TÉCNICO LISBOA

Establishing cellular models to investigate the role of LINE-1 elements in X-chromosome inactivation and genome stability

Inês Cardial Mendes Dias

Thesis to obtain the Master of Science Degree in

Biological Engineering

External supervisor: Dr. Anne-Valerie Gendrel

Internal supervisor: Prof. Cláudia Alexandra Martins Lobato da Silva

Examination Committee

Chairperson: Prof. Maria Margarida Fonseca Rodrigues Diogo

External supervisor: Dr. Anne-Valerie Gendrel

Member of the Committee: Dr. Simão José Teixeira da Rocha

November 2021

Declaration

I declare that this document is an original work of my own authorship and that it fulfills all the requirements of the Code of Conduct and Good Practices of the Universidade de Lisboa.

Preface

The work presented in this thesis was performed at the Chromatin & Epigenetic laboratory, Institute of Molecular Medicine, University of Lisbon (Lisbon, Portugal), during the period of March-October 2021, under the supervision of Dr. Anne-Valerie Gendrel. The Thesis was co-supervised at Instituto Superior Técnico by Prof. Cláudia Lobato da Silva.

Acknowledgments

A special acknowledgment to Doctor Sérgio de Almeida, for giving me the opportunity of developing my master thesis at his laboratory, and for always being available to advise me and discuss ideas and results.

A cheerful thanks to my external supervisor, Anne-Valerie Gendrel, for welcoming me in her research project and for teaching and guiding me along this journey. I am very grateful for the availability of Anne-Valerie, who listened to my endless questions about the project. Also, for the advice and support, whenever they were necessary.

I deeply thank Eduardo for always helping me with everything, for introducing me to the laboratory techniques since the very first day, and mostly, for his patience, his support, and his friendship.

I am also very thankful for all the other members of SAlmeida lab, for such a warm welcoming and for always helping me, whenever it was necessary. And to João, Madalena, Inês and Cristiana, a very special thanks for always making me laugh with your amazing jokes and for making me feel at home since the beginning.

Last but not least, I thank Professor Cláudia Lobato da Silva for accepting to be my internal supervisor and for always being available to clarify my doubts and guide me along these last months.

Resumo

Os retrotransposões representam uma parte considerável do genoma do ratinho e, entre estes, os elementos LINE-1 têm mantido uma atividade elevada na linhagem de ratinho, bem como a capacidade de retrotransposição. Apesar de este fenómeno poder originar instabilidade genómica e doença, os elementos LINE-1 têm sido considerados importantes impulsionadores da evolução genómica e uma forma alternativa de controlar a expressão génica. No entanto, apesar da sua abundância e de evidência científica que aponta para a importância dos transposões em variados processos de regulação génica, estes elementos são frequentemente "ignorados" em estudos genómicos devido à sua natureza repetitiva no genoma. Neste projeto, foram estabelecidos modelos celulares para investigar o papel dos elementos LINE-1 no processo biológico de inativação do cromossoma X e na estabilidade genómica, durante a diferenciação celular e em células diferenciadas. Os sistemas de ativação e repressão de elementos LINE-1, que já tinham sido previamente estabelecidos em células estaminais embrionárias e em células progenitoras neurais, foram parcialmente otimizados. O sistema de ativação da expressão de elementos LINE-1 foi também implementado em fibroblastos (NIH-3T3). Os resultados obtidos indicam que estes sistemas estão funcionais e permitem perturbar a expressão de elementos LINE-1, apesar de algumas melhorias futuras serem necessárias ao nível das células diferenciadas.

Palavras-chave

Transposões, LINE-1, CRISPR-Cas9, expressão génica, inativação do cromossoma X, estabilidade genómica.

Abstract

Retrotransposons represent a considerable part of the mouse genome, and, among them, long interspersed nuclear element-1 (LINE-1) has remained highly active and capable of retrotransposition. Although this can lead to genome instability and disease, we are now able to appreciate that LINE-1 elements are major drivers of genome evolution and an alternative way of controlling gene expression. Despite their abundance and evidence suggesting the importance of transposable elements (TEs) in rewiring gene regulatory networks, these elements remain frequently “ignored” in genomic studies, due to their repetitive nature in the genome. In this project, cellular models to investigate the role of LINE-1 elements in X-chromosome inactivation and genome stability, during differentiation and in differentiated cells, were established. Engineered transcriptional activation and repression systems of LINE-1 elements that were already established in female mouse embryonic stem cells (mESCs) and female neural progenitor cells (NPCs) were partially optimized and a similar LINE-1 transcriptional activation system was implemented in female mouse embryonic fibroblasts (NIH-3T3 cells). The obtained results indicate that genome-wide perturbation of LINE-1 elements can be achieved with these systems, although future improvements must be performed at the level of differentiated cells.

Keywords

Transposable elements, LINE-1, CRISPR-Cas9, gene expression, X-chromosome inactivation, genome stability.

Table of contents

List of figures

Abbreviations

ASP – Antisense promoter
BSA – Bovine serum albumin
CARGO – Chimeric array of gRNA oligos
Cas – CRISPR-associated
CBh – Constitutive hybrid human cytomegalovirus enhancer/chicken β-actin
cDNA – Complementary DNA
CRISPR – Clustered regularly interspaced short palindromic repeats
CRISPRa – CRISPR activation
CRISPRi – CRISPR inactivation
Ct – Threshold cycle
DAPI – 4',6-Diamidino-2-phenylindole
dCas9 – Nuclease-dead Cas9
DMEM – Dulbecco's Modified Eagle's Medium
DNMT – DNA methyltransferase
dNTP – Deoxynucleotide triphosphate
Dox – Doxycycline
EGF – Epidermal growth factor
ERV – Endogenous retrovirus
ES – Embryonic stem
ESC – Embryonic stem cell
FBS – Fetal bovine serum
FGF – Fibroblast growth factor-basic
gDNA – Genomic DNA
gRNA – Guide RNA
H3K9me2 – Histone H3 lysine 9 dimethylation
H4K20me3 – Histone H4 lysine 20 trimethylation
HGP – Human Genome Project
hU6 – Human U6
ICM – Inner cell mass
KRAB – Kruppel associated box
LAD – Lamina-associated domain
LIF – Leukemia inhibitory factor
LINE – Long interspersed nuclear element
LINE-1, L1 – Long interspersed nuclear element-1
lncRNA – Long noncoding RNA
LTR – Long terminal repeat
mESC – Murine embryonic stem cell
mU6 – Mouse U6
NA – Numerical aperture

NAD – Nucleolus-associated domain
NPC – Neural progenitor cell
ORF – Open reading frame
PAGE – Polyacrylamide gel electrophoresis
PAM – Protospacer adjacent motif
PBS – Phosphate-saline buffer
PCR – Polymerase chain reaction
PFA – Paraformaldehyde
PGC – Primordial germ cell
qPCR – Quantitative polymerase chain reaction
RNA Pol – RNA polymerase
RNAi – RNA interference
RNP – Ribonucleoprotein
RT – Reverse transcriptase
RT-qPCR – Reverse transcription quantitative polymerase chain reaction
SDS – Sodium dodecyl sulfate
sgRNA – Single guide RNA
TALE – Transcription-activator-like effector
TDR – Target site direct repeat
TE – Transposable element
TPRT – Target site-primed reverse transcription
UTR – Untranslated region
XCI – X-chromosome inactivation
Xi – Inactive X
Xic – X-inactivation center
Xist – X-inactive-specific-transcript
ZFP – Zinc finger proteins

1. INTRODUCTION

1.1. Overview of the human and the mouse genomes and their repeat content

The year of 1990 was marked by the beginning of a pioneer project, which goal was to determine the complete nucleotide sequence of the human genome - the Human Genome Project (HGP). In early 2001, a draft sequence of the human genome was published by the International Human Genome Sequencing Consortium (Lander et al., 2001) and although covering around 90% of the euchromatic human genome, with only 35% in finished form, this draft soon showed that coding sequences correspond only to a small part of the human genome whereas repeat sequences account for at least 50%.

This project was soon followed by the sequencing of the mouse genome (Waterston et al., 2002), revealing that 37.5% of the murine genome is transposon-derived. Further comparative genomic studies showed that transposable elements have been more active in the mouse lineage, in comparison to the human lineage (Waterston et al., 2002) and that long interspersed nuclear element (LINE)-1 (L1) makes up the single largest fraction of interspersed repeat sequence in both genomes, representing 17% of the human (Lander et al., 2001) and almost 19% of the mouse genome (Waterston et al., 2002). Since the divergence rate of the human and mouse lineages is low enough that one can still align orthologous sequences (Makałowski & Boguski, 1998), the sequencing of the mouse genome allowed to complement studies of human genes with experimental manipulations of corresponding mouse genes. Furthermore, the mouse has been considered a unique model system for probing mammalian biology and human disease, with the advantages of accumulated genetic studies, practical techniques of random mutagenesis, and directed genome engineering by transgenic, knockin, and knockout techniques (Rossant & McKerlie, 2001).

1.2. LINE-1 elements in the mouse genome

Transposable elements (TEs) are DNA sequences with the ability to change their position in the genome, comprising the most abundant class of repeat sequences. TEs can be grouped into two classes, DNA transposons (class II) that show no recent evidence of transposition events, and retrotransposons (class I) (Bourque et al., 2018). Although retrotransposons represent a considerable part of the genome, only a small portion of these elements is still able to “copy and paste”, among which LINE-1 insertions seem to be the most recent (Naas et al., 1998) and active (Akagi et al., 2008) ones in the mouse genome. Unlike most TEs that are truncated or inverted, thus showing no evidence for recent transposition events, full L1 elements seem to have remained highly active in the mouse lineage, containing at least 3,000 full-length elements that are potentially capable of retrotransposition (Goodier et al., 2001).

1.2.1. Structure, transposition mechanism and evolution

L1 elements are non-long terminal repeat (LTR) retrotransposons, which are known to rely on a unique transposition mechanism called target site-primed reverse transcription (TPRT) (Luan et al., 1993) that will be further described in this section. Full-length mouse L1 elements are around 7-kb length, and they all have 5' untranslated region (UTR), two open reading frames (ORFs), and 3' poly(A) tail as depicted in **Figure 1A**. The 5' UTR functions as a promoter and is composed of monomers, which are tandemly repeated sequences of ~200 bp situated upstream of single-copy, nonmonomeric sequence (Goodier et al., 2001). The ORFs encode for two proteins that are necessary for retrotransposition (Moran et al., 1996): ORF2p is a 150 kDa protein with endonuclease (Feng et al., 1996) and reverse transcription (Mathias et al., 1991) activities required for TPRT, and ORF1p is a 40 kDa nucleic acid chaperone[1]. Active L1 elements should be able to perform a complete retrotransposition cycle (**Figure 1B**) in the following chronological order: transcription of L1 RNA in the nucleus, export into the cytoplasm, translation of ORF1 and ORF2, association of L1 RNA with ORF1p and ORF2p to form ribonucleoprotein (RNP) particles (Doucet et al., 2010), return to the nucleus, reverse transcription, and integration at a new genomic location via TPRT. The TPRT process is initiated by the element-encoded endonuclease that cleaves the target site in genomic DNA, generating a 3' OH that will act as the primer for reverse transcription; element RNA is used as a template for the reverse transcription and the result is simultaneous reverse transcription and joining of the first-strand complementary DNA (cDNA) with the genome (Luan et al., 1993). The mechanism used to complete cDNA synthesis and integrate both ends into the chromosome is still unknown.

Although the exact role of ORF1p is not yet clear, (Martin & Bushman, 2001) have hypothesized that this protein could be involved in the strand transfer reaction that is required for TPRT once the target site DNA is cleaved. Thus, according to the model presented in **Figure 2**, T-rich sequences located at the cleaved target site would anneal with the poly(A) tail of the L1 transcript, and then the 3' end of the cleaved target DNA would act as a primer for first-strand cDNA synthesis. At this step, (Martin & Bushman, 2001) proposed that the ORF1p strand transfer activity might be crucial for the primer transfer reaction. Furthermore, (Martin & Bushman, 2001) proposed that the nucleic acid chaperone activity might also be important for a second annealing event, in which the other strand of the cleaved target DNA would base pair with first-strand L1 cDNA, with the latter acting as the primer for second-strand cDNA synthesis.

[1] ORF1p has shown nucleic acid chaperone activity by accelerating the annealing of single-stranded complementary DNA molecules, allowing the system to achieve its equilibrium position more rapidly (Martin & Bushman, 2001).

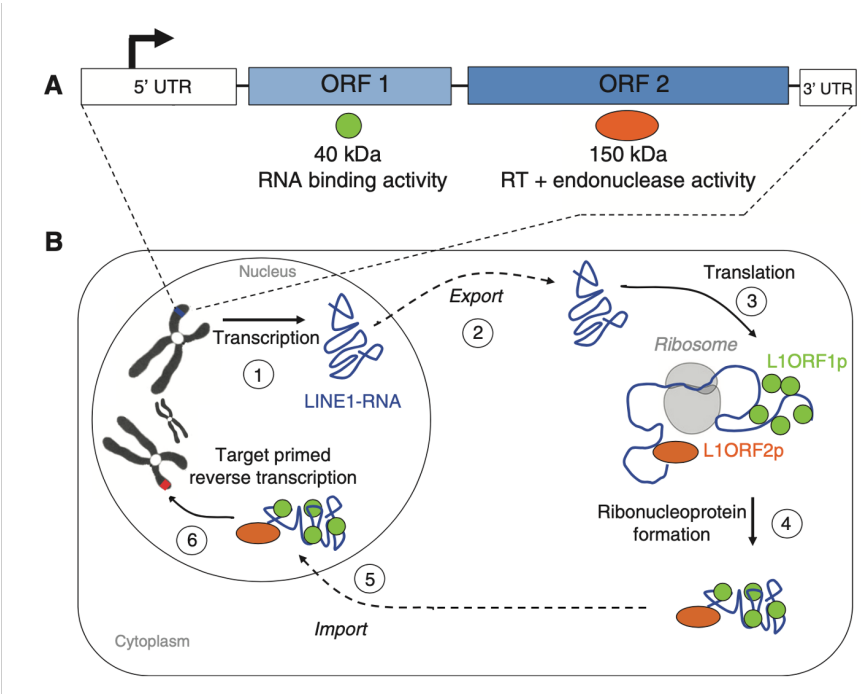

Figure 1: L1 structure and retrotransposition cycle. (A) Full-length L1 elements are composed of a 5' UTR, two ORFs (encoding for ORF1p and ORF2p), and a 3' UTR. **(B)** During a complete retrotransposition cycle, L1 RNA is transcribed in the nucleus (1) and later exported into the cytoplasm (2), where translation of ORF1 and ORF2 occurs (3). Then, ORF1p and ORF2p associate with L1 RNA to form ribonucleoprotein (RNP) particles (4) that are imported into the nucleus (5), so that the RNA can be integrated at a new genomic location via TPRT process (6). Adapted from (Bodak et al., 2014).

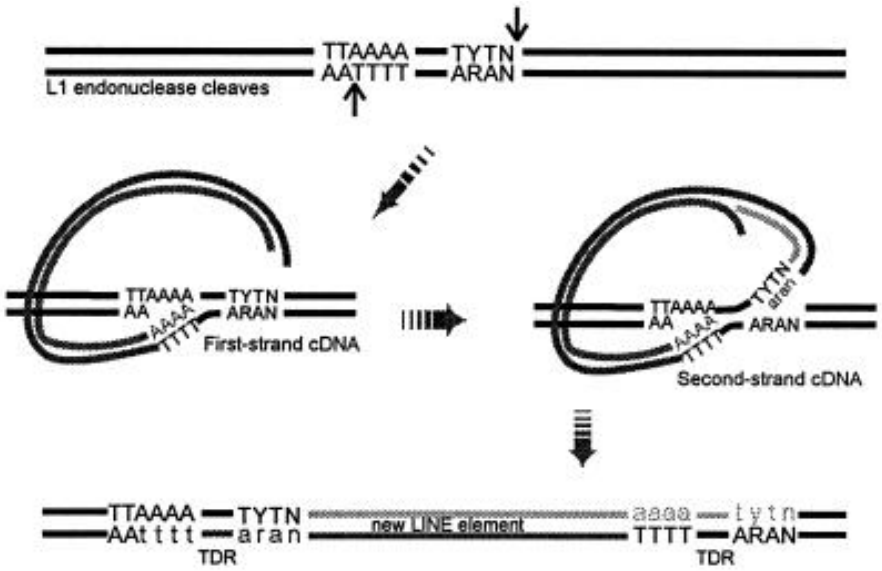

Figure 2: Proposed role for ORF1p strand-exchange activity during L1 retrotransposition. During the later stages of L1 retrotransposition, the ORF1p nucleic acid chaperone activity is proposed to facilitate the strand transfer required to anneal the DNA primer from the target site to the L1 RNA to prime reverse transcription. Furthermore, ORF1p may repeat these functions during priming and reverse transcription of the second strand. In both stages, the exchange of a DNA-DNA duplex in the target site for an RNA-DNA duplex occurs, resulting in the formation of a more stable hybrid. Newly synthesized DNA is represented with lowercase letters. Original nucleotides are represented with capital letters. TDR - target site direct repeat (direct repeat of the original sequence that is originated by the fill-in of the staggered, cleaved ends of chromosomal DNA). Adapted from (Martin & Bushman, 2001).

