Lee, M. Y. A novel PCNA-binding motif identified by the panning of a random peptide display library. *Biochemistry* **40,** 4512–4520 (2001).
205. Branciforte, D. & Martin, S. L. Developmental and cell type specificity of LINE-1 expression in mouse testis: implications for transposition. *Molecular and Cellular Biology* **14,** 2584–2592 (1994).
206. Trelogan, S. A. & Martin, S. L. Tightly regulated, developmentally specific expression of the first open reading frame from LINE-1 during mouse embryogenesis. *Proc Natl Acad Sci USA* **92,** 1520–1524 (1995).
207. Georgiou, I. *et al.* Retrotransposon RNA expression and evidence for retrotransposition events in human oocytes. *Hum. Mol. Genet.* **18,** 1221–1228 (2009).
208. Garcia-Perez, J. L., Doucet, A. J., Bucheton, A., Moran, J. V. & Gilbert, N. Distinct mechanisms for trans-mediated mobilization of cellular RNAs by the LINE-1 reverse transcriptase. *Genome Research* **17,** 602–611 (2007).
209. Macia, A. *et al.* Epigenetic Control of Retrotransposon Expression in Human Embryonic Stem Cells. *Molecular and Cellular Biology* **31,** 300–316 (2010).
210. Wissing, S. *et al.* Reprogramming somatic cells into iPS cells activates LINE-1 retroelement mobility. *Hum. Mol. Genet.* **21,** 208–218 (2011).
211. Kano, H. *et al.* L1 retrotransposition occurs mainly in embryogenesis and creates somatic mosaicism. *Genes & Development* **23,** 1303–1312 (2009).
212. van den Hurk, J. A. J. M. *et al.* L1 retrotransposition can occur early in human embryonic development. *Hum. Mol. Genet.* **16,** 1587–1592 (2007).
213. Klawitter, S. *et al.* Reprogramming triggers endogenous L1 and Alu retrotransposition in human induced pluripotent stem cells. *Nature Communications* **7,** 1–14 (2015).
214. Richardson, S. R. & Faulkner, G. J. Heritable L1 Retrotransposition Events During Development: Understanding Their Origins. *Bioessays* **40,** 1700189–14 (2018).
215. Muotri, A. R. *et al.* Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature* **435,** 903–910 (2005).
216. Coufal, N. G. *et al.* L1 retrotransposition in human neural progenitor cells. *Nature* **460,** 1127–1131 (2009).
217. Macia, A. *et al.* Engineered LINE-1 retrotransposition in nondividing human neurons. *Genome Research* **27,** 335–348 (2017).
218. Baillie, J. K. *et al.* Somatic retrotransposition alters the genetic landscape of the human brain. *Nature* **479,** 534–537 (2011).
219. Evrony, G. D. *et al.* Single-neuron sequencing analysis of L1 retrotransposition and somatic mutation in the human brain. *Cell* **151,** 483–496 (2012).

220. Evrony, G. D. *et al.* Cell Lineage Analysis in Human Brain Using Endogenous Retroelements. *Neuron* **85,** 49–59 (2015).
221. Erwin, J. A. *et al.* L1-associated genomic regions are deleted in somatic cells of the healthy human brain. *Nature Neuroscience* 1–12 (2016). doi:10.1038/nn.4388
222. Faulkner, G. J. & Garcia-Perez, J. L. L1 Mosaicism in Mammals: Extent, Effects, and Evolution. *Trends in Genetics* **33,** 802–816 (2017).
223. Bedrosian, T. A., Quayle, C., Novaresi, N. & Gage, F. H. Early life experience drives structural variation of neural genomes in mice. *Science* **359,** 1395–1399 (2018).
224. Bundo, M. *et al.* Increased L1 Retrotransposition in the Neuronal Genome in Schizophrenia. *Neuron* **81,** 306–313 (2014).
225. Belancio, V. P., Roy-Engel, A. M., Pochampally, R. R. & Deininger, P. Somatic expression of LINE-1 elements in human tissues. *Nucleic Acids Res.* **38,** 3909–3922 (2010).
226. Barbieri, D. *et al.* Thrombopoietin protects hematopoietic stem cells from retrotransposon-mediated damage by promoting an antiviral response. *J. Exp. Med.* **215,** 1463–1480 (2018).
227. O'Donnell, K. A., An, W., Schrum, C. T., Wheelan, S. J. & Boeke, J. D. Controlled insertional mutagenesis using a LINE-1 (ORFeus) gene-trap mouse model. *Proc. Natl. Acad. Sci. U.S.A.* **110,** E2706–13 (2013).
228. Grandi, F. C. *et al.* Retrotransposition creates sloping shores: a graded influence of hypomethylated CpG islands on flanking CpG sites. *Genome Research* **25,** 1135–1146 (2015).
