Neurofibromin 2 (*NF2*) is Necessary for Efficient Silencing of LINE-1 Retrotransposition Events in Human Embryonic Carcinoma Cells

by

Trenton J. Frisbie

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Human Genetics)
in the University of Michigan
2022

Doctoral Committee:

Professor John V. Moran, Chair
Associate Professor Sundeep Kalantry
Assistant Professor Jacob Kitzman
Assistant Professor Kaushik Ragunathan
Associate Professor JoAnn Sekiguchi

Trenton J. Frisbie

friz@umich.edu

ORCID iD: 0000-0002-5472-1124

Acknowledgements

First and foremost, I would like to thank my mentor, Dr. John V. Moran. John's scientific rigor, along with his passion and enthusiasm for science, is infectious. I feel fortunate to have been trained by a scientist that is truly dedicated to the success of his students. I appreciate John's patience and support throughout the PhD process. I owe a great deal of any scientific accomplishment to John.

I want to specifically thank all members of the Moran lab, past and present, for creating a collaborative work environment. To Dr. John Moldovan, Dr. Peter Larson, and Naveen Jasti, you've been incredible colleagues and friends. I would also like to thank everyone in the Department of Human Genetics, the Genetics Training Program (GTP), the Genome Science Training Program (GSTP), the Rackham Merit Fellowship (RMF) program, and the entire University of Michigan community for creating an exceptional learning environment.

Additionally, I would like to thank my thesis committee, Dr. Sundeep Kalantry, Dr. Jacob Kitzman, Dr. Kaushik Ragunathan, and Dr. JoAnn Sekiguchi for their guidance and expertise that shaped my growth as a scientist. I appreciate the invaluable feedback and lengthy discussions that drove this project forward. Thank you for always challenging me to think deeper.

Finally, I would like to thank my family and friends for their endless love and support. To my wife, Alyssa, and daughter, Eleanor, you have been my greatest inspiration and I love you more than you will ever know.

Table of Contents

List of Figures

List of Tables

Abstract

Transposable elements are DNA sequences that can mobilize to new genomic locations and have had a profound impact on human genome evolution. Although ~45% of the human genome consists of transposable element-derived sequences, Long INterspersed Element-1 (LINE-1 or L1) is the only active transposable element in the human genome. L1s can mobilize (i.e., retrotranspose) in the germline, during early development, and in select somatic cells and the resultant retrotransposition events can alter gene expression, generate structural variation, and create pathogenic mutations. Given the mutagenic potential of L1 retrotransposition, it is no surprise that a variety of cellular mechanisms have evolved to restrict unabated L1 retrotransposition.

Previous studies revealed that reporter genes integrated into the genome of human embryonic carcinoma cells (hECs) by L1 retrotransposition are efficiently silenced by a process that we have termed L1-REPEL (L1-delivered REPorter gEne siLencing). L1-REPEL is mitotically stable, reversible, and correlates with changes in chromatin status at the L1 integration site, suggesting an epigenetic mechanism that requires both initiation and maintenance phases. L1-REPEL is specific to the mechanism of L1 genomic integration (target-primed reverse transcription or TPRT), which utilizes both endonuclease and reverse transcriptase enzymatic activities to facilitate retrotransposition. Thus, we hypothesize that cellular factors recognize TPRT intermediates leading to the establishment of an epigenetic mark required for L1-REPEL in hECs.

Here, we designed and implemented a forward genetic screen using a genome-wide CRISPR/Cas9-based system to elucidate cellular factors that mediate L1-REPEL in PA-1 hECs. We identified 20 highly enriched candidate L1-REPEL factors, including our top candidate gene, neurofibromin 2 (*NF2)* – a tumor suppressor gene that encodes the NF2/merlin protein. Comprehensive validation experiments revealed that NF2/merlin expression was necessary for efficient L1-REPEL in PA-1 cells. Additionally, we determined that the expression of NF2/merlin isoform 1 efficiently re-established L1-REPEL in *NF2* knockout cells. We further demonstrated that *NF2* knockout was insufficient to reactivate L1-REPEL in cells containing a previously silenced insertion, suggesting that NF2/merlin is required, either directly or indirectly, to initiate L1-REPEL. Finally, we found that culturing cells in differentiation media further attenuated L1-REPEL, suggesting that *NF2* knockout and cellular differentiation may act independently or, perhaps, synergistically to attenuate L1-REPEL. Thus, our data indicate that NF2/merlin, a tumor suppressor gene implicated in human disease, may also play a role in silencing L1 retrotransposition events during early human development.

Chapter 1

Introduction

Overview

This thesis examines the phenomenon of L1-delivered reporter gene silencing in human embryonic carcinoma cells. Chapter one provides an overview of transposable elements in the human genome and the cellular processes that restrict LINE-1 retrotransposition. Chapter two describes the design and implementation of a genome-wide CRISPR/Cas9-based screen to identify cellular factors that mediate L1-delivered reporter gene silencing. Chapter three details our progress in defining a role for the NF2/merlin protein in L1-delivered reporter gene silencing. Chapter four provides a summary our findings and discusses possible directions for future studies.