Early phylogenetic analyses propose that mouse L1 evolution has been dominated by a single lineage (Adey et al., 1994), and that the least conserved regions within these elements are the 5´UTR as monomers differ in number and sequence among L1 elements, and their promoter activity is proportional to the number of monomers (DeBerardinis & Kazazian Jr., 1999). Thus, three different types of mouse L1 elements were defined according to the type of non-homologous promoter sequence present at the 5' end: A, F, and V. These three promoter types were inserted into the mouse genome during different time periods (Adey et al., 1994), being the L1-A and L1-F elements the youngest and most abundant ones in the murine genome (Sookdeo et al., 2013). Although early L1 phylogenies usually show a cascade structure where a single family is active until a new family emerges and replaces the pre-existing one, it is now known that, in some instances, several lineages may co-exist until one becomes extinct; in fact, this is what is currently happening in the mouse genome, where two L1 lineages have been active at the same time: L1MdA and L1MdTf/Gf (Sookdeo et al., 2013). However, this is only possible because the two currently active lineages carry different, non-homologous 5' UTR sequences: L1MdA lineage carries A type promoter, and L1Tf/Gf lineage carries F type promoter (Sookdeo et al., 2013). It has recently been proposed by (Khan et al., 2006; Sookdeo et al., 2013) that the transcription of L1 elements with different 5' UTR sequences relies on different host-factors, which allows multiple families to co-exist as they are not using the same genomic 'niche'.

1.2.2. Function and regulation

The activation and retrotransposition of L1 elements can lead to the modification of the genome in a variety of ways, either beneficial or detrimental. Besides leading to gene inactivation through insertion into exons, L1 insertions into the genome are often associated with genome instability (N. Gilbert et al., 2002; Symer et al., 2002), and homologous recombination between dispersed L1 copies can lead to chromosomal rearrangements (Burwinkel & Kilimann, 1998). They can also modify cellular transcription acting as alternative promoters or enhancers, creating or disrupting polyadenylation sites, or splicing sites (Han et al., 2004). Although the misregulation of L1 elements can lead to diseases like cancer (Xiao-Jie et al., 2015) and neurodegeneration (Thomas et al., 2012), they have also been considered one of the strongest drivers of genome evolution and an alternative way of controlling gene expression.

Although L1 elements can be found interspersed almost anywhere in the mouse genome, some regions are more prone for their presence. In general, L1 elements tend to accumulate in (A+T)-rich and gene-poor regions of the genome (Boyle et al., 1990; Waterston et al., 2002). This has been attributed to a lower tolerance to keep L1 insertions in the proximity of genes, given their potential deleterious influence including deletions or insertions that could affect regulatory regions or the structure of the genes (Jachowicz & Torres-Padilla, 2016). In the cases where L1 insertions occur close to genes or even within them, the genes in question are often lowly expressed (Han et al., 2004), which suggests that L1 insertions could play a role in the inactivation of some genomic regions (Jachowicz & Torres-Padilla, 2016). Furthermore, the localization

pattern of L1 elements in the genome is also dependent on their age, with younger L1 elements being found in closer proximity to genes than older L1 insertions (Medstrand et al., 2002). In fact, recent studies revealed that the landscape of endogenous L1 elements differs significantly from that of new insertions, which appear to broadly target all regions of the genome; on the other hand, the distribution of preexisting L1 insertions seems to be dominated by post-integration selection[2] (Sultana et al., 2019).

It has been speculated that the specific association of genes and repeat families is so important that, during evolution, selective pressures have been imposed on different classes of repeats to accumulate in specific sets of genes. For example, it was recently reported by (Lu et al., 2020) that nucleolus-associated domains (NADs) and lamina-associated domains (LADs) contain significantly higher levels of L1 compared with random genomic regions and that, similarly to L1-enriched genes, NAD- and LAD-associated genes are also enriched in highly specialized functions. Furthermore, these domains are significantly depleted of RNA Polymerase II (Pol II) and active histone modifications but are enriched in silencing H3K9me2 and H4K20me3 marks that are associated with heterochromatin. Based on these results, (Lu et al., 2020) hypothesized that the predominant localization of L1 repeats and L1-associated genes in NADs and LADs likely contributes to the genome-wide silencing of L1-associated genes (**Figure 3**). Given that L1 RNA plays an essential role in regulating the nuclear organization and repression of L1-associated genes and considering the short half-life and extensive binding of L1 RNA to its DNA sequences, (Lu et al., 2020) proposed that L1 transcripts may act in chromatin neighborhoods of their transcription sites to anchor L1-rich genomic segments to the nuclear and nucleolar peripheries, in part through L1-interacting protein partners.

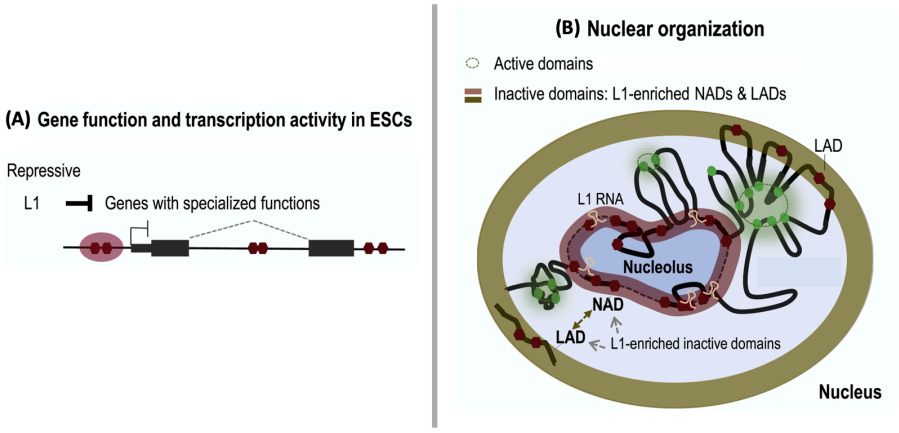

Figure 3: Repressive impact of L1 elements on specialized genes and nuclear organization of the silencing compartments. (A) L1 elements (red dots) are involved in the repression of genes with highly specialized functions. **(B)** L1 repeats (red dots) sequester L1-enriched genes in inactive nuclear domains for silencing; in particular, L1 RNA binds L1 DNA to facilitate its function in silencing L1-enriched genes that are associated with the inactive NADs and LADs in the peripheries of the nucleolus and nucleus. Adapted from (Lu et al., 2020).

[2] The predominant post-integrative process that models the distribution of endogenous L1s is purifying selection (Sultana et al., 2019), also known as natural selection, which is responsible for removing deleterious mutations that are produced in each generation.

Evolutionary-driven elimination of L1 elements by purifying selection is not the only mechanism that constrains their activity. In fact, the strongest and most significant mechanism of regulation is possibly related to their chromatin structure, whereby L1 acquires a silent chromatin configuration. The main silencing signature present on most types of TEs is DNA methylation (Meissner et al., 2008), which is related to transcriptional repression. Among these, mouse L1 elements contain particularly high levels of DNA methylation and hypermethylated canonical promoters (Meissner et al., 2008). This silencing signature is primarily established by the *de novo* DNA methyltransferases (DNMTs) DNMT3A and DNMT3B, that show a higher activity in the germ line and in early embryogenesis, and is maintained by DNMT1 in somatic tissues. Although being though to play an essential role in the silencing of L1 elements, DNA methylation is an epigenetic signature, which means that it is not always present on DNA and there are developmental time windows during which L1 elements are not methylated or not fully methylated. The most significant changes in the DNA methylation pattern occur when the genome undergoes global epigenetic reprogramming – during the formation of the primordial germ cells (PGCs) when almost a complete demethylation of the DNA occurs, and after fertilization when the newly formed zygote undergoes global DNA demethylation, reaching the lowest point in the early blastocyst (Mayer et al., 2000; Smith et al., 2012). Accordingly, the studies performed by (Smith et al., 2012) confirm that the most extreme methylation changes during the sperm to zygote transition are enriched for L1 elements, showing the largest and most consistent decrease in their methylation values. L1 activation and retrotransposition events were thought to occur predominantly in germ cells (Bourc'his & Bestor, 2004); however, recent *in vivo* studies demonstrate that retrotransposition occurring directly in germ cells is uncommon and that most of human and mouse L1 retrotransposition events occur during embryonic development (Kano et al., 2009). There is also evidence of L1 retrotransposition events from *in vitro* studies in embryonic stem cells (ESCs), suggesting once again that L1 activation and consequent retrotransposition may occur at early stages of the embryogenesis (Garcia-Perez et al., 2007). It is now evident that the time period following fertilization in mammals is a great window of opportunity for the activation of L1 elements, given the chromatin context in the embryo. This led to hypothesize if the L1 reactivation during this time window was just a result of the chromatin state or whether it could play a role in development. Indeed, recent results obtained by (Jachowicz et al., 2017) indicate that L1 activation is integral to the developmental program as L1 transcriptional activity was shown to promote global chromatin de- and recondensation after fertilization, thus affecting global chromatin accessibility. Furthermore, the latter results are independent of the coding nature of the transcript and of retrotransposition, which indicates that the functional role of L1 elements during early development is likely independent of its retrotransposition activity. Other recent study described L1 as a chromatin-associated RNA during early development (Percharde et al., 2018), thus ruling out the potential detrimental effects of L1 retrotransposition. Therefore, rather than being a vulnerability, the regulation of early development by TEs may allow both robustness, due to their repeated nature, and adaptability, due to their rapid evolution.

1.2.3. A possible role in development: X-chromosome inactivation

One of the most appealing hypotheses for the contribution of L1 elements in early development is that they could play a role in the inactivation of the X chromosome in females. Several observations led to this hypothesis, starting with early hybridization experiments performed in the genome of *Mus musculus domesticus* (Boyle et al., 1990), which revealed that "the X chromosome is exceptionally rich in LINE sequences". This observation was later confirmed by the sequencing of the mouse genome (Waterston et al., 2002), reporting that LINEs accumulate preferentially on sex chromosomes, with the mouse X chromosome containing around twice the density of L1 elements as the mouse autosomes. This bias was noticed to be even stronger on human sex chromosomes (Lander et al., 2001; Ross et al., 2005), and in both genomes this enrichment in L1 elements is still very significant after accounting for the higher (A+T) content of the sex chromosomes (Waterston et al., 2002). Furthermore, regional analysis of the X chromosome revealed that the most significant clustering of L1 elements was in Xq13-Xq21, which corresponds to the center of X inactivation, and that genomic segments containing genes that escape inactivation are significantly reduced in L1 content compared with X chromosome segments containing genes subject to X inactivation (Bailey et al., 2000). All the above-mentioned observations provided support for the association between X inactivation and L1 content in mammals.

X-chromosome inactivation (XCI) consists in the silencing of one X chromosome in females to compensate the differences in X-linked gene dosage between XX females and XY males (Lyon, 1961, 1962). Thus, XCI is a chromosome-wide silencing process that takes place early in development to ensure equivalent levels of expression from genes on the X chromosome between males and females. Although it has been reported that different eutherian mammalian species use different strategies to initiate XCI (Okamoto et al., 2011), one common feature seems to be the involvement of long noncoding RNA (lncRNA) as a central player of the XCI process. In eutherian mammals, XCI is mediated by a lncRNA called Xist (X-inactive-specific-transcript) that is expressed from the inactive X (Xi) chromosome in females. Among eutherians, the mouse has been a preferred model to study XCI due to the use of murine embryonic stem cells (mESCs), which are pluripotent cells derived from the inner cell mass (ICM) of the blastocyst. In fact, mESCs represent a useful model since in undifferentiated mESCs (as in cells from the ICM) both female X chromosomes are active, but during differentiation (as in the embryo development), the silencing of one X chromosome is triggered. Thus, random XCI can be simulated during *in vitro* differentiation of mESCs, allowing the successive steps of XCI to be followed (Chaumeil et al., 2004; Wutz & Jaenisch, 2000). In the mouse, the process of XCI is initiated by the X-inactivation center (Xic), that includes the Xist gene. The Xist transcript coats the future Xi chromosome in *cis* and triggers gene silencing. However, it has been shown that the low levels of Xist RNA in embryos and in differentiating female ES cells are incompatible with models that require Xist RNA to cover the entire Xi chromosome (Buzin et al., 1994). Thus, chromosomal elements present on the X chromosome may exist to facilitate the propagation of the inactive state along the entire Xi;

given the significant enrichment of L1 elements on the X chromosome and all the observations that were previously mentioned in this section, L1 elements were thought to play a role (Lyon, 1998), but the underlying mechanism was still largely unknown.

Only more recently, a potential mechanism of XCI based on the silencing activity of repeat elements was proposed by (Chaumeil et al., 2006). Following Xist RNA accumulation on the future Xi chromosome, Xist RNA chromosome coating leads to the rapid exclusion of RNA Pol II and associated transcription factors, thus forming a nuclear silent compartment. Initially, this silent compartment contains more repeat-rich regions, and, on the other hand, X-linked genes tend to be located outside or at the periphery of this Xist RNA domain when still active. Later, a process of gene silencing is initiated and a significant shift in location of the X-linked genes that are being silenced, from the outside toward the inside of the Xist RNA domain, is observed. This shift of genes to a more internal location within the RNA Pol II-depleted Xist RNA domain seems to be a repeat-mediated gene silencing process, and it does not appear to be due to an increase in the size of the Xist RNA domain, nor to a global condensation of the X chromosome. Thus, according to (Chaumeil et al., 2006), repeat regions are able to induce silencing of X-linked genes that then become relocated within the silent nuclear compartment delineated by Xist RNA, whereas genes that escaped from XCI remain external.

Later, the particular implication of L1 elements in this silencing process was investigated by (Chow et al., 2010). Genes in L1-poor regions were confirmed to remain at distances far from Xist RNA compartment and resist Xist-mediated silencing; on the other hand, genes in L1-rich regions show much more efficient silencing and are drawn into the repetitive Xist compartment. Thus, the high L1 density on the X chromosome likely provides a helpful environment for nucleating the heterochromatic Xist compartment, as well as for the subsequent heterochromatization of genes. Furthermore, a subset of L1 elements was shown to be specifically expressed from the Xi chromosome during XCI: L1-Tf and L1-Gf elements appear to be transcribed from their own promoters during early ES cell differentiation, but become downregulated as differentiation proceeds, except on the Xi chromosome. Indeed, Xist-mediated formation of facultative heterochromatin was shown to promote transcription of such L1 elements; given that these young and active L1 elements often lie within escape-prone regions of the X chromosome, but close to genes that are subject to XCI, their role was hypothesized to relate with the local propagation of XCI into regions that would otherwise be prone to escape. Thus, it was proposed by (Chow et al., 2010) that L1 elements may facilitate XCI at different levels, with silent LINEs participating in assembly of a heterochromatic nuclear compartment induced by Xist, and active LINEs facilitating silencing of certain regions of the Xi (**Figure 4**).

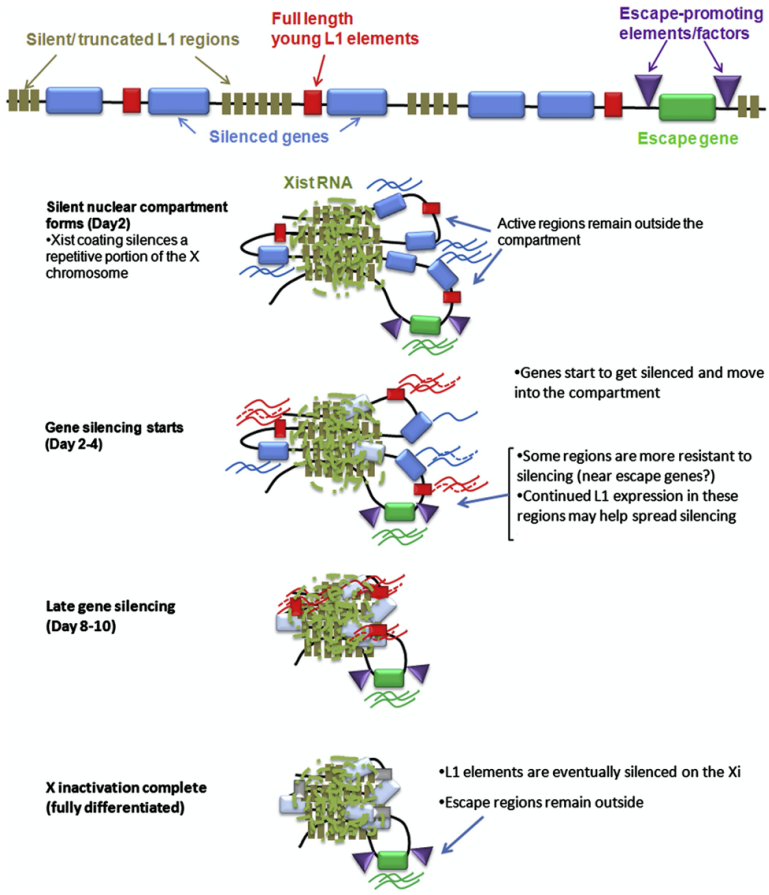

Figure 4: Model for the dual role of L1 elements during XCI. Silent L1 elements may facilitate the assembly of heterochromatic nuclear compartment early in XCI, whilst transcriptionally active L1 elements may facilitate the spread of silencing into certain regions on the X chromosome – in particular those regions that are more resistant to silencing, located around escape domains. Adapted from (Chow et al., 2010).