229. Newkirk, S. J. *et al.* Intact piRNA pathway prevents L1 mobilization in male meiosis. *Proc. Natl. Acad. Sci. U.S.A.* **114,** E5635–E5644 (2017).
230. Narita, N. *et al.* Insertion of a 5‘ truncated L1 element into the 3’ end of exon 44 of the dystrophin gene resulted in skipping of the exon during splicing in a case of Duchenne muscular dystrophy. *Journal of Clinical Investigation* **91,** 1862–1867 (1993).
231. Payer, L. M. & Burns, K. H. Transposable elements in human genetic disease. *Nat Rev Genet* **20,** 760–772 (2019).
232. Hancks, D. C. & Kazazian, H. H. Roles for retrotransposon insertions in human disease. *Mobile DNA* **7,** 1–28 (2016).
233. Rickman, K. A. *et al.* Deficiency of UBE2T, the E2 Ubiquitin Ligase Necessary for FANCD2 and FANCI Ubiquitination, Causes FA-T Subtype of Fanconi Anemia. *CellReports* **12,** 35–41 (2015).
234. Muotri, A. R. *et al.* L1 retrotransposition in neurons is modulated by MeCP2. *Nature* **468,** 443–446 (2010).
235. Zhao, B. *et al.* Somatic LINE-1 retrotransposition in cortical neurons and non-brain tissues of Rett patients and healthy individuals. *PLOS Genet* **15,** e1008043 (2019).
236. Rodić, N. *et al.* Long Interspersed Element-1 Protein Expression Is a Hallmark of Many Human Cancers. *The American Journal of Pathology* **184,** 1280–1286 (2014).
237. Lee, E. *et al.* Landscape of somatic retrotransposition in human cancers. *Science* **337,** 967–971 (2012).

238. Helman, E. *et al.* Somatic retrotransposition in human cancer revealed by whole-genome and exome sequencing. *Genome Research* **24,** 1053–1063 (2014).
239. Solyom, S. *et al.* Extensive somatic L1 retrotransposition in colorectal tumors. *Genome Research* **22,** 2328–2338 (2012).
240. Burns, K. H. Transposable elements in cancer. *Nature Publishing Group* **17,** 415–424 (2017).
241. Tang, Z. *et al.* Human transposon insertion profiling: Analysis, visualization and identification of somatic LINE-1 insertions in ovarian cancer. *Proc Natl Acad Sci USA* **114,** E733–E740 (2017).
242. Miki, Y. *et al.* Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer. *Cancer Research* **52,** 643–645 (1992).
243. Cruickshanks, H. A. *et al.* Expression of a large LINE-1-driven antisense RNA is linked to epigenetic silencing of the metastasis suppressor gene TFPI-2 in cancer. *Nucleic Acids Res.* **41,** 6857–6869 (2013).
244. Chiappinelli, K. B. *et al.* Inhibiting DNA Methylation Causes an Interferon Response in Cancer via dsRNA Including Endogenous Retroviruses. *Cell* **162,** 974–986 (2015).
245. Oberdoerffer, P. & Sinclair, D. A. The role of nuclear architecture in genomic instability and ageing. *Nature Reviews Molecular Cell Biology 2016 17:5* **8,** 692–702 (2007).
246. Wood, J. G. *et al.* Chromatin-modifying genetic interventions suppress age-associated transposable element activation and extend life span in Drosophila. *Proc Natl Acad Sci USA* **113,** 11277–11282 (2016).
247. De Cecco, M. *et al.* Genomes of replicatively senescent cells undergo global epigenetic changes leading to gene silencing and activation of transposable elements. *Aging Cell* **12,** 247–256 (2013).
248. Elgin, S. C. R. & Grewal, S. I. S. Heterochromatin: silence is golden. *Curr. Biol.* **13,** R895–8 (2003).
249. Berger, S. L. The complex language of chromatin regulation during transcription. *Nature* **447,** 407–412 (2007).
250. Iskow, R. C. *et al.* Natural mutagenesis of human genomes by endogenous retrotransposons. *Cell* **141,** 1253–1261 (2010).
251. Reilly, M. T., Faulkner, G. J., Dubnau, J., Ponomarev, I. & Gage, F. H. The Role of Transposable Elements in Health and Diseases of the Central Nervous System. *Journal of Neuroscience* **33,** 17577–17586 (2013).
252. St Laurent, G., Hammell, N. & McCaffrey, T. A. A LINE-1 component to human aging: do LINE elements exact a longevity cost for evolutionary advantage? *Mechanisms of Ageing and Development* **131,** 299–305 (2010).
253. Van Meter, M. *et al.* SIRT6 represses LINE1 retrotransposons by ribosylating KAP1 but this repression fails with stress and age. *Nature Communications* **5,** 5011–10 (2014).
254. Oberdoerffer, P. *et al.* SIRT1 redistribution on chromatin promotes genomic stability but alters gene expression during aging. *Cell* **135,** 907–918 (2008).