Transposable Elements

Transposable elements, also known as “jumping genes” or mobile elements, are DNA sequences that can move from one genomic location to another. In the 1940’s, maize geneticist Barbara McClintock discovered that transposable element activity “controlled” the expression of a pigmentation gene, leading to maize kernel color variegation (McClintock, 1950, 1951). Despite this momentous discovery, transposable elements were disparaged as nonfunctional “junk DNA” for decades (Doolittle and Sapienza, 1980; Ohno, 1972; Orgel and Crick, 1980). In 1983, McClintock’s work on

"controlling" elements and their impact on gene regulation resulted in the Nobel Prize in Physiology or Medicine.

Today, we recognize McClintock's work as foundational to our current understanding of the genome as a dynamic entity. Transposable elements are abundant in both prokaryotic and eukaryotic genomes and represent a major source of intra- and inter-organismal genetic variation. The completion of the human genome draft sequence revealed that transposable element derived sequences constitute at least ~45%, and perhaps as much as ~70%, of human genomic DNA, whereas protein-coding regions only comprise ~1.5% of genomic DNA (de Koning et al., 2011; Lander et al., 2001). Thus, it is no surprise that transposable elements continue to affect intra- and inter-individual human genetic variation (Beck et al., 2011; Cordaux and Batzer, 2009; Hancks et al., 2011; Richardson et al., 2015; Wells and Feschotte, 2020). Transposable elements are separated into two general classes based on their mechanism of mobilization (Finnegan, 1989). Class I retrotransposons mobilize (*i.e.*, retrotranspose) using an RNA intermediate by "copy-and-paste" mechanisms. Class II DNA transposons can mobilize (*i.e.*, transpose) using a DNA intermediate by a "cut-and-paste" mechanism. In aggregate, Class I and Class II elements comprise ~42% and ~3% of human genomic DNA sequences, respectively (Lander et al., 2001).

DNA Transposons

DNA transposons transpose via a DNA intermediate using non-replicative "cut and paste" or replicative "copy and paste" mechanisms (Finnegan, 1989). There are general types of DNA transposons, which can be subclassified based on their particular mechanism of mobility. They include those that mobilize using: (1) a $DD_{35}E$-type

transposase protein (Yuan and Wessler, 2011); (2) a tyrosine recombinase protein (Kojima and Jurka, 2011); (3) rolling-circle replication (Kapitonov and Jurka, 2001; Thomas and Pritham, 2015); and (4) self-synthesis (Feschotte and Pritham, 2005; Kapitonov and Jurka, 2006).

"Cut-and-paste" DNA transposons typically consist of a pair of terminal inverted repeat sequences (TIRs) that surround an open reading frame (ORF) encoding transposase. Transposase is a member of the $DD_{35}E$ superfamily of proteins that generally contain nuclear localization, DNA binding, and DNA cleavage activities (Hickman and Dyda, 2016). Transposase binds to DNA transposon TIR sequences within the nucleus and then catalyzes the excision and subsequent insertion (*i.e.* "cut-and-paste") of the DNA transposon into a new genomic location (Feschotte and Pritham, 2007; Finnegan, 1989; Jurka, 2008; Slotkin and Martienssen, 2007; Wells and Feschotte, 2020). Transposition events are generally flanked by identically sized target site target-site-duplications (TSDs) that range in size from 4-6 bp, which depend on the target site cleavage preference of a particular transposase (Ivics and Izsvak, 2015; Lander et al., 2001; Munoz-Lopez and Garcia-Perez, 2010).

The process of "cut and paste" DNA transposition is non-replicative, meaning that copy number should remain relatively constant in the genome. However, these elements can amplify during DNA synthesis and DNA repair-induced homologous recombination, resulting in their duplication (Fricker and Peters, 2014; Spradling et al., 2011). Although DNA transposons continue to flourish in simple eukaryotic genomes, they can no longer transpose in the human genome (Feschotte and Pritham, 2007; Lander et al., 2001). However, several human genes are believed to have evolved from

DNA transposons, including the recombination-activating genes *RAG1* and *RAG2*. These genes encode V(D)J recombinase enzymes, which are vital to the maturation of lymphocytes and the adaptive immune response (Huang et al., 2016; Jones and Gellert, 2004; Kapitonov and Jurka, 2005; Oettinger et al., 1990; Thompson, 1995).

DNA transposons and their encoded transposase have been used as tools in both gene discovery and gene therapy. (Ivics et al., 1997; Kebriaei et al., 2017; Moriarity and Largaespada, 2015; Munoz-Lopez and Garcia-Perez, 2010; Sandoval-Villegas et al., 2021). The nature of DNA transposition allows engineered sequences flanked by ITRs to be integrated into genomes by transposase. The Sleeping Beauty transposon system uses a highly active transposase, derived from a reanimated salmonoid DNA transposon (Sleeping Beauty), to mediate gene delivery (Ivics et al., 1997; Kebriaei et al., 2017; Munoz-Lopez and Garcia-Perez, 2010; Sandoval-Villegas et al., 2021). Similarly, the PiggyBac transposon system, derived from the cabbage looper moth, utilizes a hyperactive transposase to deliver large transgenes (Cary et al., 1989; Ding et al., 2005; Kawakami, 2007; Munoz-Lopez and Garcia-Perez, 2010; Sandoval-Villegas et al., 2021). An assay termed transposase-accessible chromatin sequencing (ATAC-seq) utilizes a hyperactive transposase (Tn5) to cuts open or accessible regions of chromatin, enabling the integration of sequencing adapters, which allows the sub-profiling of chromatin accessibility using next generation DNA sequencing technology (Buenrostro et al., 2013). Together, these studies demonstrate the utility of DNA transposons and their potential as molecular tools in biology.