1.3. Functional approaches to address the role of TEs

Despite their abundance and evidence suggesting the importance of TEs in rewiring gene regulatory networks, these elements remain frequently "ignored" in genomic studies in part because of their repetitive nature, which makes them challenging to map onto a reference genome. Thus, the study of TEs, their evolution and behavior and ultimately their impact on the host has lagged behind, given that it often requires the use of specialized tools. As genomes can harbor thousands of copies of very similar TE sequences, uniqueness or, alternatively, repetitiveness of substrings within these regions need to be taken into consideration during both experimental design and analysis; For example, short DNA oligos targeting a specific TE in the genome for polymerase chain reaction (PCR) or CRISPR-Cas9 (with CRISPR standing for Clustered Regularly Interspaced Short Palindromic Repeats, and Cas9 for a CRISPR-associated (Cas) protein containing two nuclease domains that cleave DNA) must be carefully designed and validated to ensure that they are truly specific and target unique regions of the genome. In some cases, it can even be desirable to target many elements simultaneously or an entire TE family. Nevertheless, as technology improves many of the hurdles faced by earlier studies will be

progressively removed; Recent technological advances in bioinformatics (reviewed by (Goerner-Potvin & Bourque, 2018)) including the creation of comprehensive databases of annotated repetitive elements, advances in DNA sequencing, and clearer knowledge of the genome from a great variety of organisms, have facilitated the study of repetitive elements. Furthermore, targeted genome engineering methods based on CRISPR-Cas9 nuclease may provide clean systems for studies of retrotransposon impact, for example, through the construction of cells or organisms completely lacking active retrotransposons to determine their role in genome stability and many other relevant biological processes.

However, despite these recent technological advances, most current studies are either correlative or focus on the analysis of individual insertions, rather than on systematically perturbing specific TE classes. To the best of my knowledge, only five genome-wide studies regarding the function and regulation of TEs in the genome were published until the present date. (Jachowicz et al., 2017) used transcription-activator-like effectors (TALEs) that bound the three primary locations where L1 transcription initiates – the monomers and the regions upstream of Orf1 and Orf2 – for functional perturbation of mouse L1 elements; in this study, the most conserved regions of the L1-Tf family were used to address whether specific retrotransposons families play a direct role in chromatin organization and developmental progression. (Liu et al., 2017) performed a genome-wide CRISPR-Cas9 screen in human chronic myeloid leukaemia K562 cells to systematically identify genes regulating L1 retrotransposition; here, an L1-G418^R retrotransposition reporter (in which G418^R indicates resistance to the antibiotic G418, also known as geneticin) under the control of a doxycycline (dox)-responsive promoter was used. (Fuentes et al., 2018) coupled their recently developed method for guide RNA (gRNA) multiplexing called CARGO (Chimeric Array of gRNA Oligos) (Gu et al., 2018) with nuclease dead Cas9 (dCas9) fused to an activation or repression domain (Chavez et al., 2015; L. A. Gilbert et al., 2013) to facilitate transcriptional induction or silencing of HERVK LTR5HS elements[3] en masse and test the function of LTR5HS elements as early embryonic enhancers. (Percharde et al., 2018) tested the hypothesis that L1 elements may play essential functions to orchestrate developmental progression during pre-implantation and for the self-renewal of mouse embryonic stem cells, by applying RNA interference (RNAi) techniques. Finally, more recently, a dCas9 fused with the transcriptional activator VP64 (dCas9-VP64) was transiently expressed in 293T cells together with single guide RNAs (sgRNAs) targeting the L1 antisense promoter (ASP), that drives the production of L1-gene chimeric RNAs, to investigate the biological impact of L1 ASP activity (Honda et al., 2020).

As described above, very few genome-wide studies exist to better understand the impact of TE activity in the genome. This knowledge gap is associated with technical challenges, given that a single gRNA is often insufficient for robust gene activation/silencing by CRISPRa/CRISPRi. TE families are often present in hundreds or thousands of copies, which are highly repetitive, but sufficiently sequence-divergent to prevent their recognition by a single short-sequence-dependent factor, such as a zinc finger protein or a CRISPR guide RNA (Fuentes et al., 2018).

[3] HERVK LTR5HS elements are an ape-specific class of endogenous retroviruses (ERVs).

Thus, to overcome these limitations and develop a strategy for systematic interrogation of TE function, one needs to use multiple (at least two) gRNAs to target (a considerable part of) a given subfamily of TEs.

1.3.1. Aim of this study

As stated before, there is a clear knowledge gap regarding systematic studies on TE function and consequent impact in genome stability. Thus, the main objective of this project is to establish functional cellular models that could be useful to address the impact of young LINE-1 elements genome-wide, during differentiation and in differentiated cells. Since functional perturbation systems of L1 expression in undifferentiated mouse embryonic stem cells had already been established in the lab, the implementation of such systems during differentiation and in differentiated cells would allow us to follow the cellular developmental stages and, thus, study the impact of L1 elements during this period.

2. MATERIALS AND METHODS

2.1. Cell culture

All mouse embryonic stem (ES) cell lines used in this project were previously established by Anne-Valerie Gendrel; the mouse neural progenitor cell (NPC) line was previously established by Lenka Misikova, a former MSc student in the group; the NIH-3T3 cell line (mouse embryonic fibroblasts) was kindly provided by the group of Edgar Gomes (iMM-JLA).

2.1.1.Proliferation of mouse ES cells in an undifferentiated state

ES cells were grown as a monolayer on 6-well tissue culture plates coated with 0.1% gelatin (Sigma Aldrich). To maintain them in an undifferentiated state, ES cells were grown in Dulbecco's Modified Eagle's Medium (DMEM) with sodium pyruvate (Gibco by Life Technologies), supplemented with 15% (v/v) heat-inactivated ES-grade fetal bovine serum (FBS) (Gibco by Life Technologies), 1% (v/v) 200 mM L-glutamine (Gibco by Life Technologies), 0.2% (v/v) 50 mM β-mercaptoethanol (Sigma Aldrich) and 0.01% (v/v) leukemia inhibitory factor (LIF) (Merck). Further on, the previous complete mixture will be referred as ES medium. Cells were maintained at 37ºC in a humid atmosphere containing 5% CO_2. All the handling procedures were performed in a laminar flow hood.

When a confluency of 80% was achieved, part of the undifferentiated ES cells was transferred to a new 6-well plate with fresh ES medium: wasted medium was removed, cells were gently washed with phosphate-saline buffer (PBS) 1X and detached by treatment with trypsin EDTA 0.05% (Gibco by Life Technologies) for 5 minutes at 37ºC. After trypsinization, the cells were resuspended in fresh ES medium by pipetting up and down several times to avoid the

formation of long filaments of cells. Then, part of the cell suspension was transferred to a new 0.1% gelatin-coated 6-well plate containing fresh ES medium, according to the desired dilution. Undifferentiated ES cells were passaged between 1:8 and 1:10 every two days.

2.1.2.Random differentiation of mouse ES cells by LIF withdrawal

ES cells were randomly differentiated through a process of LIF withdrawal. For this purpose, cells were grown as a monolayer on 6-well tissue culture plates containing Dulbecco's Modified Eagle's Medium (DMEM) with sodium pyruvate (Gibco by Life Technologies), supplemented with 10% (v/v) fetal bovine serum (FBS) (Gibco by Life Technologies), 1% (v/v) 200 mM L-glutamine (Gibco by Life Technologies) and 0.2% (v/v) 50 mM β-mercaptoethanol (Sigma Aldrich). Once the complete medium solution was prepared, it was stored at 4ºC. Further on, the previous complete mixture will be referred as differentiation medium. Cells were maintained at 37ºC in a humid atmosphere containing 5% CO_2. All the handling procedures were performed in a laminar flow hood.

At day 0, undifferentiated ES cells were passaged according to the protocol described in **section 2.1.1**, with the following exceptions: 0.1% gelatin coating was not necessary for differentiation purposes; instead of ES medium, ES cells were resuspended and plated in differentiation medium after being counted according to the protocol described in **section 2.1.5**. For each differentiation experiment, 5×10^5 cells/well were plated on a 6-well tissue culture plate containing differentiation medium. From day 1, wasted differentiation medium was removed and fresh differentiation medium was added every day. Two differentiation processes were performed in parallel, a 2-day and a 3-day differentiation processes.

2.1.3.Proliferation of established NPCs

Previously established NPCs were grown as a monolayer on 6-well tissue culture plates coated with 0.1% gelatin (Sigma Aldrich). These cells were maintained in N2B27 medium supplemented with murine epidermal growth factor (EGF) (Peprotech) and human fibroblast growth factor-basic (FGF) (Peprotech), at a final concentration of 10 ng/ml each. Further on, the previous complete mixture will be referred as NPC medium. N2B27 medium consists in a mixture 1:1 of Dulbecco's Modified Eagle's Medium (DMEM)/F12 (Gibco by Life Technologies) and Neurobasal Medium (Gibco by Life Technologies), supplemented with 1% (v/v) L-glutamine 200 mM (Gibco by Life Technologies), 1% (v/v) B27 (Gibco by Life Technologies), 0.5% (v/v) N2 (Millipore) and 0.2% (v/v) β-mercaptoethanol 50 mM (Sigma Aldrich); once N2B27 medium was prepared, it was filtered using a Millipore stericup system through a 0.22 μm filter. NPCs were maintained at 37ºC in a humid atmosphere containing 5% CO_2. All the handling procedures were performed in a laminar flow hood.

When a confluency of 80% was achieved, part of the cells was transferred to a new 6-well plate with fresh NPC medium: wasted medium was removed and cells were detached by

treatment with accutase (Millipore) for 1-3 minutes at 37ºC. After incubation, fresh NPC medium was added to dilute out the accutase and the cell suspension was centrifuged for 3 minutes at 1000 rpm. Then, the supernatant was removed, the cell pellet was resuspended in fresh NPC medium and part of the cell suspension was transferred to a new 0.1% gelatin-coated 6-well plate containing fresh NPC medium, according to the desired dilution. NPCs were passaged between 1:2 and 1:4 every 2-3 days.

2.1.4.Proliferation of NIH-3T3 cells

NIH-3T3 cells were grown as a monolayer in standard tissue culture petri dishes with Dulbecco's Modified Eagle's Medium (DMEM) (Gibco by Life Technologies) supplemented with 10% (v/v) fetal bovine serum (FBS) (Gibco by Life Technologies), 1% (v/v) 200 mM L-glutamine (Gibco by Life Technologies) and 1% (v/v) penicillin-streptomycin antibiotics solution (Gibco by Life Technologies). Once the complete medium solution was prepared, it was stored at 4ºC. Further on, the previous complete mixture will be referred as 3T3 medium. Cells were maintained at 37ºC in a humid atmosphere containing 5% CO_2. All the handling procedures were performed in a laminar flow hood.

When a confluency of 80-90% was achieved, part of the cells was transferred to a new petri dish with fresh 3T3 medium: wasted medium was removed, cells were gently washed with PBS 1X, and detached by treatment with trypsin EDTA 0.05% (Gibco by Life Technologies) for 3-5 minutes at 37ºC. After trypsinization, the cells were resuspended in fresh 3T3 medium and part of the cell suspension was transferred to a new petri dish containing fresh 3T3 medium, according to the desired dilution. NIH-3T3 cells were passaged between 1:8 and 1:10 every two days.

2.1.5.Cell counting

To count the cells, a hemocytometer composed by 4 quarters was used. A mixture 1:1 of cell suspension and trypan blue (Sigma Aldrich) was loaded into the counting chamber and the viable cells present on each quarter were counted under the microscope. Trypan blue enters damaged (cells with a damaged cell membrane) or dead cells, staining them blue, which enables us to distinguish viable (bright) from non-viable (blue) cells by a dye exclusion method. The viable cell concentration (N) is determined by **Equation 1**, where n_v corresponds to the total number of viable cells in the 4 quarters of the hemocytometer. To consider the counting trustworthy, the number of viable cells per quarter should be between 30 and 300. The volume of each quarter (0.1 mm^3) and the 1:1 dilution that was made with trypan blue must also be considered when determining the viable cell concentration of the original sample.

$$N\ (cells/ml) = \frac{n_v}{4} \times 2 \times 10^4 \quad (1)$$

After determining the viable cell concentration of the original sample (N), the volume of cell suspension that contains the required number of cells is determined by **Equation 2**.

$$Volume\ of\ cell\ suspension\ (ml) = \frac{Required\ number\ of\ cells}{N\ (cells/ml)} \quad (2)$$

The final mixture for plating the desired number of cells can then be prepared with the determined volume of cell suspension and the appropriate medium to make up the total working volume of the tissue culture plate.

2.2. Transfection

Transfection of plasmids was performed with the Lipofectamine 3000 Transfection Reagent (Invitrogen). To perform a transfection experiment, two solutions were prepared: solution (a), composed by 125 μl of Opti-MEM Reduced Serum Medium (Opti-MEM) (Gibco by Life Technologies), plasmid DNA and 2 μl of P3000 reagent (Invitrogen) per 1 μg of plasmid DNA; and solution (b), containing 125 μl of Opti-MEM and 1.5 μl of Lipofectamine 3000 per 1 μg of plasmid DNA. Both solutions were incubation for 5 minutes at room temperature (RT). Then, solution (a) was gently added to solution (b) and the final mix was incubated for 20 minutes at room temperature (RT). Solution (a)+(b) was added to each well drop by drop and gently distributed. Transfected cells were incubated at 37ºC in a humid atmosphere containing 5% CO_2. All the handling procedures were performed in a laminar flow hood.

Cells were plated on standard tissue culture 6-well plates on day 0. Transfection experiments (forward transfection) were performed on day 1, when the cells were 70-80% confluent. A total amount of 2.5 μg of plasmid DNA was transfected in each experiment. All transfected plasmids are described in **section 2.3**.

2.3. Plasmids and genetic constructs

To perform the transfection experiments mentioned in **section 2.2**, a number of genetic constructs in the form of recombinant plasmids, already available in the host lab, were used. The plasmids pX330-sgTIGRE-Cas9 (Addgene, #92144) and pEN366-dCas9-VPR (a kind gift from Michel Wassef, Institut Curie) were used for CRISPR-mediated homologous recombination: pX330-sgTIGRE-Cas9 expresses a single guide RNA (sgRNA) for the TIGRE safe harbor locus, under the control of a human U6 (hU6) promoter, and a nuclease Cas9 protein, under the control of a constitutive hybrid human cytomegalovirus enhancer/chicken β-actin promoter (CBh); pEN366-dCas9-VPR expresses a nuclease-dead Cas9 (dCas9) protein fused to a VPR activation complex (composed by three transcriptional activation domains – VP64, p65 and Rta), under the control of a Tet-On 3G tetracycline-inducible expression system. The plasmid pLK01-sgTfmono2-3 expresses two sgRNAs for the 5'UTR monomers of L1Md-Tf elements, respectively under the control of hU6 and mouse U6 (mU6) promoters.

2.4. Bacterial transformation

The plasmids mentioned in **section 2.3** were amplified in competent *Escherichia coli* DH5α bacteria, following a heat shock protocol. Competent bacteria were thawed on ice and plasmid DNA was mixed. The obtained mixture was incubated on ice for 20 min, followed by heat shock at 42ºC for 1 min and then, incubation on ice for 10 min. At this stage, an outgrowth step was performed at 37ºC, 220 rpm for 60 min to allow competent bacteria to produce the antibiotic resistance proteins that are encoded by the transformed plasmid. Afterwards, bacteria were plated on LB-agar medium supplemented with Ampicillin and incubated overnight at 37ºC. Next day, isolated colonies were picked from the LB-agar plate and 5 ml of LB medium supplemented with Ampicillin were inoculated. Bacteria were left to grow overnight at 37ºC, 220 rpm. Afterwards, the bacterial suspension was centrifuged at 3000 rpm, 4ºC for 8 min and the amplified plasmid DNA was separated and purified using the NZYMiniprep kit (Nzytech). Briefly, bacterial cells were lysed to allow the release of plasmid DNA, proteins were denatured, lipids were solubilized, and the RNA content was digested. Then, a centrifugation step was performed to separate the plasmid DNA from the cellular debris and the supernatant was loaded onto a silica column, where the plasmid undergoes adsorption. After several washing steps, the pure plasmid DNA was eluted from the column. The concentration and purity of the plasmid DNA were measured on the NanoDrop 2000 spectrophotometer device.

2.5. Genotyping

2.5.1.Quick cell lysis for genotyping

To genotype cell clones from a tissue culture 96-well plate, we used a quick cell lysis protocol to release DNA from cells. Briefly, medium was aspirated from each well, cells were washed twice with PBS 1X and 50 μl of lysis buffer were added to each well. Lysis buffer was composed of 100 mM tris hydrochloride (Tris-HCl) pH 8.0, 200 mM sodium chloride (NaCl), 0.5% (v/v) Tween20 and freshly added Proteinase K (Nzytech) to a final concentration of 0.2 mg/ml. The plate was then sealed and incubated overnight at 56ºC, in a humid chamber, to lyse the cells. Next day, the plate was centrifuged at 2000 rpm for a few seconds and Proteinase K was inactivated at 95ºC for 15 min. Finally, the lysates were diluted in 50 μl of DNase/RNase free water and strongly vortexed to shear the DNA. Typically, 1-2 μl of lysate were used for a 25 μl polymerase chain reaction (PCR), as described in **section 2.5.3**.