255. Abe, M. *et al.* Impact of age-associated increase in 2'-O-methylation of miRNAs on aging and neurodegeneration in Drosophila. *Genes & Development* **28,** 44–57 (2014).
256. Ge, Z.-J., Schatten, H., Zhang, C.-L. & Sun, Q.-Y. Oocyte ageing and epigenetics. *Reproduction* **149,** R103–14 (2015).

257. De Cecco, M. *et al.* L1 drives IFN in senescent cells and promotes age-associated inflammation. *Nature* **566,** 1–33 (2019).
258. Simon, M. *et al.* LINE1 Derepression in Aged Wild-Type and SIRT6- Deficient Mice Drives Inflammation. *Cell Metabolism* **29,** 1–38 (2019).
259. Heidmann, T., Heidmann, O. & Nicolas, J. F. An indicator gene to demonstrate intracellular transposition of defective retroviruses. *Proc Natl Acad Sci USA* **85,** 2219–2223 (1988).
260. Kopera, H. C. *et al.* LINE-1 Cultured Cell Retrotransposition Assay. *Methods Mol. Biol.* **1400,** 139–156 (2016).
261. Garcia-Perez, J. L. *et al.* LINE-1 retrotransposition in human embryonic stem cells. *Hum. Mol. Genet.* **16,** 1569–1577 (2007).
262. Scott, E. C. *et al.* A hot L1 retrotransposon evades somatic repression and initiates human colorectal cancer. *Genome Research* **26,** 745–755 (2016).
263. RANGWALA, S. The L1 retrotransposition assay: A retrospective and toolkit. *Methods* **49,** 219–226 (2009).
264. Rangwala, S. H., Zhang, L. & Kazazian, H. H. Many LINE1 elements contribute to the transcriptome of human somatic cells. *Genome Biol.* **10,** R100–18 (2009).
265. Ostertag, E. M., Prak, E. T., DeBerardinis, R. J., Moran, J. V. & Kazazian, H. H. Determination of L1 retrotransposition kinetics in cultured cells. *Nucleic Acids Res.* **28,** 1418–1423 (2000).
266. Xie, Y., Rosser, J. M., Thompson, T. L., Boeke, J. D. & An, W. Characterization of L1 retrotransposition with high-throughput dual-luciferase assays. *Nucleic Acids Res.* **39,** e16–e16 (2010).
267. Ewing, A. D. & Kazazian, H. H. Whole-genome resequencing allows detection of many rare LINE-1 insertion alleles in humans. *Genome Research* **21,** 985–990 (2011).
268. Wang, J. *et al.* dbRIP: a highly integrated database of retrotransposon insertion polymorphisms in humans. *Human Mutation* **27,** 323–329 (2006).
269. Shukla, R. *et al.* Endogenous Retrotransposition Activates Oncogenic Pathways in Hepatocellular Carcinoma. *Cell* **153,** 101–111 (2013).
270. Richardson, S. R., Morell, S. & Faulkner, G. J. L1 Retrotransposons and Somatic Mosaicism in the Brain. *Annu. Rev. Genet.* **48,** 1–27 (2014).
271. Mir, A. A., Philippe, C. & Cristofari, G. euL1db: the European database of L1HS retrotransposon insertions in humans. *Nucleic Acids Res.* **43,** D43–D47 (2014).
272. Perrat, P. N. *et al.* Transposition-driven genomic heterogeneity in the Drosophila brain. *Science* **340,** 91–95 (2013).
273. Treiber, C. D. & Waddell, S. Resolving the prevalence of somatic transposition in Drosophila. *Elife* **6,** 2185 (2017).
274. De Cecco, M. *et al.* Transposable elements become active and mobile in the genomes of aging mammalian somatic tissues. *Aging (Albany NY)* **5,** 867–883 (2013).
275. Doucet-O'Hare, T. T. *et al.* LINE-1 expression and retrotransposition in Barrett's esophagus and esophageal carcinoma. *Proc Natl Acad Sci USA* **112,** E4894–E4900 (2015).
276. Tieland, M., Trouwborst, I. & Clark, B. C. Skeletal muscle performance and ageing. *J Cachexia Sarcopenia Muscle* **9,** 3–19 (2018).
277. Sherwood, L. *Human physiology: from cells to systems*. (Cengage Learning, 2010).

278. Hikida, R. S. Aging changes in satellite cells and their functions. *Curr Aging Sci* **4,** 279–297 (2011).
279. Wilkins, J. T., Krivickas, L. S., Goldstein, R., Suh, D. & Frontera, W. R. Contractile properties of adjacent segments of single human muscle fibers. *Muscle Nerve* **24,** 1319–1326 (2001).
280. Thomas, G. D. Functional muscle ischemia in Duchenne and Becker muscular dystrophy. *Front Physiol* **4,** 381 (2013).