Retrotransposons

Class I retrotransposons replicate (*i.e.*, retrotranspose) through an RNA intermediate (Boeke et al., 1985; Finnegan, 1989), which is reverse transcribed into complementary DNA (cDNA) either before or during its integration into the genome (Beck et al., 2011; Cordaux and Batzer, 2009; Cost et al., 2002; Dombroski et al., 1994; Kazazian and Moran, 2017; Mager and Stoye, 2015; Richardson et al., 2015). Each retrotransposition event results in a copy of the original template donor element at a new genomic location (*i.e.*, "copy-and-paste"). This replicative nature allows the potential for functional (*i.e.*, active) retrotransposons to undergo exponential copy number expansion in the genome. There are two general types of retrotransposons: (1) long-terminal repeat (LTR) containing retroelements (Boeke and Stoye, 1997; Eickbush and Jamburuthugoda, 2008); and (2) non-LTR retroelements (Malik et al., 1999; Xiong and Eickbush, 1988, 1990). Autonomous retrotransposons encode proteins required to mediate their own mobility, whereas nonautonomous elements effectively "hijack" the proteins encoded by structurally related elements to mediate their mobility (Kazazian and Moran, 2017; Richardson et al., 2015).

LTR Retrotransposons

LTR retrotransposons, also known as endogenous retroviruses (ERVs), contain long terminal repeats (LTRs) of variable sizes (ranging from 100 bp to over 5 kb) that flank internal genes that mediate retrotransposition. Functional LTR retrotransposons typically express a polycistronic RNA containing both *gag* and *pol* genes. *Gag* encodes a capsid protein that generates a cytoplasmic virus-like particle. *Pol* encodes protease, reverse transcriptase, integrase, and RNase H enzymatic activities that are required for

converting the retrotransposon RNA into a double strand cDNA. Briefly, the LTR retrotransposon RNA is packaged within cytoplasmic virus-like particles, where the 3' end of a host tRNA binds to complementary sequences within the retrotransposon RNA to initiate (-) strand cDNA synthesis. The completion of double stranded cDNA synthesis occurs via a template switching mechanism and the resultant double stranded cDNA is integrated into the genome using the element encoded integrase activity by a mechanism similar to that used by class II transposases, resulting in the generation of 4-6 bp TSDs that flank the new retrotransposon insertion (Boeke and Stoye, 1997; Telesnitsky and Goff, 1997).

Retroviruses are similar to LTR retrotransposons in both their structure and mobility mechanism, but have a functional envelope (*env*) gene. In general, the envelope protein allows the retrovirus to exit the host cell. Retroviruses that lose their functional *env* gene become endogenized (*i.e.,* endogenous retroviruses), such as mouse intracisternal A particles (IAPs) and MusD elements (Magiorkinis et al., 2012; Ribet et al., 2008).

Human-specific ERVs (HERVs)-derived sequences from both autonomous and non-autonomous elements comprise ~8% of the human genome, but to date, no autonomously active HERVs have been identified in the human genome (Garcia-Montojo et al., 2018; Lander et al., 2001). However, some HERV-K sequences (where "K" indicates the lysine tRNA that is required to initiate [-] strand cDNA synthesis) are polymorphic with respect to presence/absence in the human population, suggesting that they were autonomously mobile since the divergence of humans and chimpanzees around six million years ago (MYA) (Belshaw et al., 2005; Mager and Stoye, 2015; Medstrand and Mager, 1998; Moyes et al., 2007).

Despite being inactive, it is clear that sequences derived from HERVs have been co-opted by the host to play roles in gene regulation. For example, HERV-derived sequences have evolved to act as *cis*-acting sequences that orchestrate a transcriptional network mediating the interferon response (Chuong et al., 2016). Similarly, the expression of HERV-derived proteins can influence embryonic development (Dupressoir et al., 2012; Grow et al., 2015).

Non-LTR retrotransposons

Autonomous non-LTR retrotransposons contain one or two open reading frames (ORFs) followed by a 3' poly(A) sequence (Richardson et al., 2015). One of the ORFs encodes endonuclease and reverse transcriptase activities that are critical for retrotransposition, which occurs by a mechanism termed target-primed reverse transcription (TPRT) (discussed below) (Feng et al., 1996; Luan et al., 1993; Moran et al., 1996; Xiong and Eickbush, 1990).

Autonomous non-LTR retrotransposons are the only active transposable elements in the human genome and their sequences account at least 17% of human genomic DNA (Cordaux and Batzer, 2009; Lander et al., 2001). Non-autonomous non-LTR retrotransposons do not encode proteins; thus, they must "hijack" the protein machinery encoded by related autonomous non-LTR retrotransposons, in *trans*, to mediate their retrotransposition (Dewannieux and Heidmann, 2005). Non-autonomous non-LTR retrotransposons comprise at least 11% of the human genomic DNA (Lander et al., 2001) and include Short INterspersed Elements (SINEs) (Deininger et al., 1981; Dewannieux et al., 2003; Smit and Riggs, 1995) and SINE-R/VNTR/Alu elements (SVAs) (Hancks et al., 2011; Ostertag et al., 2003; Raiz et al., 2012). Other cellular

RNAs, such as U6 spliceosomal RNA, U3 small nucleolar RNA, and messenger RNAs also can be retrotransposed by proteins encoded by autonomous non-LTR retrotransposons, with the latter leading to the formation of processed pseudogenes (Buzdin et al., 2002; Esnault et al., 2000; Garcia-Perez et al., 2007a; Gilbert et al., 2005; Moldovan et al., 2019; Weber, 2006; Wei et al., 2001).