2.5.2.Phenol/chloroform DNA extraction

When higher quality DNA samples were required, a phenol/chloroform DNA purification protocol was followed. First, cell pellet was resuspended in 500 μl of lysis buffer, composed by 50 mM Tris-HCl pH 8.0, 10 mM NaCl, 10 mM ethylenediamine tetraacetic acid (EDTA) pH 8.0,

0.5% (v/v) sodium dodecyl sulfate (SDS) and freshly added Proteinase K (Nzytech) to a final concentration of 1.0 mg/ml. Samples were incubated for at least 2 hours at 50ºC to lyse the cells. Then, an equal volume (500 μl) of phenol-chloroform was added to the lysate and the mix was strongly vortexed to shear the DNA. The aqueous and organic phases were separated by centrifugation at maximum speed (15,000 rpm), 4ºC for 5 min and then, the aqueous phase (containing the genomic DNA) was gently transferred to a fresh DNase/RNase free tube. The previous steps were repeated, replacing the phenol-chloroform with chloroform only. The aqueous phase was, once again, transferred to a new tube. To precipitate the DNA that was in solution, 1/25 volumes of 5 M NaCl and 2.5 volumes of 100% ethanol were added to the aqueous phase and mixed gently. The DNA pellet was then collected by centrifugation at maximum speed, 4ºC for 30 min, followed by a washing step with 70% ethanol and a final centrifugation step at maximum speed, 4ºC for 10 min. Finally, the supernatant was removed, the DNA pellet was left to air dry for 30 min at RT and then, resuspended in DNase/RNase free water. The concentration and purity of the DNA samples were measured on the NanoDrop 2000 spectrophotometer device. Typically, 50 ng of pure DNA were used for a 25 μl PCR, as described in **section 2.5.3**. All the steps that required the handling of phenol-chloroform, chloroform and 100% ethanol were performed under the fume hood.

2.5.3.Polymerase chain reaction (PCR)

Polymerase chain reaction (PCR) was used for genotyping purposes. Each reaction was set up with 50 ng of pure DNA, 0.5 μl of deoxynucleotide triphosphates (dNTPs) mix at 10 mM each (Thermo Fisher Scientific), 1 μl of the forward primer at 10 μM, 1 μl of the reverse primer at 10 μM, 0.2 μl of HotStarTaq DNA polymerase (Qiagen), 2.5 μl of the corresponding 10X PCR buffer with magnesium chloride ($MgCl_2$) (Qiagen) and DNase/RNase free water to a final reaction volume of 25 μl. The amplification program recommended by the Taq DNA polymerase supplier was followed.

2.6. Reverse transcription quantitative PCR

2.6.1.Total RNA extraction

Total RNA isolation was performed using TRIzol reagent (Life Technologies). First, cell pellet was resuspended in 500 μl of TRIzol reagent and incubated at RT for 5 min with rotation, to homogenize the mixture. Then, 100 μl of chloroform were added, the samples were mixed well by shaking and then, incubated at RT for 3 min. The aqueous and organic phases were separated by centrifugation at 12,000 rcf, 4ºC for 15 min. The aqueous phase (containing the RNA) was gently transferred to a fresh DNase/RNase free tube and 250 μl of isopropanol were added to precipitate the RNA. The samples were incubated at RT for 10 min and then, the RNA pellet was collected by centrifugation at 12,000 rcf, 4ºC for 10 min. After removing the supernatant, the RNA

pellet was washed with 75% ethanol and centrifuged at 7500 rcf, 4ºC for 10 min. Afterwards, the supernatant was discarded and once the RNA pellet was dried, it was resuspended in DNase/RNase free water. At this point, the concentration and purity of the RNA samples were measured on the NanoDrop 2000 spectrophotometer device. To remove any residual DNA that may be in the RNA samples, a treatment with DNase I was performed. For this purpose, 5 μg of RNA were treated with 5 μl of DNase I recombinant enzyme (Nzytech), along with 5 μl of the corresponding 10X buffer DNase I (Nzytech) and 1 μl of Glycoblue (Invitrogen), in a final reaction volume of 50 μl. After 2 hours of incubation at 37ºC, 150 μl of 100% ethanol and 5 μl of 3M sodium acetate were added, mixed well, and the samples were incubated at -20ºC overnight to allow the RNA to precipitate. Next day, samples were centrifuged at maximum speed (15,000 rpm), 4ºC for 60 min, the supernatant was discarded and the blue RNA pellet was gently washed with 75% ethanol and centrifuged at maximum speed, 4ºC for 10 min. Finally, the RNA pellet was left to air dry at RT and then, resuspended in DNase/RNase free water. The concentration and purity of the final RNA samples were measured on the NanoDrop 2000 spectrophotometer device. All steps (excluding the ones where RT is specifically mentioned) were performed on ice. Additionally, all the steps that required the handling of TRIzol reagent, chloroform, isopropanol and 100% ethanol were performed under the fume hood. The obtained RNA samples were diluted with DNase/RNase free water to a final concentration of 200 ng/μl to decrease the variability among samples in downstream applications.

2.6.2.Reverse transcription (RT)

To convert the previously extracted RNA into complementary DNA (cDNA), NZY First-Strand cDNA Synthesis Kit (Nzytech) was used. On ice, 800 ng of RNA were mixed with 10 μl of NZYRT 2X Master Mix (oligo(dT)$_{18}$, random hexamers, $MgCl_2$ and dNTPs), 2 μl of NZYRT Enzyme Mix (NZY Reverse Transcriptase and NZY Ribonuclease Inhibitor) and RNase-free water up to a final volume of 20 μl. The previous mixture was incubated at 25ºC for 10 min, followed by 50ºC for 30 min. The reaction was then inactivated by heating at 85ºC for 5ºC and samples were chilled on ice. At this point, 1 μl of RNase H (Nzytech) was added to degrade the RNA that was bond to cDNA and samples were incubated at 37ºC for 20 min. The addition of RNase H is mainly recommended when using cDNA in PCR amplification, since it will increase the sensitivity of the downstream PCR step. The obtained cDNA samples were diluted 1:5 with DNase/RNase free water, i.e., 80 μl of water were added to 20 μl of cDNA, before being used in downstream applications.

2.6.3.Quantitative PCR (qPCR)

In this project, quantitative PCR (qPCR) was used in association with RT (mentioned in **section 2.6.2**). Thus, reverse transcription quantitative polymerase chain reaction (RT-qPCR) was used for gene expression analysis, allowing for the quantification of a specific transcript in a

sample. For that purpose, 2 μl of diluted sample (cDNA or negative control) were used in each reaction, as template for the quantitative PCR. Negative control samples consisted in 800 ng of RNA diluted in DNase/RNase free water to a final volume of 100 μl. Additionally, 5 μl of PowerUp SYBR Green Master Mix 2X (Thermo Fisher Scientific), 0.25 μl of forward/reverse primers mix (at 10 μM each) and 2.75 μl of DNase/RNase free water were added to the template, making a final volume of 10 μl per reaction. Each reaction was carried out in triplicates, except for the negative control samples, for which only one reaction was carried out per primers pair. The qPCR reactions were carried out on 384-well reaction plates and the amplification data was acquired on ViiA 7 Real-Time PCR System. The standard cycling program (including a melt curve stage) recommended by the PowerUp SYBR Green suppliers was followed.

A relative quantification assay following the $2^{-\Delta\Delta C_t}$ method was performed to determine the changes in gene expression of each sample relative to a reference sample. In this project, the cycle number in which the fluorescent signal generated by the qPCR reaction exceeded the background threshold (threshold cycle, Ct) was averaged for each triplicate reaction and was normalized against the averaged Ct of housekeeping genes (internal controls) with different expression levels – Gapdh, Beta-actin and Rrm2. Some technical replicates were considered outliers whenever the standard deviation was higher than 0.25.

2.7. Western blot

2.7.1.Total protein extraction

To extract the total protein content, cell pellets were resuspended in 100-300 μl (depending on the size of the cell pellet) of cold RIPA buffer supplemented with Pierce Protease and Phosphatase Inhibitor Mini Tablet (Thermo Fisher Scientific) (referred as RIPA+PI) to lyse the cells and solubilize proteins while avoiding protein degradation and interference with the biological activity of proteins. RIPA buffer was composed of 150 mM NaCl, 1% (v/v) NP-40, 0.5% (m/v) sodium deoxycholate (DOC), 0.1% (v/v) SDS, 25 mM Tris-HCl and 1 mM EDTA. Samples were then passed 5x through a 21G needle and sonicated at 45 kHz, 4x 1 min with 30 sec on ice between each sonication, on an ultrasonic bath USC 300TH to shear the DNA. Afterwards, a pellet containing cell debris and other impurities was collected by centrifugation at 12,000 rpm, 4ºC for 20 min and the supernatant (containing the total protein extract) was transferred to a new tube. All steps were performed on ice (or on water with ice, in the case of the sonication step).

2.7.2.Bradford assay

To determine the total protein concentration of the extracts collected in **section 2.7.1**, Bradford Assay was performed with NZYBradford Reagent (Nzytech). For that purpose, standard solutions (i.e., solutions with a known concentration of total proteins) were prepared with bovine serum albumin (BSA) to the final concentrations of 0.05, 0.1, 0.2, 0.3, 0.4, 0.5 and 0.6 mg/ml.

Total protein extracts were diluted in RIPA+PI to fit the range 0.05-0.6 mg/ml. A flat transparent 96-well plate was prepared with 10 μl of standard solutions, diluted protein extracts or blank solution (RIPA+PI) in triplicates. Then, 200 μl of NZYBradford Reagent was added to each well, the plate was incubated with linear agitation for 2 min and absorbance was measured at 595 nm using Microplate Reader TECAN Infinite M200.

Total protein concentration of each sample was determined by a standard curve equation and the appropriate dilution factor.

2.7.3.Western blot

After determining the total protein concentration as described in **section 2.7.2**, 10 μg of proteins (from **section 2.7.1**) were mixed with 2X Sample Buffer and RIPA+PI until a final volume of 10 μl. The prepared samples were boiled at 100ºC for 5 min and then, centrifuged at 15,000 rpm, RT for 2 min. An extract containing linear and negatively charged proteins was obtained at this point. Equal amounts of protein extract and NZYColour Protein Marker (Nzytech) were loaded on 10% polyacrylamide gel, which was resolved by SDS-PAGE at the constant current of 30 mA, applied by a Mini-PROTEANT Tetra Electrophoresis System (Biorad). Next, proteins were transferred to a nitrocellulose membrane using the dry transfer iBlot System (Thermo Fisher Scientific), through a 7 min standard program at 20 V. The transfer process was confirmed by staining the nitrocellulose membrane with Ponceau Red and rinsing it with water to check the appearance of total protein bands. At this point, the membrane was cut by the desired molecular weights, if several primary antibodies were aimed to be used simultaneously. The membranes were washed twice with tap water and PBS 1X containing 0.05% (v/v) Tween20 (PBS-Tween), followed by incubation with a blocking solution composed of 5% (m/v) milk in PBS-Tween for 1 hour, rocking at RT to prevent unspecific binding. The membranes were then incubated overnight, rocking at 4ºC, with the appropriate primary antibodies diluted in blocking solution. Next day, membranes were washed 3x 5 min, rocking in PBS-Tween and incubated with the appropriate horseradish peroxidase (HRP)-coupled secondary antibody diluted in blocking solution, for 1 hour, rocking at RT. After washing the membranes 3x 5 min, rocking in PBS-Tween, detection was achieved by using enhanced chemiluminescence substrates and images were acquired with Amersham 680. As primary antibodies, rabbit anti-Lamin B1 (Abcam, ab16048) and rabbit anti-ORF1p (Abcam, ab216324) were used, at a 1:1,000 dilution in blocking solution. Lamin B1 was used as loading control. The secondary antibody anti-rabbit-HRP (Biorad, 170-6515) was used at a 1:5,000 dilution in blocking solution.

Protein bands were relatively quantified on ImageJ, following the protocol from (Davarinejad, n.d.). The quantification reflects the relative amounts as a ratio of each protein band relative to loading control lane.

2.8. Immunofluorescence

For immunofluorescence assays, cells were passaged and grown (as described in **section 2.1.4**) on sterile 22x22 mm coverslips in a tissue culture 6-well plate for 48 hours. Afterwards, the medium was aspirated, the cells were gently washed with PBS 1X and fixed with a solution of 3.7% (v/v) paraformaldehyde (PFA) in PBS 1X for 10 min at RT. Cells were then washed 3x with PBS 1X and permeabilized with ice-cold solution of 0.5% (v/v) Triton-X100 in PBS 1X for 5 min on ice. Afterwards, coverslips were washed 3x with PBS 1X and blocked with a solution of 1% (v/v) BSA in PBS 1X (blocking solution) for 30 min at RT. Incubation with the appropriate primary antibodies diluted in blocking solution was performed in a humid light-tight box for 1 hour at RT. Then, coverslips were washed 3x 5 min, rocking with PBS-Tween and incubation with the appropriate conjugated secondary antibodies diluted in blocking solution was performed also in a humid light-tight box for 45 min at RT. The coverslips were washed 3x 5 min, rocking with PBS-Tween in the dark (secondary antibody is light sensitive) and then, washed with PBS 1X, rocking for 5 min. Coverslips were incubated in the dark with 1 μg/ml solution of 4',6-diamidino-2-phenylindole (DAPI) (Sigma Aldrich) in PBS 1X for 10 min at RT and then, washed 3x with PBS 1X. Finally, coverslips were mounted on microscope slides with Vectashield Mounting Medium (Baptista Marques) and left to air dry in the dark. All the steps that required the handling of PFA were performed under the fume hood. As primary antibodies, mouse anti-Cas9 (clone 8C1-F10, Active Motif, #61957) and rabbit anti-Cas9 (from the recombinant antibody platform at Institut Curie, Paris, France) were used, at a 1:100 dilution in blocking solution. As secondary antibodies, goat anti-mouse-Dy488 (Bethyl Laboratories, A90-244D2) and goat anti-rabbit-Cy3 (Bethyl Laboratories, A120-201C3) were used at a 1:200 dilution in blocking solution. The wavelengths corresponding to the maximum excitation and emission peak for each fluorophore are 358/461 nm for DAPI, 493/518 nm for Dy488 and 552/570 nm for Cy3. Zeiss Cell Observer, a widefield fluorescence microscope, was used to visualize the cells and acquire images, either with a 20x objective with 0.80 numerical aperture (NA) or with a 63x oil immersion objective with 1.40 NA, depending on the experiment.

Image analysis was performed on ImageJ software, where at least 150 cells from each experiment were manually counted, and the percentage of positive cells was determined.

3. RESULTS

The development of cellular models in which we are able to interfere with L1 expression is an important first step to enable the investigation of the contribution of L1 elements for X-chromosome inactivation and genome stability. As described in **section 1.2.2**, the strongest and most significant mechanism of L1 regulation is related to their chromatin structure, which was shown to undergo significant changes during early development stages. Thus, engineered systems for L1 transcriptional activation and/or repression were established in cells with different

states of differentiation, from undifferentiated mES cells to fully differentiated somatic NIH-3T3 cells, in which the efficiency of these systems was characterized.

3.1. CRISPRa-mediated activation levels of L1Md-Tf expression are not maintained during mES cells differentiation nor in differentiated NPCs

Before I came into the lab, a doxycycline-inducible CRISPR activation (CRISPRa) system had already been established in female mES cells by Anne-Valerie Gendrel (**Figure 5**). In the presence of doxycycline, this system encodes for a dCas9 protein fused to a VPR transcriptional activation complex (including VP64, p65 and Rta transcriptional activators). Later, this dCas9-VPR complex binds two guide RNAs that are complementary to monomers 2 and 3 of the 5'UTR region of L1Md-Tf elements and, in theory, this should lead to genome-wide activation/increase of L1Md-Tf expression.

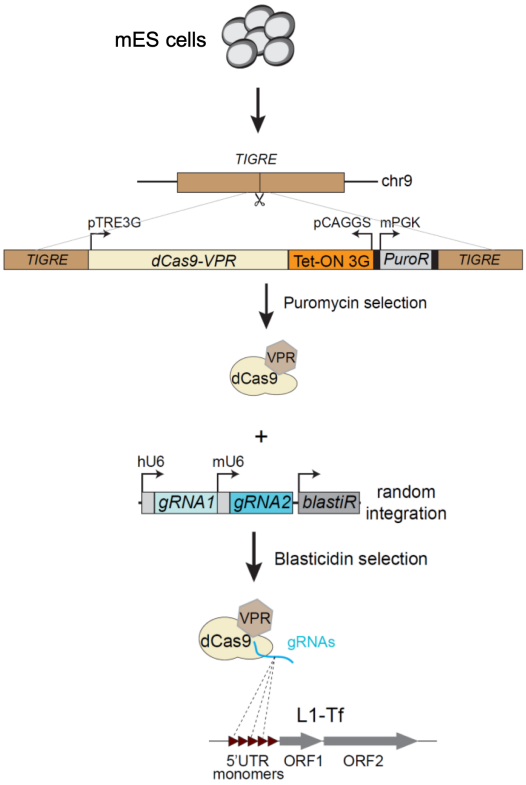

Figure 5: L1Md-Tf activation system implemented in female mES cells. Doxycycline-inducible CRISPR activation system composed by a dCas9 protein fused to VPR transcriptional activation complex. The dCas9-VPR complex then binds two sgRNAs that are complementary to 5'UTR monomers of L1Md-Tf elements, leading to genome-wide activation/increase of L1MdTf expression. The dCas9-VPR transgene was integrated by CRISPR-mediated homologous recombination into TIGRE safe harbor locus and the guide RNAs were randomly integrated into the *Mus musculus* genome.