281. Jayasinghe, I. D. & Launikonis, B. S. Three-dimensional reconstruction and analysis of the tubular system of vertebrate skeletal muscle. *J Cell Sci* **126,** 4048–4058 (2013).
282. Lamboley, C. R., Murphy, R. M., McKenna, M. J. & Lamb, G. D. Sarcoplasmic reticulum Ca2+ uptake and leak properties, and SERCA isoform expression, in type I and type II fibres of human skeletal muscle. *J. Physiol. (Lond.)* **592,** 1381–1395 (2014).
283. Dahl, R. *et al.* Three-dimensional reconstruction of the human skeletal muscle mitochondrial network as a tool to assess mitochondrial content and structural organization. *Acta Physiol (Oxf)* **213,** 145–155 (2015).
284. Frontera, W. R. & Ochala, J. Skeletal muscle: a brief review of structure and function. *Calcif. Tissue Int.* **96,** 183–195 (2015).
285. HUNT, C. C. & KUFFLER, S. W. Motor innervation of skeletal muscle: multiple innervation of individual muscle fibres and motor unit function. *J. Physiol. (Lond.)* **126,** 293–303 (1954).
286. Korthuis, R. J. Skeletal Muscle Circulation. (2011).
287. Clark, K. A., McElhinny, A. S., Beckerle, M. C. & Gregorio, C. C. Striated muscle cytoarchitecture: an intricate web of form and function. *Annu. Rev. Cell Dev. Biol.* **18,** 637–706 (2002).
288. Maruyama, K. & Ebashi, S. Alpha-actinin, a new structural protein from striated muscle. II. Action on actin. *J. Biochem.* **58,** 13–19 (1965).
289. Tajsharghi, H. Thick and thin filament gene mutations in striated muscle diseases. *IJMS* **9,** 1259–1275 (2008).
290. Brenner, B. Effect of Ca2+ on cross-bridge turnover kinetics in skinned single rabbit psoas fibers: implications for regulation of muscle contraction. *Proc Natl Acad Sci USA* **85,** 3265–3269 (1988).
291. HUXLEY, H. & HANSON, J. Changes in the cross-striations of muscle during contraction and stretch and their structural interpretation. *Nature* **173,** 973–976 (1954).
292. Hynes, T. R., Block, S. M., White, B. T. & Spudich, J. A. Movement of myosin fragments in vitro: domains involved in force production. *Cell* **48,** 953–963 (1987).
293. Spudich, J. A. The myosin swinging cross-bridge model. *Nature Reviews Molecular Cell Biology 2016 17:5* **2,** 387–392 (2001).
294. LORAND, L. Adenosine triphosphate-creatine transphosphorylase as relaxing factor of muscle. *Nature* **172,** 1181–1183 (1953).
295. Lehman, W., Craig, R. & Vibert, P. Ca(2+)-induced tropomyosin movement in Limulus thin filaments revealed by three-dimensional reconstruction. *Nature* **368,** 65–67 (1994).

296. Etgen, G. J., Wilson, C. M., Jensen, J., Cushman, S. W. & Ivy, J. L. Glucose transport and cell surface GLUT-4 protein in skeletal muscle of the obese Zucker rat. *Am. J. Physiol.* **271,** E294–301 (1996).
297. Larance, M., Ramm, G. & James, D. E. The GLUT4 code. *Molecular Endocrinology* **22,** 226–233 (2008).
298. Otto Buczkowska, E. & Dworzecki, T. [The role of skeletal muscle in the regulation of glucose homeostasis]. *Endokrynol Diabetol Chor Przemiany Materii Wieku Rozw* **9,** 93–97 (2003).
299. Stump, C. S., Henriksen, E. J., Wei, Y. & Sowers, J. R. The metabolic syndrome: role of skeletal muscle metabolism. *Ann. Med.* **38,** 389–402 (2006).
300. Wolfe, R. R. The underappreciated role of muscle in health and disease. *Am. J. Clin. Nutr.* **84,** 475–482 (2006).
301. Schnyder, S. & Handschin, C. Skeletal muscle as an endocrine organ: PGC-1α, myokines and exercise. *Bone* **80,** 115–125 (2015).
302. Chaillou, T. Skeletal Muscle Fiber Type in Hypoxia: Adaptation to High-Altitude Exposure and Under Conditions of Pathological Hypoxia. *Front Physiol* **9,** 1450 (2018).
303. Qaisar, R., Bhaskaran, S. & Van Remmen, H. Muscle fiber type diversification during exercise and regeneration. *Free Radical Biology and Medicine* **98,** 56–67 (2016).
304. Schiaffino, S. & Reggiani, C. Fiber Types in Mammalian Skeletal Muscles. *Physiological Reviews* **91,** 1447–1531 (2011).
305. Bentzinger, C. F., Wang, Y. X. & Rudnicki, M. A. Building muscle: molecular regulation of myogenesis. *Cold Spring Harb Perspect Biol* **4,** a008342–a008342 (2012).