Human L1 Retrotransposons

Long INterspersed Element-1 (LINE-1 or L1) sequences began amplifying prior to the eutherian-marsupial split approximately 150 MYA (Lander et al., 2001; Scott et al., 1987; Smit, 1996). L1 is the only active autonomous retrotransposon in the human genome. L1-derived sequences account for ~17% of human DNA and are present at greater than 500,000 copies per haploid genome (Lander et al., 2001). Most L1 sequences are inactive due to 5' truncation mutations, internal structural rearrangements, and/or point mutations within the ORFs (Beck et al., 2011; Grimaldi et al., 1984; Kazazian and Moran, 1998; Lander et al., 2001). However, the average human genome contains at least 80-100 L1s capable of retrotransposition (Brouha et al., 2003; Moran et al., 1996; Sassaman et al., 1997).

L1 Structure

Full-length retrotransposition-competent human L1s are ~6 kilobases in length (Dombroski et al., 1991; Scott et al., 1987) and consist of a 5' untranslated region (UTR) containing an internal RNA polymerase II promoter (Speek, 2001; Swergold, 1990), two ORFs that are separated by a 63 bp inter-ORF region that contains an in-frame stop codon (Alisch et al., 2006), and a 3'UTR ending in a poly-(A) tract (Doucet et al., 2015; Grimaldi et al., 1984; Scott et al., 1987) (Figure 1.1). Genomic L1s also typically exhibit

characteristic ~7-20 bp TSDs that flank TPRT-mediated non-LTR retrotransposition events (Gilbert et al., 2005; Gilbert et al., 2002; Symer et al., 2002) (Figure 1: TSDs).

The first L1 open reading frame (ORF1) encodes a ~40 kD protein (ORF1p) with nucleic acid binding (Hohjoh and Singer, 1996, 1997; Holmes et al., 1992; Martin, 1991; Martin and Branciforte, 1993) and nucleic acid chaperone activities (Figure 1.1) (Khazina et al., 2011; Khazina and Weichenrieder, 2009; Martin and Bushman, 2001; Naufer et al., 2016) (Figure 1.1). L1 ORF2 encodes a ~150 kD protein (ORF2p), which exhibits both endonuclease (EN) (Cost and Boeke, 1998; Cost et al., 2001; Feng et al., 1996) and reverse transcriptase (RT) activities (Figure 1.1) (Dombroski et al., 1994; Hattori et al., 1986; Mathias et al., 1991). Both ORF1p and ORF2p are necessary for L1 retrotransposition in cultured human cells (Moran et al., 1996).

L1 5' UTR

The L1 5'UTR is ~910 bp and contains several transcription factor DNA binding sites that influence L1 transcription (Athanikar et al., 2004; Becker et al., 1993; Minakami et al., 1992; Swergold, 1990; Tchenio et al., 2000; Yang et al., 2003). The human L1 5'UTR exhibits both sense (Swergold, 1990) and antisense (Speek, 2001) RNA polymerase II activity. The sense promoter directs transcription of full-length L1 RNA (Swergold, 1990). The weaker antisense promoter lacks a clear role in L1 retrotransposition, but can drive the transcription of neighboring genes in human cells (Macia et al., 2011; Matlik et al., 2006; Nigumann et al., 2002; Speek, 2001). The antisense promoter also encodes an open reading frame (ORF0) with unknown function (Denli et al., 2015).

There are several transcription factor binding sites within the L1 5'UTR. A YY1 transcription factor binding site is located near the beginning of the element (+13 to +20) and can position L1 sense strand transcriptional initiation at or near the first base of a full-length element (Athanikar et al., 2004; Becker et al., 1993). There are at least two RUNX3 binding sites (+90 and +510) that are thought to influence both sense and antisense transcription (Yang et al., 2003). Additionally, multiple SRY-like binding sites can influence promoter activity and retrotransposition (Muotri et al., 2005; Tchenio et al., 2000). Recently, TP53 also was demonstrated to bind the 5'UTR of L1 in human cells and mediate transcriptional repression (Tiwari et al., 2020).

L1 ORF1

Following the 5'UTR is a ~1,017 bp open reading frame, ORF1. ORF1 encodes a 40 kD protein (ORF1p) with three well-defined domains (Callahan et al., 2012; Hohjoh and Singer, 1996, 1997; Holmes et al., 1992; Khazina et al., 2011; Martin, 1991, 2006; Naufer et al., 2016). The amino-terminus of ORF1p contains a coiled-coil domain necessary for ORF1p trimerization and retrotransposition (Basame et al., 2006; Doucet et al., 2010; Khazina et al., 2011; Khazina and Weichenrieder, 2009, 2018; Martin et al., 2003). Following the coiled-coil domain is the RNA recognition motif (RRM) domain (Khazina et al., 2011; Khazina and Weichenrieder, 2009) and a carboxyl-terminal domain (CTD), which are important for nucleic acid binding and chaperone activity (Januszyk et al., 2007; Martin, 2010; Martin and Bushman, 2001). Conserved residues within the RRM and CTD domains are required for L1 retrotransposition in a cell culture-based assay (Doucet et al., 2010; Khazina and Weichenrieder, 2009; Moran et al., 1996).