The transcriptional activation system presented in **Figure 5** was previously validated and characterized by Anne-Valerie Gendrel. For this purpose, doxycycline-treated (1 μg/ml of

doxycycline for 48 hours) and untreated samples of B7, B4 and D4 clones were used. B7 is a control clone that lacks the guide RNAs for 5'UTR monomers of L1Md-Tf elements. B4 and D4 are two clones that contain a functional CRISPRa system, including the guide RNAs for 5'UTR monomers of L1Md-Tf elements. Although in undifferentiated mES cells (**Figure 6A**) it was possible to observe a great increase in expression of dCas9-VPR transgene (~ 200 to 600-fold change) and in genome-wide L1Md-Tf expression (~ 3-fold change) when doxycycline was present, these expression levels were not maintained during mES cells differentiation (**Figure 6B**). In fact, after 4 days of mES cells differentiation, the expression levels of dCas9-VPR transgene become much lower (~ 50 to 100-fold change) than in undifferentiated mES cells and the expression levels of L1Md-Tf elements no longer change in the presence of doxycycline. Furthermore, differentiated NPCs (**Figure 6C**) showed even lower values of dCas9-VPR transgene expression (~ 2.5-fold change) and, although there is some increase in the expression levels of L1Md-Tf elements (~ 1.5-fold change) in the presence of doxycycline, these are much lower than in undifferentiated mES cells. Altogether, these results led us to hypothesize that some chromatin changes could occur during the differentiation process of mES cells that would decrease the accessibility of transcriptional machinery to the dCas9-VPR transgene and the guide RNAs and thus, lead to a decrease of expression. This decrease in expression of the dCas9-VPR transgene and the guide RNAs is not desirable, since it impairs the final goal of the CRISPRa system which is to impact the transcriptional levels of L1Md-Tf elements, by increasing their expression.

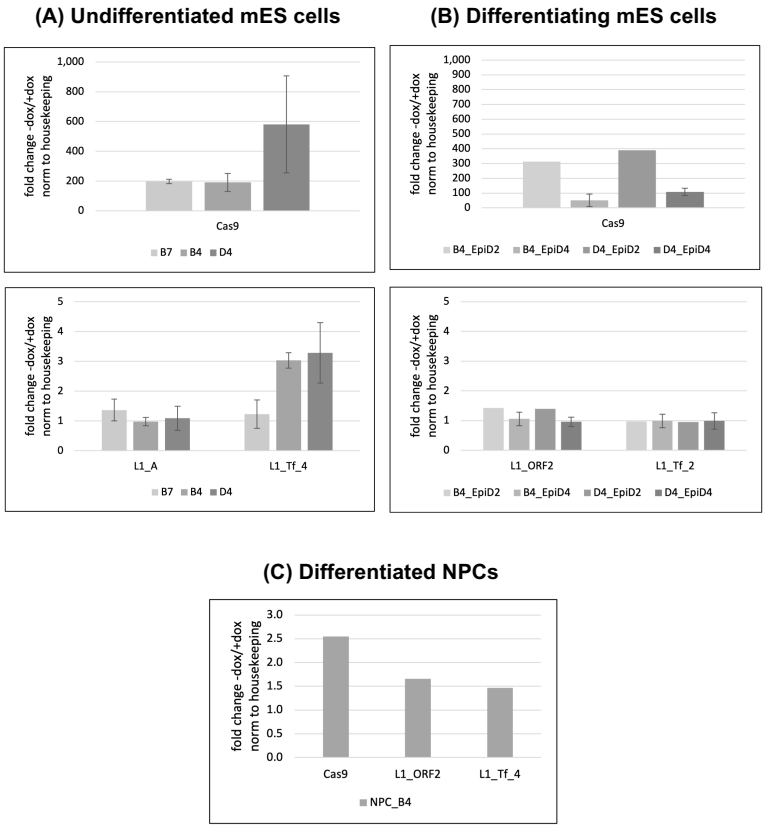

Figure 6: CRISPRa-mediated activation levels of L1Md-Tf expression are not maintained during mES cells differentiation nor in differentiated NPCs. Relative expression levels of dCas9-VPR transgene and L1Md-Tf elements

after 48h treatment with 1 μg/ml of doxycycline in **(A)** undifferentiated female mES cells **(B)** 2-day and 4-day differentiated mES cells and **(C)** differentiated NPCs. Expression levels were obtained by qPCR and analyzed with the $2^{-\Delta\Delta C_t}$ method, using beta-actin as housekeeping gene. Data is normalized to the housekeeping gene and represented as mean of the fold change over the control (doxycycline-untreated cells), +/- standard deviation of the mean (n=3 biological replicates). Cas9, L1_A, L1_ORF2 and L1_Tf correspond to primer pairs that amplify a genomic region of the dCas9-VPR transgene, L1Md-A elements, ORF2 coding region of L1 elements and L1Md-Tf elements, respectively. L1_A was used as a negative control, since the implemented CRISPRa system does not have guide RNAs for L1Md-A elements, and L1_ORF2 as a positive control, given that ORF2 is a conserved region among all L1 elements. B7, B4 and D4 are undifferentiated mES cell clones. EpiD2 and EpiD4 are mES cell clones (B4 or D4) after 2-day and 4-day epiblast stem cell differentiation, respectively. NPC_B4 represents a mES cell clone after differentiation into NPC.

3.2. KRAB-ZFP-mediated repression levels of L1Md-Tf expression are not maintained during mES cells differentiation

The same tendency was observed on a doxycycline-inducible transcriptional repression system that had already been established by Anne-Valerie Gendrel in female mES cells (**Figure 7**). In the presence of doxycycline, this system encodes for zinc finger proteins (ZFP) with an engineered DNA binding domain that is complementary to 5'UTR monomers of L1Md-Tf elements, fused to a Kruppel associated box (KRAB) transcriptional repressor. In theory, this system should lead to genome-wide repression/decrease of L1Md-Tf expression.

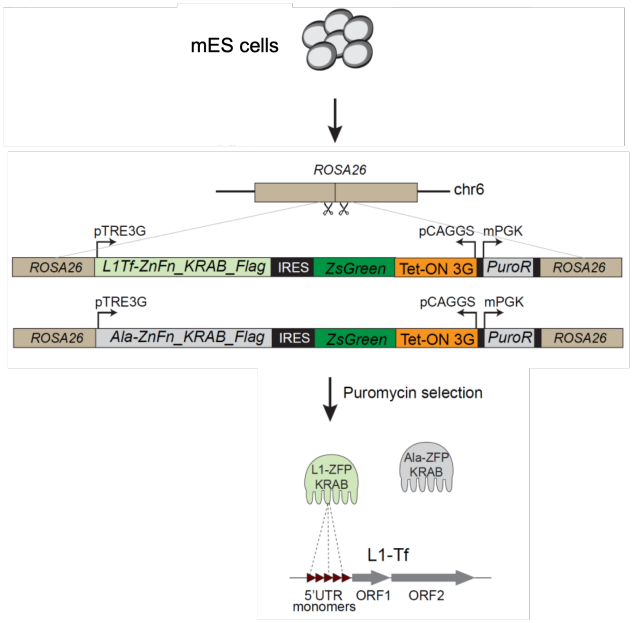

Figure 7: L1Md-Tf repression system implemented in female mES cells. Doxycycline-inducible KRAB-ZFP repression system composed by a KRAB transcriptional repressor fused to engineered ZFPs that bind to 5'UTR monomers of L1Md-Tf elements, leading to genome-wide repression/decrease of L1MdTf expression. The KRAB-ZFP transgene was integrated by CRISPR-mediated homologous recombination into the ROSA26 safe harbor locus, located on chromosome 6 of the *Mus musculus* genome.

Doxycycline-treated (1 μg/ml of doxycycline for 48 hours) and untreated samples of B5, A5 and A6 clones were previously used to validate and characterize this repression system. B5 is a control clone that carries mutations in the DNA binding domain of ZFP, preventing it from binding to 5'UTR monomers of L1Md-Tf elements. A5 and A6 clones are two clones with a functional KRAB-ZFP repression system for L1Md-Tf elements. Similarly to the CRISPRa system, although in undifferentiated mES cells (**Figure 8A**) we are able to observe a great increase in expression of KRAB-ZFP transgene (~ 25 to 65-fold change) and a considerable decrease in genome-wide L1Md-Tf expression (~ 0.2 to 0.4-fold change) when doxycycline was present, these expression levels were not maintained during mES cells differentiation (**Figure 8B**). After 4 days of mES cells differentiation, the expression levels of KRAB-ZFP transgene become lower (~ 20-fold change) than in undifferentiated mES cells and the expression levels of L1Md-Tf elements no longer change in the presence of doxycycline, which indicates that the repression system is no longer able to interfere with the transcriptional levels of these elements. As mentioned before, these results support the hypothesis that some chromatin changes may occur during the differentiation process of mES cells, making the KRAB-ZFP transgene less accessible or even inaccessible to the transcriptional machinery of mES cells thus, impairing the final goal of L1Md-Tf genome-wide repression.

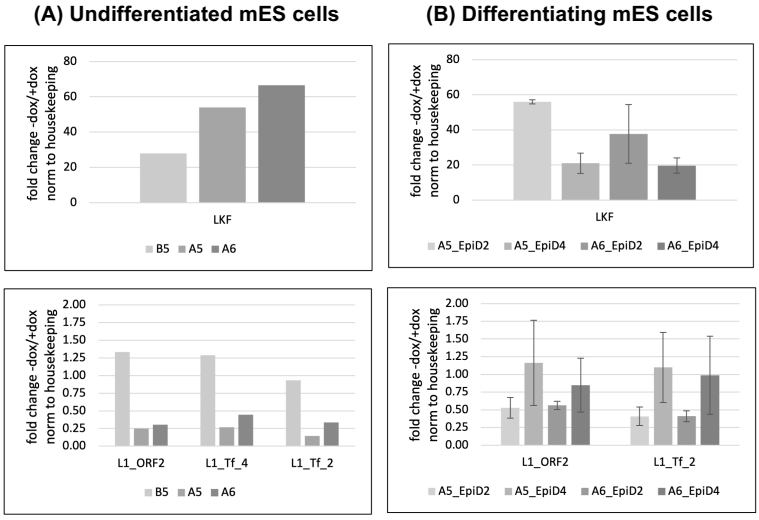

Figure 8: KRAB-ZFP-mediated repression levels of L1Md-Tf expression are not maintained during mES cells differentiation. Relative expression levels of KRAB-ZFP transgene and L1Md-Tf elements after 48h treatment with 1 μg/ml of doxycycline in **(A)** undifferentiated female mES cells and **(B)** 2-day and 4-day differentiated mES cells. Expression levels were obtained by qPCR and analyzed with the $2^{-\Delta\Delta C_t}$ method, using beta-actin as housekeeping gene. Data is normalized to the housekeeping gene and represented as mean of the fold changes over the control (doxycycline-untreated cells), +/- standard deviation of the mean (n=3 biological replicates). LKF, L1_ORF2 and L1_Tf correspond to primer pairs that amplify a genomic region of the KRAB-ZFP transgene, ORF2 coding region of L1 elements and L1Md-Tf elements, respectively. L1_ORF2 was used as a positive control, given that ORF2 is a conserved region among all L1 elements. B5, A5 and A6 are undifferentiated mES cell clones. EpiD2 and EpiD4 are mES cell clones (A5 or A6) after 2-day and 4-day epiblast stem cell differentiation, respectively.

Considering the above-mentioned hypothesis, we wanted to test if we could improve both systems by creating an open chromatin configuration near the transgenes of interest (dCas9-VPR transgene and guide RNAs for the activation system; KRAB-ZFP transgene for the repression system) and thus, establish functional cellular models in which we would be able to interfere with the expression levels of L1Md-Tf elements during differentiation and in differentiated cells, such as NPCs. For that purpose, we decided to grow mES cells (undifferentiated and during differentiation) and differentiated NPCs with the drugs that had previously been used for selection purposes. Since the resistance marker for the selection drugs is located adjacent of the transgenes of interest (puromycin for dCas9-VPR and KRAB-ZFP transgenes and blasticidin for the sgRNAs; **Figure 5** and **Figure 7**), we hypothesize that by growing the cells with these drugs we will force the expression of the resistance genes and thus, create an open chromatin configuration in that region that could also favor the transcription of the adjacent genes – the transgenes of interest – since they would benefit from the imposed open chromatin configuration.

3.3. CRISPRa-mediated activation levels of L1Md-Tf expression do not change after treatment with selection drugs, in undifferentiated mES cells

The impact of the selection drugs was first studied in undifferentiated female mES cells containing the transcriptional activation system. For that purpose, undifferentiated mES cells were grown in ES medium, as described in **section 2.1.1**, supplemented with 1 μg/ml of puromycin and 10 μg/ml of blasticidin. Doxycycline-treated (1 μg/ml of doxycycline for 48 hours) and untreated samples of the mES B7, B4 and D4 clones were then used for expression analysis, at both the RNA and the protein levels.

At the RNA level (**Figure 9**), although it was possible to observe a great increase in expression of dCas9-VPR transgene in the 3 clones analyzed (~ 120 to 270-fold change) and in genome-wide L1Md-Tf expression in B4 and D4 clones (~ 4-fold change) when doxycycline was present, these expression levels are not considerably different from the expression levels obtained in undifferentiated mES cells that were grown in ES medium without selection drugs (**Figure 6A**). More biological replicates would be necessary to increase the robustness of this conclusion, however, as a preliminary result, we can consider that at undifferentiated state, growing mES cells in the presence of selection drugs does not seem to impact the efficiency of the transcriptional activation system, which was already functional.

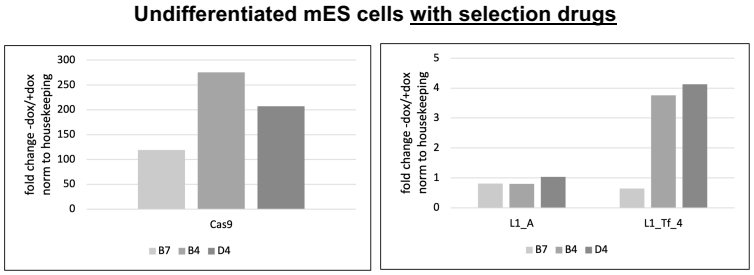

Figure 9: CRISPRa-mediated activation levels of L1Md-Tf expression do not increase after treatment with selection drugs, in undifferentiated mES cells. Relative expression levels of dCas9-VPR transgene and L1Md-Tf

elements after 48h treatment with 1 μg/ml of doxycycline in undifferentiated female mES cells that were grown in ES medium supplemented with 1 μg/ml of puromycin and 10 μg/ml of blasticidin. Expression levels were obtained by qPCR and analyzed with the $2^{-\Delta\Delta Ct}$ method, using Gapdh as housekeeping gene. Data is normalized to the housekeeping gene and represented as fold change over the control (doxycycline-untreated cells). Cas9, L1_A and L1_Tf correspond to primer pairs that amplify a genomic region of the dCas9-VPR transgene, L1Md-A elements and L1Md-Tf elements, respectively. L1_A was used as a negative control, since the implemented CRISPRa system does not have guide RNAs for L1Md-A elements. B7, B4 and D4 are undifferentiated mES cell clones.

At the protein level, immunofluorescence and western blot techniques were used to further assess the impact of growing undifferentiated mES cells in ES medium supplemented with 1 μg/ml of puromycin and 10 μg/ml of blasticidin. Immunofluorescence slides were prepared, according to the protocol described in **section 2.8**, with two different anti-Cas9 primary antibodies (mouse and rabbit) and images were then acquired on Zeiss Cell Observer (**Figure 10B**). The rabbit anti-Cas9 primary antibody was shown to be much more specific than the mouse anti-Cas9 primary antibody, showing less background signal. However, for quantification purposes (**Figure 10C**), only mouse anti-Cas9 images were considered since this was the only anti-Cas9 primary antibody that was also present in previously obtained images (without selection drugs, **Figure 10A**). The percentage of Cas9-positive cells was determined for both conditions (with and without selection drugs), as described in **section 2.8**, and although a higher percentage of Cas9-positive cells might be observed on D4 clone after being grown with selection drugs, more biological replicates would be necessary to assess if this would be a significant change or not. Thus, preliminary results presented in **Figure 10C** led us to conclude that, indeed dCas9 nuclease is being translated in presence of doxycycline, but no considerable enrichment seems to be observed when growing undifferentiated mES cells with selection drugs.