306. Chal, J. & Pourquié, O. Making muscle: skeletal myogenesis in vivo and in vitro. *Development* **144,** 2104–2122 (2017).
307. Hirsinger, E., Jouve, C., Dubrulle, J. & Pourquié, O. Somite formation and patterning. *Int. Rev. Cytol.* **198,** 1–65 (2000).
308. Chevallier, A. Role of the somitic mesoderm in the development of the thorax in bird embryos. II. Origin of thoracic and appendicular musculature. *J Embryol Exp Morphol* **49,** 73–88 (1979).
309. Jacob, M., Christ, B. & Jacob, H. J. The migration of myogenic cells from the somites into the leg region of avian embryos. An ultrastructural study. *Anat. Embryol.* **157,** 291–309 (1979).
310. Berkes, C. A. & Tapscott, S. J. MyoD and the transcriptional control of myogenesis. *Seminars in Cell and Developmental Biology* **16,** 585–595 (2005).
311. Pownall, M. E., Gustafsson, M. K. & Emerson, C. P. Myogenic regulatory factors and the specification of muscle progenitors in vertebrate embryos. *Annu. Rev. Cell Dev. Biol.* **18,** 747–783 (2002).
312. Rudnicki, M. A. *et al.* MyoD or Myf-5 is required for the formation of skeletal muscle. *Cell* **75,** 1351–1359 (1993).
313. Tapscott, S. J. The circuitry of a master switch: Myod and the regulation of skeletal muscle gene transcription. *Development* **132,** 2685–2695 (2005).
314. Biressi, S., Molinaro, M. & Cossu, G. Cellular heterogeneity during vertebrate skeletal muscle development. *Developmental Biology* **308,** 281–293 (2007).
315. Stockdale, F. E. Myogenic cell lineages. *Developmental Biology* **154,** 284–298 (1992).

316. Kelly, R. G. *et al.* Embryonic and fetal myogenic programs act through separate enhancers at the MLC1F/3F locus. *Developmental Biology* **187,** 183–199 (1997).
317. White, R. B., Biérinx, A.-S., Gnocchi, V. F. & Zammit, P. S. Dynamics of muscle fibre growth during postnatal mouse development. *BMC Dev. Biol.* **10,** 21–11 (2010).
318. Gros, J., Manceau, M., Thomé, V. & Marcelle, C. A common somitic origin for embryonic muscle progenitors and satellite cells. *Nature* **435,** 954–958 (2005).
319. Kassar-Duchossoy, L. *et al.* Pax3/Pax7 mark a novel population of primitive myogenic cells during development. *Genes & Development* **19,** 1426–1431 (2005).
320. Relaix, F., Rocancourt, D., Mansouri, A. & Buckingham, M. A Pax3/Pax7-dependent population of skeletal muscle progenitor cells. *Nature* **435,** 948–953 (2005).
321. MAURO, A. Satellite cell of skeletal muscle fibers. *J Biophys Biochem Cytol* **9,** 493–495 (1961).
322. Bentzinger, C. F. *et al.* Fibronectin regulates Wnt7a signaling and satellite cell expansion. *Cell Stem Cell* **12,** 75–87 (2013).
323. Yin, H., Price, F. & Rudnicki, M. A. Satellite Cells and the Muscle Stem Cell Niche. *Physiological Reviews* **93,** 23–67 (2013).
324. Castiglioni, A. *et al.* FOXP3+ T Cells Recruited to Sites of Sterile Skeletal Muscle Injury Regulate the Fate of Satellite Cells and Guide Effective Tissue Regeneration. *PLoS ONE* **10,** e0128094 (2015).
325. Burzyn, D. *et al.* A special population of regulatory T cells potentiates muscle repair. *Cell* **155,** 1282–1295 (2013).
326. Joe, A. W. B. *et al.* Muscle injury activates resident fibro/adipogenic progenitors that facilitate myogenesis. *Nat Cell Biol* **12,** 153–163 (2010).
327. White, J. D. *et al.* Myotube formation is delayed but not prevented in MyoD-deficient skeletal muscle: studies in regenerating whole muscle grafts of adult mice. *J. Histochem. Cytochem.* **48,** 1531–1544 (2000).
328. Yablonka-Reuveni, Z. *et al.* The transition from proliferation to differentiation is delayed in satellite cells from mice lacking MyoD. *Developmental Biology* **210,** 440–455 (1999).
329. Cooper, R. N. *et al.* In vivo satellite cell activation via Myf5 and MyoD in regenerating mouse skeletal muscle. *J Cell Sci* **112 (Pt 17),** 2895–2901 (1999).
330. Faralli, H. & Dilworth, F. J. Turning on myogenin in muscle: a paradigm for understanding mechanisms of tissue-specific gene expression. *Comp. Funct. Genomics* **2012,** 836374–10 (2012).