ORF1p binds back to its encoding RNA in *cis*, contributing to L1-ribonucleoprotein particle (L1-RNP) formation (Doucet et al., 2010; Hohjoh and Singer, 1996, 1997; Kulpa and Moran, 2005; Martin, 1991; Wei et al., 2001). Binding of ORF1p facilitates nucleic acid remodeling, which is thought to be necessary for L1 retrotransposition (Martin, 2010). In addition to single-stranded RNA, ORF1p also can bind unstructured single-strand and double-stranded DNA (Callahan et al., 2012; Khazina and Weichenrieder, 2009; Martin and Bushman, 2001; Martin et al., 2005). Phosphorylation of ORF1p at proline-directed protein kinase target sites is essential for L1 retrotransposition (Cook et al., 2015).

Other non-LTR retrotransposons encoding an ORF1p, such as the zebrafish LINE (Zfl2-1), do not require ORF1p for retrotransposition (Kajikawa et al., 2012; Nakamura et al., 2012). Similarly, Alu elements, the predominant SINE in the human genome, do not require ORF1p in order to retrotranspose (Dewannieux et al., 2003; Wallace et al., 2008). Thus, why ORF1p is essential for L1 retrotransposition remains unclear. It is possible that L1 ORF1p may protect L1 mRNA from degradation, facilitate the import of L1 RNPs into the nucleus through an uncharacterized process, and/or may play an active role during TPRT. However, additional studies are required to assess the role of ORF1p in specific steps of the L1 retrotransposition pathway.

L1 ORF2

Following ORF1 is a 63 bp inter-ORF spacer that contains an in-frame stop codon before the second open reading frame (ORF2). ORF2 is ~3,828 bp and encodes a ~150 kD protein (ORF2p) with endonuclease and reverse transcriptase activities, which are required for L1 retrotransposition (Cost et al., 2002; Dombroski et al., 1994; Ergun et al.,

2004; Feng et al., 1996; Mathias et al., 1991; Moran et al., 1996). ORF2p is translated from a bicistronic RNA using an unconventional ribosomal termination/reinitiation mechanism (Alisch et al., 2006). ORF2p has an endonuclease (EN) domain, which resembles that of an apurinic/apyrimidinic (AP) endonuclease (Feng et al., 1996; Martin et al., 1995), that can generate a single-strand endonucleolytic nick at the preferred 5'-TTTTT/AA-3' genomic DNA cleavage site (Feng et al., 1996; Flasch et al., 2019; Jurka, 1997; Morrish et al., 2002). ORF2p-mediated cleavage of genomic DNA liberates a 3'-hydroxyl group and a 5'- monophosphate (Feng et al., 1996).

Downstream of the EN domain is the ORF2p reverse transcriptase (RT) domain. The ORF2p RT domain is similar to that of telomerase, LTR-retrotransposons, and group II introns (Eickbush, 1997; Hattori et al., 1986; Kopera et al., 2011; Malik et al., 1999; Xiong and Eickbush, 1990). Purified recombinant ORF2p has both RNA-dependent and DNA-dependent DNA polymerase activities (Cost et al., 2002; Piskareva et al., 2003; Piskareva and Schmatchenko, 2006). Similar to ORF1p, ORF2p preferentially binds back to its encoding RNA in *cis*, contributing to L1-RNP formation (Doucet et al., 2010; Doucet et al., 2015; Kulpa and Moran, 2005, 2006). Furthermore, isolated L1-RNPs can reverse transcribe L1 RNA *in vitro*, demonstrating the presence of ORF2p RT activity within the L1-RNPs (Doucet et al., 2010; Kopera et al., 2016a; Kulpa and Moran, 2006).

A cysteine-rich (C) domain with unknown function resides within the carboxyl-terminus of ORF2p (Fanning and Singer, 1987; Moran et al., 1996). Mutations within the ORF2p C-domain hinder L1 retrotransposition efficiency in human cells (Clements and Singer, 1998; Doucet et al., 2010; Moran et al., 1996). A study using a synthetic codon

optimized L1 (ORFeus-Hs) demonstrated that replacement of the native ORF2p C-domain with ORFeus-HS sequence increased ORF2p activity, suggesting more efficient translation of ORF2p from the chimeric construct (An et al., 2011). Therefore, sequences within the region of L1 RNA encoding the C-domain may influence ORF2p translational efficiency, perhaps by generating RNA structures that impede ribosome processivity.

L1 3'UTR

The L1 3'UTR is ~206 bp and contains a evolutionarily conserved polypurine tract, which is dispensable for L1 retrotransposition in human cultured cells (Moran et al., 1996), as well as a functional RNA polymerase II polyadenylation (poly(A)) signal (Dombroski et al., 1991; Lander et al., 2001). The polypurine tract is predicted to fold into a G-quadruplex structure (Howell and Usdin, 1997; Sahakyan et al., 2017; Usdin and Furano, 1989). Recently, the gamma interferon inhibitor of translation (GAIT) complex was demonstrated to restrict L1 retrotransposition in human cells (Ward et al., 2017). The GAIT complex binds structured RNA stem-loops in the 3'UTR of interferon-related mRNAs (Arif et al., 2018). Thus, the conserved polypurine tract within the L1 3'UTR may interact with host factors that modulate ORF2p translation.