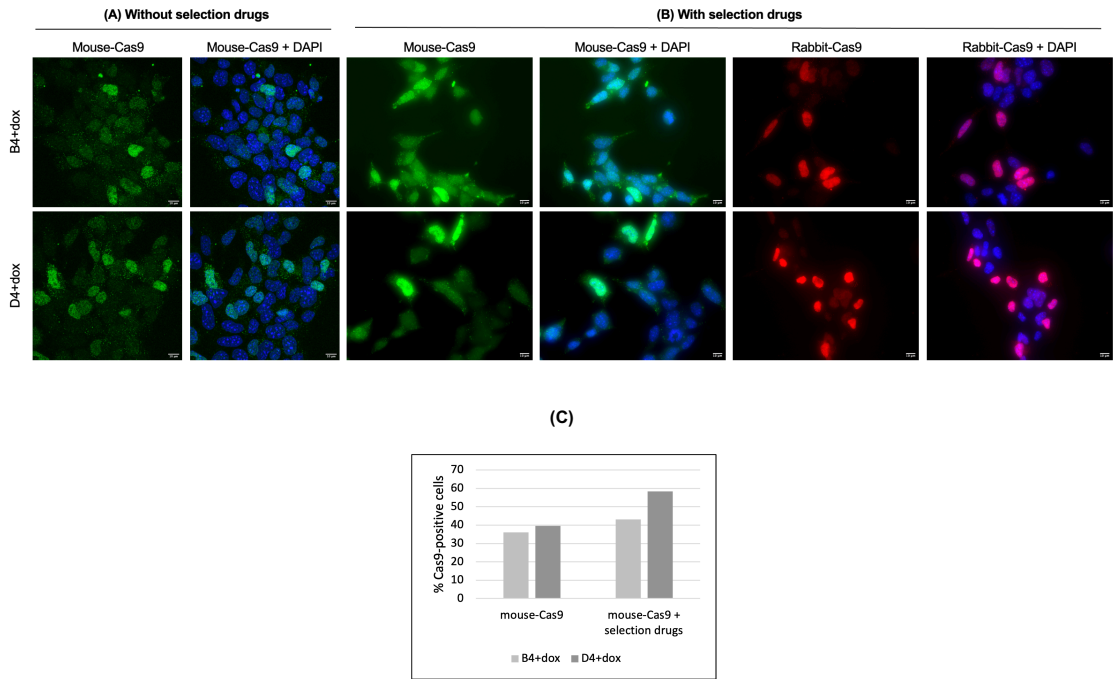

Figure 10: Undifferentiated mES cells population is not enriched in dCas9 nuclease after treatment with selection drugs. Representative immunofluorescence images of doxycycline-treated (1 μg/ml for 48h) mES cells that were grown

in ES medium **(A)** without and **(B)** with 1 μg/ml of puromycin and 10 μg/ml of blasticidin. **(C)** Percentage of Cas9-positive cells in the mES cells population, with and without selection drugs. DAPI is presented here as a nuclear marker. Quantification was only performed on images that were acquired with mouse anti-Cas9 primary antibody. Scale bars: 10 μm.

The expression levels of L1Md-Tf elements were indirectly assessed at the protein level, by performing a western blot (as described in **section 2.7**) with anti-ORF1p primary antibody. The ORF1p protein has two isoforms (41 and 43 kDa), being the upper isoform the one that is mainly expressed from the L1Md-Tf subfamily of L1 elements (Kolosha & Martin, 1995). Thus, the increased intensity of the top ORF1p band in the presence of doxycycline (**Figure *11*A**) indicates that, indeed, the expression levels of L1Md-Tf elements increase in B4 and D4 clones when CRISPRa system is activated. To further confirm these observations, intensity quantification of the upper and lower ORF1p bands was performed (**Figure *11*B**), as described in **section 2.7.3**.

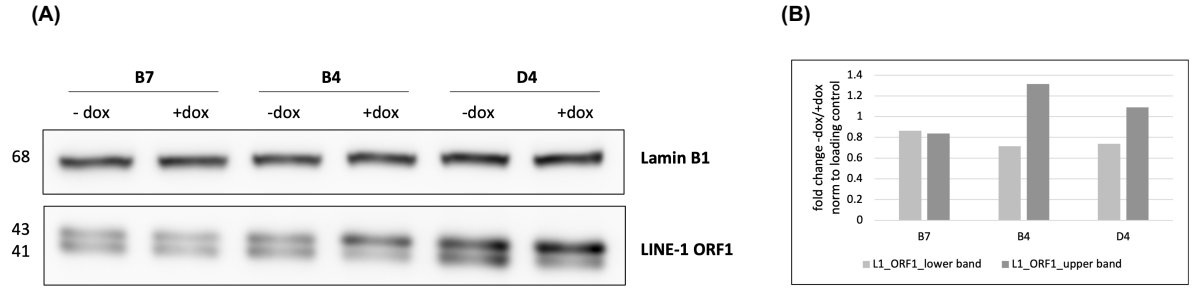

Figure 11: CRISPRa-mediated activation of L1Md-Tf expression confirmed by western blot, in undifferentiated mES cells after treatment with selection drugs. (A) Total protein extracts from doxycycline-treated (1 μg/ml of doxycycline for 48h) and untreated mES cells, that were grown in ES medium supplemented with 1 μg/ml of puromycin and 10 μg/ml of blasticidin, were analysed by western blot. Primary antibodies against Lamin B1 and ORF1p proteins were used. Lamin B1 was used as loading control. Molecular weight is indicated at the left, in kDa. **(B)** Relative quantification of protein bands was performed on the upper and lower bands of ORF1p. Data is normalized to the loading control and represented as fold change over the control (doxycycline-untreated cells).

3.4. KRAB-ZFP-mediated repression of L1Md-Tf expression is maintained after 3 days of random mES cells differentiation, in the presence of selection drugs

The impact of the selection drugs was further studied during differentiation, in female mES cells containing the transcriptional repression system, KRAB-ZFP, in one mES clone (A6 clone). For that purpose, mES cells underwent a 2-day and 3-day random differentiation process by LIF withdrawal (described in **section 2.1.2**) in ES differentiation medium supplemented with 1 μg/ml of puromycin. Doxycycline-treated (1 μg/ml of doxycycline for 48 hours) and untreated samples were harvested for RNA extraction, as described in **section 2.6.1**, and the expression levels of KRAB-ZFP transgene and L1Md-Tf elements during the differentiation process were assessed by qPCR (**Figure 12A**). Although a decrease in KRAB-ZFP transgene expression is observed between days 2 and 3 of differentiation, the levels of transgene expression are still considerably high after 3 days of differentiation ($\sim$ 300-fold change). At the level of L1Md-Tf expression, a genome-wide L1Md-Tf de-repression is observed between days 2 and 3 of differentiation, but

again, a considerable level of repression is still observed after 3 days of differentiation (~ 0.3-fold change). The decrease in expression of Nanog and Klf4 pluripotency markers, as well as the increase in expression of Xist and Fgf5 differentiation markers, during differentiation (**Figure 12B**) show that the differentiation process was successful. Altogether, these results led us to conclude that, after 3 days of mES cells random differentiation by LIF withdrawal, in the presence of 1 μg/ml of puromycin, the L1Md-Tf transcriptional repression system seems to be functional and thus, being able to induce a considerable genome-wide repression of L1Md-Tf expression. More biological replicates and a longer differentiation process – of at least 4 days, to be comparable with the data presented in **Figure 8B** – would be necessary to draw a robust conclusion about the impact of selection drugs on the performance of the transcriptional repression system during mES cell differentiation. However, the expression levels presented in **Figure 12** are similar to the ones that were previously obtained for undifferentiated mES cells without selection drugs (**Figure 8A**), which seems to be a good indication that indeed the use of selection drugs during mES cells differentiation may increase the accessibility of KRAB-ZFP transgene by the transcriptional machinery and thus, increase the performance of the L1Md-Tf transcriptional repression system.

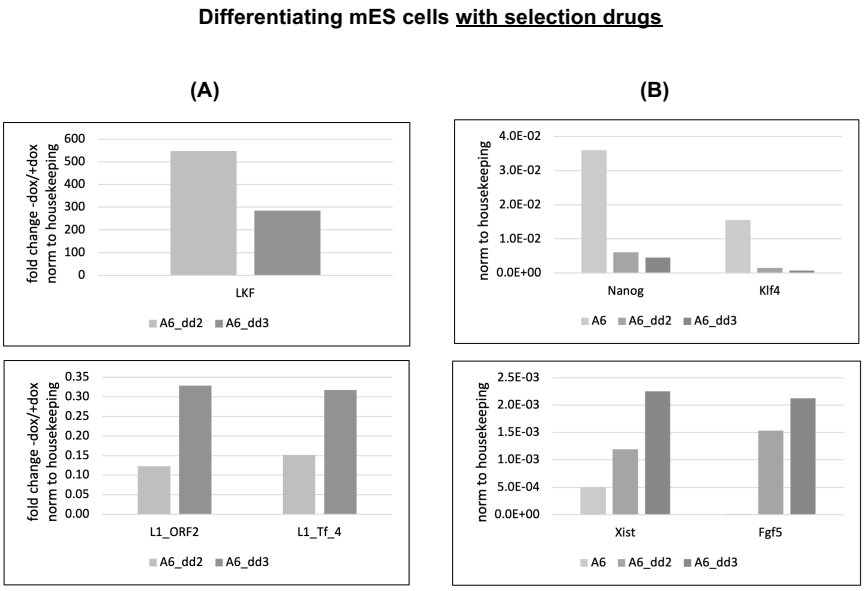

Figure 12: KRAB-ZFP-mediated repression of L1Md-Tf expression is maintained after 3 days of random mES cells differentiation, in the presence of selection drugs. (A) Relative expression levels of KRAB-ZFP transgene and L1Md-Tf elements during mES cells random differentiation by LIF withdrawal in ES differentiation medium supplemented with1 μg/ml of puromycin. **(B)** Relative expression levels of pluripotency markers (Nanog and Klf4) and differentiation markers (Xist and Fgf5), before and during mES cells random differentiation by LIF withdrawal in ES differentiation medium supplemented with1 μg/ml of puromycin. Expression levels were obtained by qPCR and analyzed with the $2^{-\Delta\Delta C_t}$ method, using Gapdh as housekeeping gene. Data is normalized to the housekeeping gene and, in (A), represented as fold change over the control (doxycycline-untreated cells). Doxycycline treatment was performed with 1 μg/ml of doxycycline for 48 hours. LKF, L1_ORF2 and L1_Tf correspond to primer pairs that amplify a genomic region of the KRAB-ZFP transgene, ORF2 coding region of L1 elements and L1Md-Tf elements, respectively. L1_ORF2 was used as a positive control, given that ORF2 is a conserved region among all L1 elements. A6 corresponds to undifferentiated mES cell clone. A6_dd2 and A6_dd3 correspond to A6 clone after 2-day and 3-day random differentiation process by LIF withdrawal, respectively.

3.5. CRISPRa-mediated activation levels of L1Md-Tf expression increase after treatment with selection drugs, in differentiated NPCs

Finally, the impact of the selection drugs was studied in differentiated female NPCs containing the transcriptional activation CRISPRa system, obtained after differentiation of the mES clone B4 (NPC_B4 clone). For that purpose, NPCs were grown in NPC medium, as described in **section 2.1.3**, supplemented with 1 μg/ml of puromycin and 10 μg/ml of blasticidin. Doxycycline-treated (1 μg/ml of doxycycline for 48 hours) and untreated samples were harvested for RNA extraction, as described in **section 2.6.1**, and the expression levels of dCas9-VPR transgene and L1Md-Tf elements were assessed by qPCR (**Figure 13**). A considerable increase in expression of dCas9-VPR transgene (~ 17.5-fold change) and a modest increase in expression of L1Md-Tf elements (~ 2.5-fold change) were observed after growing the cells with selection drugs. Although these expression levels continue to be lower than the expression levels of undifferentiated mES cells without selection drugs (**Figure 6A**), they are higher than the previously obtained ones for differentiated NPCs without selection drugs (**Figure 6C**). Therefore, the results presented in **Figure 13** seem to be a good indication that selection drugs may increase the accessibility of the transgenes (dCas9-VPR and guide RNAs) by the transcriptional machinery and thus, increase the performance of the L1Md-Tf transcriptional activation system in differentiated NPCs. Nevertheless, more biological replicates would be necessary to draw a robust conclusion on this topic.

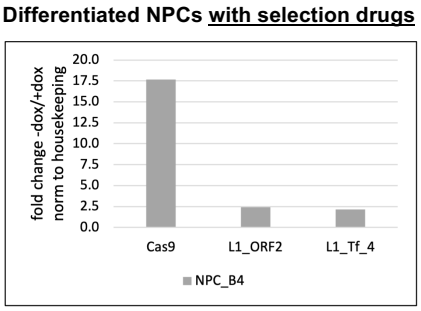

Figure 13: CRISPRa-mediated activation levels of L1Md-Tf expression increase after treatment with selection drugs, in differentiated NPCs. Relative expression levels of dCas9-VPR transgene and L1Md-Tf elements after 48h treatment with 1 μg/ml of doxycycline in differentiated female NPCs that were grown in NPC medium supplemented with 1 μg/ml of puromycin and 10 μg/ml of blasticidin. Expression levels were obtained by qPCR and analyzed with the $2^{-\Delta\Delta C_t}$ method, using Gapdh as housekeeping gene. Data is normalized to the housekeeping gene and represented as fold change over the control (doxycycline-untreated cells). Cas9, L1_ORF2 and L1_Tf correspond to primer pairs that amplify a genomic region of the dCas9-VPR transgene, ORF2 coding region of L1 elements and L1Md-Tf elements, respectively. L1_ORF2 was used as a positive control, given that ORF2 is a conserved region among all L1 elements. B7, B4 and D4 are undifferentiated mES cell clones. NPC_B4 represents a mES cell clone after differentiation into NPC.

3.6. Establishing a CRISPRa system for L1Md-Tf elements in fully differentiated, somatic NIH-3T3 cells

3.6.1. CRISPR-mediated homologous recombination of dCas9-VPR transgene

Next, we sought to implement the L1Md-Tf transcriptional activation system (previously presented in **Figure 5**) in fully differentiated, somatic cells. For that purpose, female NIH-3T3 cell line, which is a standard cell line of mouse embryonic fibroblasts, was used. To begin, the dCas9-VPR transgene was integrated into the TIGRE safe harbor locus by CRISPR-mediated homologous recombination, i.e., by co-transfection of pX330-sgTIGRE-Cas9 and pEN366-dCas9-VPR plasmids as described in **section 2.2**. The cells where the dCas9-VPR transgene was successfully integrated were selected with 2.5 μg/ml of puromycin for one week. Afterwards, 48 puromycin-resistant clonal colonies were picked into a 96-well plate and the remaining puromycin-resistant colonies were pooled and harvested to serve as a genotyping control at the population level. After DNA extraction (as described in **section 2.5.2**), PCR was performed on the population of cells (as described in **section 2.5.3**) with primers that amplify a region of the TIGRE locus that was not perturbed by the CRISPR-mediated homologous recombination event (WT), as a positive control for the PCR reaction, and primers that amplify the left transgene/TIGRE junction after CRISPR-mediated homologous recombination (LeftJ) (**Figure 14**). The fact that the intensity of the band obtained after PCR with LeftJ primers was much lower than the intensity of WT positive control band was already an indication that, although we had been able to integrate the dCas9-VPR transgene at the desired location within the TIGRE safe harbor locus, the efficiency of the CRISPR-mediated homologous recombination at the population level was probably low. Indeed, after performing quick cell lysis (as described in **section 2.5.1**) and PCR genotyping with LeftJ primer pair on the 48 cell clones, only 4 of them seemed to have the dCas9-VPR transgene integrated at the desired location, the TIGRE safe harbor locus (**Figure 15**).

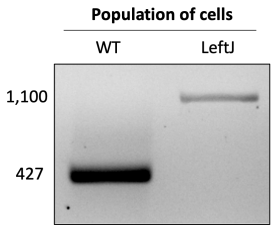

Figure 14: CRISPR-mediated homologous recombination of dCas9-VPR transgene into the TIGRE safe harbor locus was successful but showed low efficiency, at the NIH-3T3 cells population level. PCR genotyping on the population of NIH-3T3 cells after co-transfection of pX330-sgTIGRE-Cas9 and pEN366-dCas9-VPR plasmids. WT primer pair was used as a positive control for the PCR reaction. LeftJ primer pair amplifies the left junction between the dCas9-VPR transgene and the TIGRE locus after CRISPR-mediated homologous recombination event. Molecular weights of interest are presented at the left side, in base pairs (bp).