331. Sacco, A., Doyonnas, R., Kraft, P., Vitorovic, S. & Blau, H. M. Self-renewal and expansion of single transplanted muscle stem cells. *Nature* **456,** 502–506 (2008).
332. Blau, H. M., Cosgrove, B. D. & Ho, A. T. V. The central role of muscle stem cells in regenerative failure with aging. *Nat Med* **21,** 854–862 (2015).
333. Sousa-Victor, P., García-Prat, L., Serrano, A. L., Perdiguero, E. & Muñoz-Cánoves, P. Muscle stem cell aging: regulation and rejuvenation. *Trends Endocrinol. Metab.* **26,** 287–296 (2015).
334. Rosenblatt, J. D., Lunt, A. I., Parry, D. J. & Partridge, T. A. Culturing satellite cells from living single muscle fiber explants. *In Vitro Cell. Dev. Biol. Anim.* **31,** 773–779 (1995).

335. Schmalbruch, H. & Hellhammer, U. The number of nuclei in adult rat muscles with special reference to satellite cells. *Anat. Rec.* **189,** 169–175 (1977).
336. Dumont, N. A., Bentzinger, C. F., Sincennes, M.-C. & Rudnicki, M. A. *Satellite Cells and Skeletal Muscle Regeneration*. **139,** 1027–1059 (John Wiley & Sons, Inc., 2011).
337. McGreevy, J. W., Hakim, C. H., McIntosh, M. A. & Duan, D. Animal models of Duchenne muscular dystrophy: from basic mechanisms to gene therapy. *Dis Model Mech* **8,** 195–213 (2015).
338. Cornelison, D. D. W. Context matters: in vivo and in vitro influences on muscle satellite cell activity. *J. Cell. Biochem.* **105,** 663–669 (2008).
339. Truskey, G. A. Development and application of human skeletal muscle microphysiological systems. *Lab Chip* **18,** 3061–3073 (2018).
340. Murphy, M. M., Lawson, J. A., Mathew, S. J., Hutcheson, D. A. & Kardon, G. Satellite cells, connective tissue fibroblasts and their interactions are crucial for muscle regeneration. *Development* **138,** 3625–3637 (2011).
341. Liu, L., Cheung, T. H., Charville, G. W. & Rando, T. A. Isolation of skeletal muscle stem cells by fluorescence-activated cell sorting. *Nat Protoc* **10,** 1612–1624 (2015).
342. Conboy, M. J., Cerletti, M., Wagers, A. J. & Conboy, I. M. Immuno-analysis and FACS sorting of adult muscle fiber-associated stem/precursor cells. *Methods Mol. Biol.* **621,** 165–173 (2010).
343. Pasut, A., Oleynik, P. & Rudnicki, M. A. Isolation of muscle stem cells by fluorescence activated cell sorting cytometry. *Methods Mol. Biol.* **798,** 53–64 (2012).
344. Cerletti, M. *et al.* Highly efficient, functional engraftment of skeletal muscle stem cells in dystrophic muscles. *Cell* **134,** 37–47 (2008).
345. Rao, N. *et al.* Fibroblasts influence muscle progenitor differentiation and alignment in contact independent and dependent manners in organized co-culture devices. *Biomed Microdevices* **15,** 161–169 (2013).
346. Hindi, L., McMillan, J. D., Afroze, D., Hindi, S. M. & Kumar, A. Isolation, Culturing, and Differentiation of Primary Myoblasts from Skeletal Muscle of Adult Mice. *Bio Protoc* **7,** (2017).
347. Jiwlawat, S. *et al.* Differentiation and sarcomere formation in skeletal myocytes directly prepared from human induced pluripotent stem cells using a sphere-based culture. *Differentiation* **96,** 70–81 (2017).
348. Trotter, J. A. & Nameroff, M. Myoblast differentiation in vitro: morphological differentiation of mononucleated myoblasts. *Developmental Biology* **49,** 548–555 (1976).
349. Gilbert, P. M. *et al.* Substrate elasticity regulates skeletal muscle stem cell self-renewal in culture. *Science* **329,** 1078–1081 (2010).
350. Guilak, F. *et al.* Control of stem cell fate by physical interactions with the extracellular matrix. *Cell Stem Cell* **5,** 17–26 (2009).
351. Discher, D. E., Janmey, P. & Wang, Y.-L. Tissue cells feel and respond to the stiffness of their substrate. *Science* **310,** 1139–1143 (2005).
352. Calve, S., Odelberg, S. J. & Simon, H.-G. A transitional extracellular matrix instructs cell behavior during muscle regeneration. *Developmental Biology* **344,** 259–271 (2010).

353. Denes, L. T. *et al.* Culturing C2C12 myotubes on micromolded gelatin hydrogels accelerates myotube maturation. *Skeletal Muscle* **9,** 17–10 (2019).
354. Lacraz, G. *et al.* Increased Stiffness in Aged Skeletal Muscle Impairs Muscle Progenitor Cell Proliferative Activity. *PLoS ONE* **10,** e0136217 (2015).