The L1 3'UTR also contains a weak RNA polymerase II poly(A) signal. Transcriptional bypassing of the L1 poly(A) signal and using a fortuitous poly(A) signal present in 3' flanking genomic DNA can result in the formation L1-mediated 3' transductions upon retrotransposition (Holmes et al., 1994; Moran et al., 1999; Moran et al., 1996). Replacement of the L1 poly(A) tail with a stable RNA structure that lacks a poly(A) tract (*i.e.*, the 3' end of Metastasis Associated Lung Adenocarcinoma Transcript

1 [MALAT 1] non-coding RNA) allows for the production of a stable L1 RNA and efficient ORF2p translation, but prevents L1 retrotransposition. Thus, the L1 poly(A) tract is necessary for ORF2p to bind, either directly or indirectly, to its encoding L1 RNA in *cis* (Doucet et al., 2015) to mediate retrotransposition.

L1 Retrotransposition Cycle

Retrotransposition requires reverse transcription of L1 RNA into cDNA and integration of the copied element into the genome. The L1 retrotransposition cycle commences with transcription of L1 RNA from an internal promoter located within the 5'UTR of a full-length genomic L1 (Figure 1.2). The bicistronic polyadenylated mRNA is exported to the cytoplasm, where it undergoes translation (Figure 1.2). The L1-encoded proteins ORF1p and ORF2p exert *cis*-preference and preferentially bind their encoding RNA (Esnault et al., 2000; Wei et al., 2001), forming L1-RNPs (Hohjoh and Singer, 1996; Kulpa and Moran, 2005, 2006; Martin, 1991) (Figure 1.2). Components of the L1-RNP enter the nucleus by a process that may or may not require cell division (Kubo et al., 2006; Mita et al., 2018) (Figure 1.2). Upon nuclear entry, L1 undergoes genomic integration by target-primed reverse transcription (TPRT), which minimally requires ORF2p and L1 RNA (Figure 1.2).

During TPRT, the ORF2p EN makes a single-strand endonucleolytic nick of chromosomal DNA at the EN consensus cleavage site 5'-TTTTT/AA-3' (Cost and Boeke, 1998; Cost et al., 2001; Feng et al., 1996; Flasch et al., 2019; Gilbert et al., 2002; Morrish et al., 2002; Symer et al., 2002; Szak et al., 2002). ORF2p-mediated cleavage liberates a 3' hydroxyl group that can be used as a primer by ORF2p RT to initiate L1 cDNA synthesis (Cost et al., 2002; Feng et al., 1996; Kulpa and Moran, 2006;

Luan et al., 1993). Although the downstream steps of TPRT are not completely understood, a second endonucleolytic nick likely generates a 3' hydroxyl group that primes DNA-dependent DNA synthesis of second-strand L1 cDNA by the L1 RT using the (-) strand L1 cDNA as a template (Christensen and Eickbush, 2005). The completion of TPRT likely requires cellular factors that mediate DNA repair. As a result of TPRT, L1s typically are flanked by variable-length TSDs that range from ~7-20 bp in length (Figure 1.2) (Gilbert et al., 2005; Gilbert et al., 2002; Symer et al., 2002).

ORF2p can be used to retrotranspose other cellular RNAs in *trans* (Ahl et al., 2015; Doucet et al., 2015; Wei et al., 2001). This process, termed *trans*-complementation, is required to mediate the retrotransposition of non-autonomous SINEs, including Alu (Deininger et al., 1981; Dewannieux et al., 2003) and SVA elements (Hancks et al., 2011; Ostertag et al., 2003; Raiz et al., 2012). With over one million copies, Alu is the most abundant transposable element in the human genome, comprising ~11% of human DNA (Lander et al., 2001). Alu elements are short (~300 bp) RNA polymerase III transcribed sequences derived from 7SL RNA, which is part of the signal recognition particle (SRP) that recognizes proteins destined for co-translational import into the endoplasmic reticulum (Ahl et al., 2015; Bennett et al., 2008; Chu et al., 1995; Deininger et al., 1981; Kriegs et al., 2007; Sinnett et al., 1991; Ullu and Tschudi, 1984; Weichenrieder et al., 2001). Thus, localization to the ribosome may allow Alu to efficiently steal ORF2p during translational elongation arrest (Ahl et al., 2015; Doucet et al., 2015; Doudna and Rath, 2002; Walter and Blobel, 1983). ORF2p also can retrotranspose other cellular mRNAs, leading to the formation of processed pseudogenes (Buzdin et al., 2002; Esnault et al., 2000; Garcia-Perez et al., 2007a;

Gilbert et al., 2005; Moldovan et al., 2019; Wei et al., 2001). Thus, in aggregate, L1-mediated retrotransposition events are responsible one-third of the human genome (Lander et al., 2001; Richardson et al., 2015).