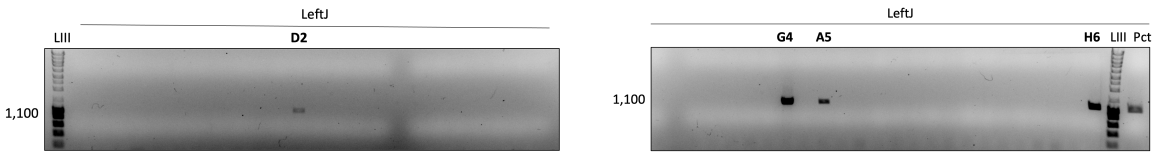

Figure 15: PCR genotyping on cell lysates indicates that four NIH-3T3 cell clones may have the dCas9-VPR transgene integrated at the TIGRE safe harbor locus. PCR genotyping on 48 puromycin-resistant NIH-3T3 cell clones after co-transfection of pX330-sgTIGRE-Cas9 and pEN366-dCas9-VPR plasmids. LeftJ primer pair amplifies the left junction between the dCas9-VPR transgene and the TIGRE locus after CRISPR-mediated homologous recombination event. D2, G4, A5 and H6 are DNA samples of NIH-3T3 cell clones. Pct represents the positive control DNA sample, which was the population of cells after co-transfection of pX330-sgTIGRE-Cas9 and pEN366-dCas9-VPR plasmids. LIII represents the molecular weight marker. Molecular weights of interest are presented at the left side, in base pairs (bp).

Next, the 4 positive clones from **Figure 15** were further expanded and cells were harvested for proper DNA extraction, according to the protocol described in **section 2.5.2**. PCR with WT and LeftJ primer pairs was performed on these DNA samples (**Figure 16**) and, indeed, only two (G4 and H6) out of the four clones were confirmed to have the dCas9-VPR transgene integrated at the TIGRE safe harbor locus.

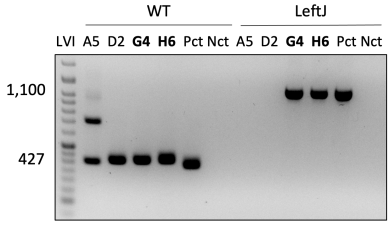

Figure 16: Two NIH-3T3 cell clones were confirmed to have the dCas9-VPR transgene integrated at the TIGRE safe harbor locus by PCR genotyping. PCR genotyping on 4 puromycin-resistant NIH-3T3 cell clones after co-transfection of pX330-sgTIGRE-Cas9 and pEN366-dCas9-VPR plasmids. WT primer pair was used as a positive control for the PCR reaction. LeftJ primer pair amplifies the left junction between the dCas9-VPR transgene and the TIGRE locus after CRISPR-mediated homologous recombination event. A5, D2, G4 and H6 are DNA samples of NIH-3T3 cell clones. Pct represents the positive control DNA sample. Nct represents the negative control DNA sample. LVI represents the molecular weight marker. Molecular weights of interest are presented at the left side, in base pairs (bp).

3.6.2. Induction of dCas9-VPR expression with doxycycline

The confirmed positive clones, G4 and H6, were grown in 3T3 medium and expression of dCas9-VPR transgene was inducted by treating the cells with doxycycline for 48 hours. Since it was the first time using NIH-3T3 cell line, two different concentrations of doxycycline were tested, 1 μg/ml and 2 μg/ml. Then, doxycycline-treated and untreated cells were harvested for RNA extraction, as described in **section 2.6.1**, and the expression levels of dCas9-VPR transgene were assessed by qPCR (**Figure 17A**). At this point, H6 clone was chosen for further studies, given that it showed considerably higher levels of dCas9-VPR transgene expression (~ 26-fold change) after induction with doxycycline, and 1 μg/ml was considered to be the best doxycycline concentration for transgene induction purposes in this cell line. The expression levels

of L1-ORF2 coding region, L1Md-A, L1Md-Gf and L1Md-Tf elements were also assessed in both clones by qPCR (**Figure 17B**). For H6 clone, no activation of L1 elements was observed after treatment with doxycycline, which was the expected behavior for cells lacking guide RNAs complementary with 5'UTR monomers of L1 elements. G4 clone, on the other hand, showed an unexpected behavior and, thus, it was not considered for further studies.

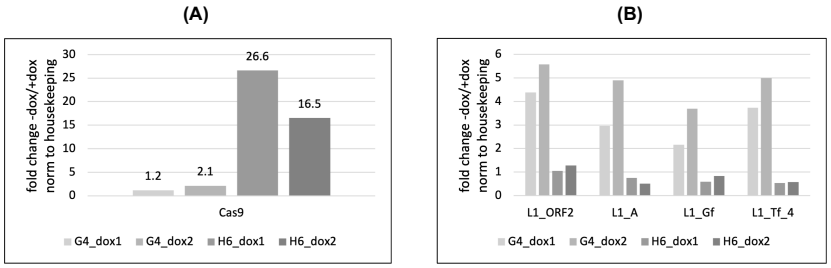

Figure 17: dCas9-VPR transgene expression levels increase, unlike L1 expression levels, in H6 cell clone after treatment with 1 μg/ml of doxycycline. Relative expression levels of **(A)** dCas9-VPR transgene and **(B)** L1 elements, after 48h treatment with 1 μg/ml (dox1) or 2 μg/ml (dox2) of doxycycline in puromycin-resistant NIH-3T3 cell clones. Expression levels were obtained by qPCR and analyzed with the $2^{-\Delta\Delta C_t}$ method, using Gapdh as housekeeping gene. Data is normalized to the housekeeping gene and represented as fold change over the control (doxycycline-untreated cells). Cas9, L1_ORF2, L1_A, L1_Gf and L1_Tf correspond to primer pairs that amplify a genomic region of the dCas9-VPR transgene, ORF2 coding region of L1 elements, L1Md-A elements, L1Md-Gf elements and L1Md-Tf elements, respectively. G4 and H6 are puromycin-resistant NIH-3T3 cell clones.

At the protein level, immunofluorescence was performed (as described in **section 2.8**) with rabbit anti-Cas9 primary antibody to check the dCas9 nuclease levels of H6 clone, before and after treatment with 1 μg/ml of doxycycline for 48h (**Figure 18**). The images were then analyzed, and the percentage of Cas9-positive cells was determined for both conditions, as described in **section 2.8**. Doxycycline-untreated population of cells showed no Cas9-positive cells, which indicates that the system is not leaky, and the population of doxycycline-treated cells presented approximately 15% of Cas9-positive cells. Although this is not a high percentage of Cas9-positive cells, as a preliminary result, it does indicate that dCas9 nuclease is only being translated in the presence of doxycycline and that the implemented system, so far, is functional.

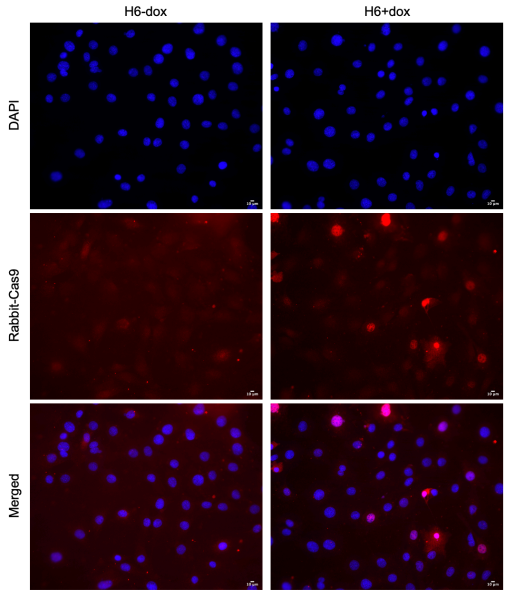

Figure 18: dCas9 nuclease translation is activated in NIH-3T3 cell clone H6 after induction with doxycycline. Representative immunofluorescence images of doxycycline-untreated and treated (1 μg/ml for 48h) NIH-3T3 cells of H6 clone. DAPI is presented here as a nuclear marker. Quantification was performed on both conditions, resulting in 0% and 15% of Cas9-positive cells for untreated and treated cell populations, respectively. Scale bars: 10 μm.

3.6.3.Random integration of sgRNAs for 5'UTR of L1Md-Tf elements

To achieve a complete CRISPRa system, the next step was to transfect sgRNAs that were complementary to 5'UTR monomers of L1Md-Tf elements and contained appropriate PAM sequences, allowing them to be recognized by the dCas9 nuclease. For this purpose, H6 clone was grown in 3T3 medium and transfected with pLK01-sgTfmono2-3 plasmid, as described in **section 2.2**. The cells where the transgene was successfully integrated, by random integration, were selected with 10 μg/ml of blasticidin for one week. Then, a 48-hour treatment with 1 μg/ml of doxycycline was performed in part of the blasticidin-resistant cells to induce the expression of the dCas9-VPR transgene and study the impact of the implemented CRISPRa system on L1Md-Tf elements, at the population level. For that purpose, doxycycline-treated and untreated cells were harvested for RNA extraction, as described in **section 2.6.1**, and the expression levels of dCas9-VPR transgene and L1Md-Tf elements were assessed by qPCR (**Figure 19**). At this point, a great increase in expression of dCas9-VPR transgene (~ 55-fold change) and only a modest activation of L1Md-Tf elements (~ 1.6-fold change) was observed at the population of cells after transfection of sgRNAs. Although at sub-optimal levels, the preliminary results presented in **Figure 19** indicate that the transcriptional activation system is, at least, functional and leading to some L1Md-Tf activation, when compared with the control clone (clone H6 without sgRNAs for L1Md-Tf elements). To draw a robust conclusion on the present efficiency of the transcriptional activation system at the NIH-3T3 population level, more biological replicates would be necessary. Nevertheless, to establish a stable and clonal CRISPRa cell line, blasticidin-resistant colonies should be individually picked at this stage and the expression levels of dCas9-VPR transgene

and L1Md-Tf elements should then be reassessed by qPCR, at the cell clone level. Protein-level techniques such as immunofluorescence and western blot would further complement the validation and characterization of the clonal cell line at that point.

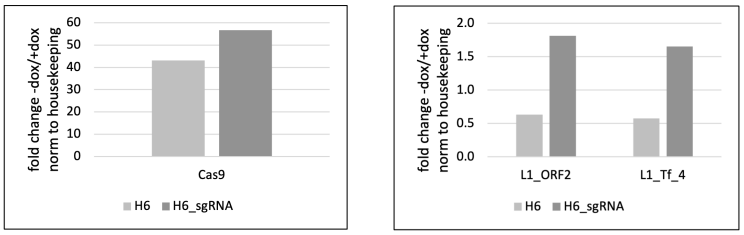

Figure 19: CRISPRa-mediated activation of L1Md-Tf expression was observed by qPCR, in fully differentiated and somatic NIH-3T3 cells. Relative expression levels of dCas9-VPR transgene and L1Md-Tf elements after 48h treatment with 1 μg/ml of doxycycline in female somatic NIH-3T3 cells. Expression levels were obtained by qPCR and analyzed with the $2^{-\Delta\Delta C_t}$ method, using Gapdh as housekeeping gene. Data is normalized to the housekeeping gene and represented as fold change over the control (doxycycline-untreated cells). Cas9, L1_ORF2 and L1_Tf correspond to primer pairs that amplify a genomic region of the dCas9-VPR transgene, ORF2 coding region of L1 elements and L1Md-Tf elements, respectively. L1_ORF2 was used as a positive control, given that ORF2 is a conserved region among all L1 elements. H6 is a NIH-3T3 cell clone without sgRNAs for L1Md-Tf elements. H6_sgRNA represents the population of cells after transfection of sgRNAs for L1Md-Tf elements into H6 clone.

Even with preliminary results only, we can already observe that the L1Md-Tf activation levels obtained in fully differentiated, somatic NIH-3T3 cells are very similar to the ones that were previously obtained for differentiated NPCs without selection drugs (**Figure 6C**) and, thus, considerably lower than the previously obtained expression levels in undifferentiated mES cells without selection drugs (**Figure 6A**). This preliminary observation may lead us to hypothesize that, indeed, the level of cellular differentiation might hamper the transcriptional interference that one is able to induce on L1Md-Tf elements.

4. DISCUSSION

After treating differentiated cells – such as NPCs – with selection drugs, although a great increase in expression of dCas9-VPR transgene was observed (**Figure 13**) in comparison to the results without selection drugs (**Figure 6C**), it only translated into a small increase in L1Md-Tf expression. This suggests that the selection drugs may, indeed, be increasing the accessibility of the transgenes (dCas9-VPR and sgRNAs) to the transcriptional machinery by creating an open chromatin configuration near each transgene of interest. However, the increased efficiency of the CRISPRa system does not translate into a proportional increase in the expression levels of L1Md-Tf elements, which indicates that the level of silencing occurring locally, at the 5'UTR of L1Md-Tf elements, is hampering the binding between the activation complex and the promoter region of these elements. Following induction of dCas9-VPR with doxycycline, somatic NIH-3T3 cells showed L1Md-Tf expression levels (**Figure 19**) similar to the ones observed in differentiated NPCs without selection drugs (**Figure 6C**), which supports the previously-mentioned hypothesis.

The gradual decrease in expression of dCas9-VPR transgene that is observed during the mESC differentiation process (**Figure 6B**) suggests that, indeed, some structural changes might occur during differentiation, possibly at the chromatin level, that may lead to a closed chromatin configuration. The same trend was observed for the transcriptional repression system, in which the expression levels of KRAB-ZFP transgene decrease during differentiation (**Figure 8B**). Nevertheless, the expression levels of KRAB-ZFP transgene after 4 days of differentiation should still be enough to induce, at least, some repression of L1Md-Tf elements, which is not happening since the repression system after 4 days of differentiation is no longer able to interfere with L1Md-Tf expression levels. Only after increasing greatly the expression levels of KRAB-ZFP transgene, by using the selection drugs during differentiation, we are able to observe that some L1Md-Tf repression is maintained after 3 days of differentiation (**Figure 12A**). These results suggest, once again, that the implemented interference system is functional and becomes more efficient when selection drugs are used, but not being able to proportionally interfere with L1Md-Tf expression levels, while the cells become more differentiated.

Undifferentiated mES cells, on the other hand, show a completely different behavior, with high expression levels of activation/repression transgenes and a considerably higher level of interference on L1Md-Tf expression (**Figure 6** and **Figure 8**). Furthermore, no significant changes are observed when selection drugs are used, neither at the expression level of the transgene nor of L1Md-Tf elements (**Figure 9**). These results suggest that the implemented translational activation/repression systems in undifferentiated mES cells are functional and efficient, even when selection drugs are not used. Thus, we may hypothesize that, at the undifferentiated state, a closed chromatin configuration is not present near the transgenes of interest and that the promoter region of L1Md-Tf elements is more accessible to the activation/repression complexes, than it is in further differentiated cells, which would explain the greater levels of L1 interference that one can achieve in undifferentiated mES cells.

After analyzing all the obtained results, we conclude that, although the implemented interference systems show a great potential in interfering with L1Md-Tf expression in cellular models with different stages of differentiation, their efficiency during differentiation and in differentiated cells is still sub-optimal. Here, two problems may arise that decrease the efficiency of these systems: i) low transcription levels of the activation/repression transgenes due to closed chromatin configuration, that is possibly established during the differentiation process, and ii) low accessibility of L1Md-Tf 5'UTR to the activation/repression complexes, due to epigenetic silencing marks that are deposited on these regions. As previously mentioned, growing the cells in the presence of selection drugs may impose an open chromatin configuration in the drugs resistance genes, that are located adjacent to the activation/repression transgenes, favoring the accessibility of the transgenes by the transcriptional machinery of the cells. Thus, the use of selection drugs seems to solve problem (i). Regarding problem (ii), (Wiznerowicz et al., 2007) have shown that epigenetic silencing of L1 elements occur once mESCs undergo differentiation. Thus, to be able to considerably interfere with L1Md-Tf expression levels during differentiation and in differentiated cells, besides increasing the expression levels of activation/repression transgenes, one must also

remove the “epigenetic armor” from the 5’UTR of L1Md-Tf elements to increase the accessibility of these regions. Only then, one would be able to establish a stronger L1 interference system that could be used to study the impacts of L1 expression levels in X-chromosome inactivation and genome stability.

4.1. Constraints and troubleshooting

To establish a CRISPRa system in fully differentiated cells, the use of NIH-3T3 cells appeared initially to be a good option, since this is a standard mouse cell line of female somatic cells. Thus, a system identical to the one that had previously been established in mES cells (**Figure 5**) could be implemented in fully differentiated, somatic cells. However, after the first attempt of integrating a dCas9-VPR transgene into the TIGRE safe harbor locus by CRISPR-mediated homologous recombination, 16 puromycin-resistant colonies were picked and none of them was confirmed to be a positive clone by genotyping with LeftJ primers. After this first try, the amount of transfected DNA (pX330-sgTIGRE-Cas9 and pEN366-dCas9-VPR plasmids) was increased and 48 puromycin-resistant colonies were picked, as described in **section 3.6.1**. Even then, only two clones were confirmed positive by genotyping (**Figure 16**) and only one of them showed the expected behavior by qPCR, after induction of dCas9-VPR expression with doxycycline (**Figure 17**). Thus, although we were able to obtain one positive clone, i.e., with the dCas9-VPR transgene integrated at the TIGRE safe harbor locus and a functional CRISPRa system in general (**Figure 19**), the efficiency of CRISPR-mediated homologous recombination in NIH-3T3 cells seems to be sub-optimal.