355. Thorley, M. *et al.* Skeletal muscle characteristics are preserved in hTERT/cdk4 human myogenic cell lines. *Skeletal Muscle* **6,** 43–12 (2016).
356. Mamchaoui, K. *et al.* Immortalized pathological human myoblasts: towards a universal tool for the study of neuromuscular disorders. *Skeletal Muscle* **1,** 34–11 (2011).
357. Yaffe, D. & Saxel, O. Serial passaging and differentiation of myogenic cells isolated from dystrophic mouse muscle. *Nature* **270,** 725–727 (1977).
358. Li, Z. *et al.* Electrical pulse stimulation induces GLUT4 translocation in C2C12 myotubes that depends on Rab8A, Rab13, and Rab14. *Am. J. Physiol. Endocrinol. Metab.* **314,** E478–E493 (2018).
359. Hsu, D. K. *et al.* Identification of a murine TEF-1-related gene expressed after mitogenic stimulation of quiescent fibroblasts and during myogenic differentiation. *J. Biol. Chem.* **271,** 13786–13795 (1996).
360. Kessler, P. D. *et al.* Gene delivery to skeletal muscle results in sustained expression and systemic delivery of a therapeutic protein. *Proc Natl Acad Sci USA* **93,** 14082–14087 (1996).
361. Krom, Y. D. *et al.* Intrinsic epigenetic regulation of the D4Z4 macrosatellite repeat in a transgenic mouse model for FSHD. *PLOS Genet* **9,** e1003415 (2013).
362. Dandapat, A. *et al.* Dominant lethal pathologies in male mice engineered to contain an X-linked DUX4 transgene. *CellReports* **8,** 1484–1496 (2014).
363. Bosnakovski, D. *et al.* Muscle pathology from stochastic low level DUX4 expression in an FSHD mouse model. *Nature Communications* **8,** 1–9 (2017).
364. Kang, P. B. & Griggs, R. C. Advances in Muscular Dystrophies. *JAMA Neurol* **72,** 741–742 (2015).
365. Mah, J. K. *et al.* A systematic review and meta-analysis on the epidemiology of Duchenne and Becker muscular dystrophy. *Neuromuscul. Disord.* **24,** 482–491 (2014).
366. Lapidos, K. A., Kakkar, R. & McNally, E. M. The dystrophin glycoprotein complex: signaling strength and integrity for the sarcolemma. *Circ. Res.* **94,** 1023–1031 (2004).
367. Rahimov, F. & Kunkel, L. M. The cell biology of disease: cellular and molecular mechanisms underlying muscular dystrophy. *J Cell Biol* **201,** 499–510 (2013).
368. Padberg, G. W. *et al.* Facioscapulohumeral muscular dystrophy in the Dutch population. *Muscle Nerve Suppl* S81–4 (1995).
369. Mostacciuolo, M. L. *et al.* Facioscapulohumeral muscular dystrophy: epidemiological and molecular study in a north-east Italian population sample. *Clin. Genet.* **75,** 550–555 (2009).
370. Lamperti, C. *et al.* A standardized clinical evaluation of patients affected by facioscapulohumeral muscular dystrophy: The FSHD clinical score. *Muscle Nerve* **42,** 213–217 (2010).
371. Padberg, G. W. Facioscapulohumeral disease. **22,** (1982).
372. Lassche, S. *et al.* Sarcomeric dysfunction contributes to muscle weakness in facioscapulohumeral muscular dystrophy. *Neurology* **80,** 733–737 (2013).

373. Celegato, B. *et al.* Parallel protein and transcript profiles of FSHD patient muscles correlate to the D4Z4 arrangement and reveal a common impairment of slow to fast fibre differentiation and a general deregulation of MyoD-dependent genes. *Proteomics* **6,** 5303–5321 (2006).
374. van Deutekom, J. C. *et al.* FSHD associated DNA rearrangements are due to deletions of integral copies of a 3.2 kb tandemly repeated unit. *Hum. Mol. Genet.* **2,** 2037–2042 (1993).
375. Wijmenga, C. *et al.* Chromosome 4q DNA rearrangements associated with facioscapulohumeral muscular dystrophy. *Nat. Genet.* **2,** 26–30 (1992).
376. Snider, L. *et al.* Facioscapulohumeral Dystrophy: Incomplete Suppression of a Retrotransposed Gene. *PLOS Genet* **6,** e1001181–14 (2010).
377. van der Maarel, S. M., Tawil, R. & Tapscott, S. J. Facioscapulohumeral muscular dystrophy and DUX4: breaking the silence. *Trends Mol Med* **17,** 252–258 (2011).
378. Lemmers, R. J. L. F. *et al.* Digenic inheritance of an SMCHD1 mutation and an FSHD-permissive D4Z4 allele causes facioscapulohumeral muscular dystrophy type 2. *Nat. Genet.* **44,** 1370–1374 (2012).