L1s containing missense mutations in the active sites of the L1 EN domain can undergo endonuclease-independent (ENi) retrotransposition in cultured XRCC4- and XRCC7-deficient Chinese hamster ovary (CHO) cells. Unlike canonical TPRT-mediated insertions, ENi L1 retrotransposition events typically lack TSDs and exhibit genomic deletions and other rearrangements at the site of L1 integration (Morrish et al., 2002). ENi L1 retrotransposition is thought to utilize genomic lesions (*e.g.*, DNA lesions and dysfunctional telomeres) to prime ORF2p RT activity (Kopera et al., 2011; Morrish et al., 2007; Morrish et al., 2002). Recently, intermediates of DNA replication (*i.e.*, lagging strand 3'-hydroxyl groups) were proposed to facilitate ENi L1 retrotransposition (Flasch et al., 2019). The above data suggest that ENi may be an ancestral mechanism of retrotransposition prior to the acquisition of a functional L1 ORF2p EN domain (Flasch et al., 2019).

A Cell Culture Assay for L1 Retrotransposition

The L1 retrotransposition assay (Figure 1.3) (Moran et al., 1996) was developed based on earlier assays to detect *Saccharomyces cerevisiae* Ty1 and mouse IAP retrotransposition events (Boeke et al., 1985; Garfinkel et al., 1989; Heidmann and Heidmann, 1991). In the L1 retrotransposition assay, an engineered L1 containing a retrotransposition reporter cassette is transiently transfected into cultured cells. Each retrotransposition reporter cassette is cloned into the LINE-1 3' UTR to not disrupt ORF1p and ORF2p expression (Figure 1.3). The retrotransposition reporter cassettes

contain an antisense selectable (*neomycin* or *blasticidin*) or screenable (*EGFP*) reporter gene whose transcription is driven by an exogenous promoter and terminates at a heterologous polyadenylation signal (Freeman et al., 1994; Moran et al., 1996; Morrish et al., 2002; Ostertag et al., 2000). The reporter gene is interrupted by an intron cloned in the same transcriptional orientation as the L1 sequence. This arrangement ensures that a functional reporter gene is only expressed after successful round of retrotransposition, where the engineered L1 is transcribed from an episomal expression vector (pCEP4), the intron is spliced from sense strand L1 mRNA containing the "backward" reporter gene, and L1 is integrated into genomic DNA by TPRT (Figure 1.3) (Kopera et al., 2016b; Moran et al., 1996; Ostertag et al., 2000; Wei et al., 2001). Quantification of retrotransposed reporter gene expression (*e.g.,* using a focus formation assay or flow cytometry) then can be used to quantify the L1 retrotransposition efficiency (Figure 1.3) (Kopera et al., 2016b; Moran et al., 1996; Ostertag et al., 2000; Wei et al., 2001).

Consequences of L1 Retrotransposition

Disease

In 1988, *de novo* L1 retrotransposition was demonstrated to cause hemophilia A (Kazazian et al., 1988). Kazazian and colleagues conducted a screen in 240 male patients with X-linked hemophilia A. Two unrelated patients contained mutagenetic truncated L1 insertions disrupting the factor VIII gene. Detailed characterization of one mutation revealed that the L1 insertion in the factor VIII gene was absent in both parents and was not somatic mosaic in the patient, suggesting that a *de novo* L1

retrotransposition occurred in the mother's germline or during early embryonic development.

The full-length L1 responsible for the mutagenic insertion in one patient (L1.2) was subsequently identified and demonstrated to be capable of retrotransposition in cultured mammalian cells (Dombroski et al., 1991; Moran et al., 1996). Since this time, over 130 pathogenic L1-mediated (*i.e.*, L1, Alu, SVA, or processed pseudogene) retrotransposition events have been implicated in human disease (Hancks and Kazazian, 2016; Kazazian and Moran, 2017). Indeed, L1-mediated retrotransposition events are estimated to account for 1 in 250 disease-causing mutations in humans (Wimmer et al., 2011).

L1-mediated retrotransposition has also been implicated in the etiology of various cancers (Burns, 2017, 2020; Scott and Devine, 2017). In 1992, the first documented somatic L1 retrotransposition event was discovered in a patient with colorectal cancer (Miki et al., 1992). This 750 bp insertion was located within the *APC* tumor suppressor gene and exhibited characteristics of TPRT-mediated integration (*i.e.*, it ended in a 3' poly (A) tail followed by a short 3' transduction and was flanked by variable length TSDs) (Miki et al., 1992; Moran, 1999). Subsequent studies demonstrated that somatic L1 retrotransposition predominantly occurs in human epithelial cell derived cancers (Doucet-O'Hare et al., 2015; Doucet-O'Hare et al., 2016; Ewing et al., 2015; Iskow et al., 2010; Scott and Devine, 2017) and that the rate of L1 retrotransposition is highly variable among tumors (Doucet-O'Hare et al., 2015; Lee et al., 2012; Rodic et al., 2015; Tubio et al., 2014). Furthermore, hypomethylation of full-length genomic L1s also was common in malignant cells, suggesting that L1 expression can become derepressed in

tumor cells (Burns, 2017; Iskow et al., 2010; Scott and Devine, 2017). Thus, L1-mediated retrotransposition may produce driver or passenger mutations in cancers. Future studies are necessary to determine the extent to which L1 driver mutations play a role in promoting tumorigenesis and/or tumor progression.