4.2. Future work

To correlate the translation of dCas9 and ORF1p, at the protein level, immunofluorescence images of doxycycline-treated and untreated B4 and D4 mES cell clones could be acquired with anti-dCas9 and anti-ORF1p primary antibodies, simultaneously. Then, intensity quantification could be performed on ImageJ to i) determine the level of ORF1p signal enhancement that is induced by the doxycycline treatment, ii) evaluate if the cells that show a higher enhancement of ORF1p signal are, simultaneously, dCas9-positive cells and also iii) determine the percentage of dCas9-positive cells that show a strong ORF1p signal enhancement.

Although the L1Md-Tf activation levels obtained in NIH-3T3 cells with dCas9-VPR were only modest, preliminary qPCR results suggested that the CRISPRa system is functional (**Figure 19**). Thus, a clonal and stable CRISPRa cell line should be established by picking individual colonies of cells transfected with both dCas9-VPR and sgRNAs transgenes. Then, further validation and characterization should be performed by qPCR, western blot and immunofluorescence, in accordance with what was previously done in mES cells. Additionally, more biological replicates of the experiments should be obtained to increase the robustness of the data, not only in NIH-3T3 cells but also in the previously mentioned experiments with mES cells and NPCs.

Furthermore, a higher level of perturbation could be achieved if we were able to interfere simultaneously with all currently active subfamilies of L1 elements (L1Md-A, L1Md-Tf and L1Md-Gf) instead of interfering with L1Md-Tf elements only. Briefly, this could be achieved by cloning two additional pairs of sgRNAs, one against L1Md-A elements and another against L1Md-Gf elements, into the pLK01-sgTfmono2-3 plasmid. This new construct could then be used to establish a stable and clonal CRISPRa cell line.

4.3. Concluding remarks

Transposable elements are DNA sequences with the ability to change their position in the genome, accounting for more than 50% of the mouse genome. Among these, L1 retrotransposons have remained highly active in the mouse lineage. Although L1 elements are usually silenced by epigenetic marks, there are developmental time windows, like the early developmental stages, during which L1 elements are not silenced or fully silenced. This time windows represent a great window of opportunity for the activation of L1 elements, which has been hypothesized to play an important role in development. Here, one appealing hypotheses is that the activation of L1 elements during early development may facilitate the process of X-chromosome inactivation. Thus, the development of functional cellular models in which we are able to interfere (by activation or repression) with L1 expression genome-wide is essential to further investigate the role of L1 elements in XCI and in genome stability.

In the present study, an attempt to optimize L1 perturbation systems during differentiation of female mES cells and in female differentiated NPCs was tested, by culturing the cells with selection drugs. The obtained preliminary results suggest that the use of selection drugs leads to a considerable improvement in the expression levels of the activation/repression transgenes, but not in the levels of L1Md-Tf perturbation. Nevertheless, both transcriptional activation and repression systems are apparently functional and lead to L1Md-Tf expression interference, in comparison to previous results without the use of selection drugs. Furthermore, a CRISPRa system for L1Md-Tf activation was also implemented in fully differentiated mouse embryonic fibroblasts (female NIH-3T3 cells). Preliminary results suggest that the activation system is functional in NIH-3T3 cells, although only a modest level of L1Md-Tf perturbation has been achieved.

We hypothesize that L1Md-Tf perturbation levels may be considerably improved if one is able to specifically remove the “epigenetic armor” that is deposited on the promoter region of these elements, once differentiation is induced. Thus, cellular models with optimized levels of genome-wide L1Md-Tf perturbation could possibly be obtained, allowing for the investigation of the role of L1 elements in XCI and in genome stability, during the several stages of development – from undifferentiated to fully differentiated states.

5. REFERENCES

Adey, N. B., Schichman, S. A., Graham, D. K., Peterson, S. N., Edgell, M. H., & Hutchison, C. A. (1994). Rodent L1 evolution has been driven by a single dominant lineage that has repeatedly acquired new transcriptional regulatory sequences. *Molecular Biology and Evolution*, *11*(5), 778–789.

Akagi, K., Li, J., Stephens, R. M., Volfovsky, N., & Symer, D. E. (2008). Extensive variation between inbred mouse strains due to endogenous L1 retrotransposition. *Genome Research*, *18*(6), 869.

Bailey, J. A., Carrel, L., Chakravarti, A., & Eichler, E. E. (2000). Molecular evidence for a relationship between LINE-1 elements and X chromosome inactivation: The Lyon repeat hypothesis. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(12), 6634.

Bodak, M., Yu, J., & Ciaudo, C. (2014). Regulation of LINE-1 in mammals. *Biomolecular Concepts*, *5*(5), 409–428.

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

Bourque, G., Burns, K. H., Gehring, M., Gorbunova, V., Seluanov, A., Hammell, M., Imbeault, M., Izsvák, Z., Levin, H. L., Macfarlan, T. S., Mager, D. L., & Feschotte, C. (2018). Ten things you should know about transposable elements. *Genome Biology*, *19*(1).

Boyle, A. L., Ballard, S. G., & Ward, D. C. (1990). Differential distribution of long and short interspersed element sequences in the mouse genome: chromosome karyotyping by fluorescence in situ hybridization. *Proceedings of the National Academy of Sciences of the United States of America*, *87*(19), 7757.

Burwinkel, B., & Kilimann, M. (1998). Unequal homologous recombination between LINE-1 elements as a mutational mechanism in human genetic disease. *Journal of Molecular Biology*, *277*(3), 513–517.

Buzin, C. H., Mann, J. R., & Singer-Sam, J. (1994). Quantitative RT-PCR assays show Xist RNA levels are low in mouse female adult tissue, embryos and embryoid bodies. *Development*, *120*(12), 3529–3536.

Chaumeil, J., Baccon, P. le, Wutz, A., & Heard, E. (2006). A novel role for Xist RNA in the formation of a repressive nuclear compartment into which genes are recruited when silenced. *Genes & Development*, *20*(16), 2223.

Chaumeil, J., Okamoto, I., & Heard, E. (2004). X-chromosome inactivation in mouse embryonic stem cells: analysis of histone modifications and transcriptional activity using immunofluorescence and FISH. *Methods in Enzymology*, *376*, 405–419.

Chavez, A., Scheiman, J., Vora, S., Pruitt, B. W., Tuttle, M., Iyer, E., Lin, S., Kiani, S., Guzman, C. D., Wiegand, D. J., Ter-Ovanesyan, D., Braff, J. L., Davidsohn, N., Housden, B. E., Perrimon, N., Weiss, R., Aach, J., Collins, J. J., & Church, G. M. (2015). Highly-efficient Cas9-mediated transcriptional programming. *Nature Methods*, *12*(4), 326.

Chow, J. C., Ciaudo, C., Fazzari, M. J., Mise, N., Servant, N., Glass, J. L., Attreed, M., Avner, P., Wutz, A., Barillot, E., Greally, J. M., Voinnet, O., & Heard, E. (2010). LINE-1 Activity in Facultative Heterochromatin Formation during X Chromosome Inactivation. *Cell*, *141*(6), 956–969.

Davarinejad, H. (n.d.). *Quantification of Western Blot with ImageJ*. Retrieved November 11, 2021, from http://rsb.info.nih.gov/ij/

DeBerardinis, R. J., & Kazazian Jr., H. H. (1999). Analysis of the promoter from an expanding mouse retrotransposon subfamily. *Genomics*, *56*(3), 317–323.

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., & Gilbert, N. (2010). Characterization of LINE-1 Ribonucleoprotein Particles. *PLOS Genetics*, *6*(10).

Feng, Q., Moran, J. v, Kazazian Jr., H. H., & Boeke, J. D. (1996). Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell*, *87*(5), 905–916.

Fuentes, D. R., Swigut, T., & Wysocka, J. (2018). Systematic perturbation of retroviral LTRs reveals widespread long-range effects on human gene regulation. *ELife*, *7*.

Garcia-Perez, J. L., Marchetto, M. C. N., Muotri, A. R., Coufal, N. G., Gage, F. H., O'Shea, K. S., & Moran, J. v. (2007). LINE-1 retrotransposition in human embryonic stem cells. *Human Molecular Genetics*, *16*(13), 1569–1577.

Gilbert, L. A., Larson, M. H., Morsut, L., Liu, Z., Brar, G. A., Torres, S. E., Stern-Ginossar, N., Brandman, O., Whitehead, E. H., Doudna, J. A., Lim, W. A., Weissman, J. S., & Qi, L. S. (2013). CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes. *Cell*, *154*(2), 442.

Gilbert, N., Lutz-Prigge, S., & Moran, J. v. (2002). Genomic Deletions Created upon LINE-1 Retrotransposition. *Cell*, *110*(3), 315–325.

Goerner-Potvin, P., & Bourque, G. (2018). Computational tools to unmask transposable elements. *Nature Reviews. Genetics*, *19*(11), 688–704.

Goodier, J. L., Ostertag, E. M., Du, K., Kazazian, H. H., & Jr. (2001). A Novel Active L1 Retrotransposon Subfamily in the Mouse. *Genome Research*, *11*(10), 1677.

Gu, B., Swigut, T., Spencley, A., Bauer, M. R., Chung, M., Meyer, T., & Wysocka, J. (2018). Transcription-coupled changes in nuclear mobility of mammalian cis-regulatory elements. *Science (New York, N.Y.)*, *359*(6379), 1050.

Han, J. S., Szak, S. T., & Boeke, J. D. (2004). Transcriptional disruption by the L1 retrotransposon and implications for mammalian transcriptomes. *Nature 2004 429:6989*, *429*(6989), 268–274.

Honda, T., Nishikawa, Y., Nishimura, K., Teng, D., Takemoto, K., & Ueda, K. (2020). Effects of activation of the LINE-1 antisense promoter on the growth of cultured cells. *Scientific Reports 2020 10:1*, *10*(1), 1–8.

Jachowicz, J. W., Bing, X., Pontabry, J., Bošković, A., Rando, O. J., & Torres-Padilla, M.-E. (2017). LINE-1 activation after fertilization regulates global chromatin accessibility in the early mouse embryo. *Nature Genetics 2017 49:10*, *49*(10), 1502–1510.

Jachowicz, J. W., & Torres-Padilla, M. E. (2016). LINEs in mice: features, families, and potential roles in early development. *Chromosoma*, *125*(1), 29–39.

Kano, H., Godoy, I., Courtney, C., Vetter, M. R., Gerton, G. L., Ostertag, E. M., Kazazian, H. H., & Jr. (2009). L1 retrotransposition occurs mainly in embryogenesis and creates somatic mosaicism. *Genes & Development*, *23*(11), 1303.

Khan, H., Smit, A., & Boissinot, S. (2006). Molecular evolution and tempo of amplification of human LINE-1 retrotransposons since the origin of primates. *Genome Research*, *16*(1), 78–87.

Kolosha, V. O., & Martin, S. L. (1995). Polymorphic Sequences Encoding the First Open Reading Frame Protein from LINE-1 Ribonucleoprotein Particles. *Journal of Biological Chemistry*, *270*(6), 2868–2873.

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., Fitzhugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., Levine, R., McEwan, P., … Morgan, M. J. (2001). Initial sequencing and analysis of the human genome. *Nature*, *409*(6822), 860–921.

Liu, N., Lee, C. H., Swigut, T., Grow, E., Gu, B., Bassik, M. C., & Wysocka, J. (2017). Selective silencing of euchromatic L1s revealed by genome-wide screens for L1 regulators. *Nature*, *553*(7687), 228–232.

Lu, J. Y., Shao, W., Chang, L., Yin, Y., Li, T., Zhang, H., Hong, Y., Percharde, M., Guo, L., Wu, Z., Liu, L., Liu, W., Yan, P., Ramalho-Santos, M., Sun, Y., & Shen, X. (2020). Genomic Repeats Categorize Genes with Distinct Functions for Orchestrated Regulation. *Cell Reports*, *30*(10), 3296-3311.

Luan, D. D., Korman, M. H., Jakubczak, J. L., & 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*(4), 595–605.

Lyon, M. F. (1961). Gene Action in the X-chromosome of the Mouse (Mus musculus L.). *Nature*, *190*(4773), 372–373.

Lyon, M. F. (1962). Sex Chromatin and Gene Action in the Mammalian X-Chromosome. *American Journal of Human Genetics*, *14*(2), 135.

Lyon, M. F. (1998). X-chromosome inactivation: a repeat hypothesis. *Cytogenetics and Cell Genetics*, *80*(1–4), 133–137.

Makałowski, W., & Boguski, M. S. (1998). Evolutionary parameters of the transcribed mammalian genome: An analysis of 2,820 orthologous rodent and human sequences. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(16), 9407.

Martin, S. L., & Bushman, F. D. (2001). Nucleic Acid Chaperone Activity of the ORF1 Protein from the Mouse LINE-1 Retrotransposon. *Molecular and Cellular Biology*, *21*(2), 467.

Mathias, S., Scott, A., Kazazian, H., Boeke, J., & Gabriel, A. (1991). Reverse transcriptase encoded by a human transposable element. *Science*, *254*(5039), 1808–1810.

Mayer, W., Niveleau, A., Walter, J., Fundele, R., & Haaf, T. (2000). Demethylation of the zygotic paternal genome. *Nature*, *403*(6769), 501–502.

Medstrand, P., Lagemaat, L. N. van de, & Mager, D. L. (2002). Retroelement Distributions in the Human Genome: Variations Associated With Age and Proximity to Genes. *Genome Research*, *12*(10), 1483.

Meissner, A., Mikkelsen, T. S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B. E., Nusbaum, C., Jaffe, D. B., Gnirke, A., Jaenisch, R., & Lander, E. S. (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature*, *454*(7205), 766.

Moran, J. v, Holmes, S. E., Naas, T. P., DeBerardinis, R. J., & Boeke, J. D. (1996). High frequency retrotransposition in cultured mammalian cells. *Cell*, *87*(5), 917–927.

Naas, T. P., DeBerardinis, R. J., Moran, J. v, Ostertag, E. M., Kingsmore, S. F., Seldin, M. F., Hayashizaki, Y., Martin, S. L., & Kazazian, H. H. (1998). An actively retrotransposing, novel subfamily of mouse L1 elements. *The EMBO Journal*, *17*(2), 590.

Okamoto, I., Patrat, C., Thépot, D., Peynot, N., Fauque, P., Daniel, N., Diabangouaya, P., Wolf, J.-P., Renard, J.-P., Duranthon, V., & Heard, E. (2011). Eutherian mammals use diverse strategies to initiate X-chromosome inactivation during development. *Nature*, *472*(7343), 370–374.

Percharde, M., Lin, C.-J., Yin, Y., Guan, J., Peixoto, G. A., Bulut-Karslioglu, A., Biechele, S., Huang, B., Shen, X., & Ramalho-Santos, M. (2018). A LINE1-Nucleolin Partnership Regulates Early Development and ESC Identity. *Cell*, *174*(2), 391-405.

Ross, M. T., Grafham, D. v., Coffey, A. J., Scherer, S., McLay, K., Muzny, D., Platzer, M., Howell, G. R., Burrows, C., Bird, C. P., Frankish, A., Lovell, F. L., Howe, K. L., Ashurst, J. L., Fulton, R. S., Sudbrak, R., Wen, G., Jones, M. C., Hurles, M. E., … Bentley, D. R. (2005). The DNA sequence of the human X chromosome. *Nature*, *434*(7031), 325.

Rossant, J., & McKerlie, C. (2001). Mouse-based phenogenomics for modelling human disease. *Trends in Molecular Medicine*, *7*(11), 502–507.

Smith, Z. D., Chan, M. M., Mikkelsen, T. S., Gu, H., Gnirke, A., Regev, A., & Meissner, A. (2012). A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature, 484*(7394), 339–344.

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

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., Gilbert, N., & Cristofari, G. (2019). The Landscape of L1 Retrotransposons in the Human Genome Is Shaped by Pre-insertion Sequence Biases and Post-insertion Selection. *Molecular Cell*, *74*(3), 555-570.

Symer, D. E., Connelly, C., Szak, S. T., Caputo, E. M., Cost, G. J., Parmigiani, G., & Boeke, J. D. (2002). Human L1 Retrotransposition Is Associated with Genetic Instability In Vivo. *Cell*, *110*(3), 327–338.

Thomas, C. A., Paquola, A. C. M., & Muotri, A. R. (2012). LINE-1 Retrotransposition in the Nervous System. *Annual Review of Cell and Developmental Biology*, *28*, 555–573.

Waterston, R. H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J. F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P., Antonarakis, S. E., Attwood, J., Baertsch, R., Bailey, J., Barlow, K., Beck, S., Berry, E., Birren, B., Bloom, T., … Lander, E. S. (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature*, *420*(6915), 520–562.

Wiznerowicz, M., Jakobsson, J., Szulc, J., Liao, S., Quazzola, A., Beermann, F., Aebischer, P., & Trono, D. (2007). The Krüppel-associated box repressor domain can trigger de novo promoter methylation during mouse early embryogenesis. *Journal of Biological Chemistry*, *282*(47), 34535–34541.

Wutz, A., & Jaenisch, R. (2000). A Shift from Reversible to Irreversible X Inactivation Is Triggered during ES Cell Differentiation. *Molecular Cell*, *5*(4), 695–705.

Xiao-Jie, L., Hui-Ying, X., Qi, X., Jiang, X., & Shi-Jie, M. (2015). LINE-1 in cancer: multifaceted functions and potential clinical implications. *Genetics in Medicine*, *18*(5), 431–439.