379. van Overveld, P. G. M. *et al.* Hypomethylation of D4Z4 in 4q-linked and non-4q-linked facioscapulohumeral muscular dystrophy. *Nat. Genet.* **35,** 315–317 (2003).
380. Lemmers, R. J. L. F. *et al.* Facioscapulohumeral muscular dystrophy is uniquely associated with one of the two variants of the 4q subtelomere. *Nat. Genet.* **32,** 235–236 (2002).
381. Shadle, S. C. *et al.* DUX4-induced bidirectional HSATII satellite repeat transcripts form intranuclear double stranded RNA foci in human cell models of FSHD. *Hum. Mol. Genet.* **6,** 941 (2019).
382. Whiddon, J. L., Langford, A. T., Wong, C.-J., Zhong, J. W. & Tapscott, S. J. Conservation and innovation in the DUX4-family gene network. *Nat. Genet.* **25,** 4577–8 (2017).
383. De Iaco, A. *et al.* DUX-family transcription factors regulate zygotic genome activation in placental mammals. *Nature Publishing Group* **49,** 941–945 (2017).
384. Percharde, M. *et al.* A LINE1-Nucleolin Partnership Regulates Early Development and ESC Identity. *Cell* **174,** 1–35 (2018).
385. Lek, A., Rahimov, F., Jones, P. L. & Kunkel, L. M. Emerging preclinical animal models for FSHD. *Trends Mol Med* **21,** 295–306 (2015).
386. Doucet, A. J., Basyuk, E. & Gilbert, N. Cellular Localization of Engineered Human LINE-1 RNA and Proteins. *Methods Mol. Biol.* **1400,** 281–297 (2016).
387. Dewannieux, M. & Heidmann, T. L1-mediated retrotransposition of murine B1 and B2 SINEs recapitulated in cultured cells. *Journal of Molecular Biology* **349,** 241–247 (2005).
388. Block, G. J. *et al.* Wnt/β-catenin signaling suppresses DUX4 expression and prevents apoptosis of FSHD muscle cells. *Hum. Mol. Genet.* **22,** 4661–4672 (2013).
389. Deininger, P. *et al.* A comprehensive approach to expression of L1 loci. *Nucleic Acids Res.* **45,** e31–e31 (2016).
390. Mumford, P. W. *et al.* Skeletal muscle LINE-1 retrotransposon activity is upregulated in older versus younger rats. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **317,** R397–R406 (2019).

391. Roberson, P. A. *et al.* Skeletal muscle LINE-1 ORF1 mRNA is higher in older humans but decreases with endurance exercise and is negatively associated with higher physical activity. *J. Appl. Physiol.* **127,** 895–904 (2019).
392. Zhong, X. *et al.* Ten-Eleven Translocation-2 (Tet2) Is Involved in Myogenic Differentiation of Skeletal Myoblast Cells in Vitro. *Sci Rep* **7,** 43539–11 (2017).
393. Tsumagari, K. *et al.* Early de novo DNA methylation and prolonged demethylation in the muscle lineage. *Epigenetics* **8,** 317–332 (2013).
394. Dewannieux, M., Esnault, C. & Heidmann, T. LINE-mediated retrotransposition of marked Alu sequences. *Nat. Genet.* **35,** 41–48 (2003).
395. Winokur, S. T. *et al.* Facioscapulohumeral muscular dystrophy (FSHD) myoblasts demonstrate increased susceptibility to oxidative stress. *Neuromuscular Disorders* **13,** 322–333 (2003).
396. Dmitriev, P. *et al.* DUX4-induced constitutive DNA damage and oxidative stress contribute to aberrant differentiation of myoblasts from FSHD patients. *Free Radical Biology and Medicine* **99,** 244–258 (2016).
397. Yu, Y., Cui, Y., Niedernhofer, L. J. & Wang, Y. Occurrence, Biological Consequences, and Human Health Relevance of Oxidative Stress-Induced DNA Damage. *Chem. Res. Toxicol.* **29,** 2008–2039 (2016).
398. Finkel, T. & Holbrook, N. J. Oxidants, oxidative stress and the biology of ageing. *Nature* **408,** 239–247 (2000).
399. Wallace, L. M. *et al.* DUX4, a candidate gene for facioscapulohumeral muscular dystrophy, causes p53-dependent myopathy in vivo. *Ann Neurol.* **69,** 540–552 (2011).
400. Young, J. M. *et al.* DUX4 binding to retroelements creates promoters that are active in FSHD muscle and testis. *PLOS Genet* **9,** e1003947 (2013).
401. Wylie, A. *et al.* p53 genes function to restrain mobile elements. *Genes & Development* **30,** 64–77 (2016).
402. Agamanolis, D. P., Dasu, S. & Krill, C. E. Tumors of skeletal muscle. *Hum. Pathol.* **17,** 778–795 (1986).