Structural Variation

L1-mediated retrotransposition events can lead to genomic structural variation (Beck et al., 2011). Structural analysis of engineered L1 retrotransposition events in somatic cancer cell lines demonstrated that DNA recombination processes such as single-strand annealing (SSA), synthesis-dependent strand annealing (SDSA), or nonhomologous end joining (NHEJ) may generate chromosomal rearrangements such as deletions, duplications, and perhaps, translocations upon retrotransposition (Beck et al., 2011; Gilbert et al., 2005; Gilbert et al., 2002; Symer et al., 2002). Moreover, comparative analyses between the human and chimpanzee genomes revealed nonallelic homologous event recombination events between L1 or Alu sequences can lead to human-specific structural variation either during or after L1 integration (Callinan et al., 2005; Han et al., 2008; Han et al., 2005; Hayakawa et al., 2001).

Aberrant L1 transcription can result in the addition of 5' or 3' genomic DNA to the L1 mRNA transcript. Subsequent retrotransposition of these mRNAs can result in L1-mediated transductions (Beck et al., 2011; Richardson et al., 2015). L1 5'-transductions are rare given that most L1s are 5' truncated. However, approximately 1 in 5 human-specific L1s contain 3' transductions, which likely is due to the ability of RNA polymerase II to bypass the weak L1 poly(A) signal (Beck et al., 2010; Goodier et al., 2000; Holmes et al., 1994; Kidd et al., 2010; Moran, 1999; Moran et al., 1999; Pickeral

et al., 2000; Tubio et al., 2014). Consequently, L1-mediated transductions can serve as molecular tags to identify actively expanding L1 subfamilies as well as tools to infer progenitor/offspring L1s relationships (Beck et al., 2010). Indeed, LRE2, LRE3, and $L1_{RP}$ are disease-producing retrotransposition-competent L1s that each contain L1-mediated 3'-transductions (Brouha et al., 2002; Holmes et al., 1994; Kimberland et al., 1999; Tubio et al., 2014).

Gene Evolution

L1-mediated 3'-transduction can provide a vehicle to promote exon shuffling and possibly the formation of new genes (Moran, 1999; Moran et al., 1999; Xing et al., 2006). L1 retrotransposition assays in cultured human cells provided proof-of-principle evidence that L1 can transduce 3' DNA sequences to new genomic locations, thereby offering the potential to create novel cellular genes (Moran et al., 1999). Similarly, L1 retrotransposition into a cellular gene can generate cryptic poly(A) or cryptic splicing sequences, resulting in the premature polyadenylation or mis-splicing of a genic transcript (Han et al., 2004; Perepelitsa-Belancio and Deininger, 2003). Moreover, mobilization of cellular mRNAs in *trans* can result in expressed fusion genes such as TRIM/Cyp, which arose by retrotransposition of the *CypA* gene into the 3'UTR of the macaque *TRIM5* gene (Liao et al., 2007; Virgen et al., 2008). Interestingly, TRIM5α was recently reported to recognize cytoplasmic L1-RNPs, leading to the induction of innate immune signaling and restriction of L1 retrotransposition (Volkmann et al., 2020). In addition to L1, other transposable elements have been co-opted for gene regulation (Chuong et al., 2017). For example, ERVs contain interferon-inducible enhancers that facilitate the mammalian innate immune response (Chuong et al., 2016). Thus,

transposable elements may serve as a system to distribute regulatory sequences to new genomic locations, supporting the prescient hypothesis that sequences within mobile elements can be co-opted by the host over evolutionary time to regulate gene expression and create transcriptional networks (Davidson and Britten, 1979).

L1 Retrotransposition During Human Development

Germline and Early Development

In order to survive, L1s must retrotranspose in cells that contribute genetic material to future generations such as germ cells or during early embryonic development. Pedigree-based estimations of germline retrotransposition rates using whole-genome sequencing suggests that 1 in 63 human births contain a *de novo* L1 retrotransposition event (Feusier et al., 2019), which is consistent with previous estimates (1:20-1:200 births) (Kazazian, 1999; Xing et al., 2009). Similarly, 1 in 63 births contained a SVA retrotransposition event, whereas 1 in 40 contained an Alu retrotransposition event (Feusier et al., 2019). Interestingly, *de novo* Alu retrotransposition events exhibited a paternal bias (Feusier et al., 2019).

An example of a heritable L1 retrotransposition event was discovered in the *CMH* gene of a male patient with X-linked choroideremia, a form of hereditary retinal degeneration (van den Hurk et al., 2003). Further work revealed that the mother of the patient was both a somatic and germline mosaic with respect to this L1 insertion, suggesting that L1 retrotransposition event occurred post-zygotically during early embryonic development in the mother of the patient (van den Hurk et al., 2007).

Experiments in mice demonstrated that L1 can retrotranspose in the germline and during early embryonic development (Ostertag et al., 2002; Richardson et al., 2017). An

engineered L1 transgene containing an *EGFP* retrotransposition indicator cassette was shown to retrotranspose in the germ cells of male and female transgenic mice (Ostertag et al., 2002). Similarly, pedigree-based analyses using mouse retrotransposon capture and whole-genome sequencing revealed an endogenous L1 retrotransposition rate of ~1 in 8 births (Richardson et al., 2017). Heritable insertions were identified in both primordial germ cells and pluripotent embryonic cells (Richardson et al., 2017). Moreover, 1 in 3 transgenic mice harboring a synthetic mouse L1 (ORFeus) exhibited a germline L1 retrotransposition event (An et al., 2006; Han and Boeke, 2004).

In humans, germline L1 expression has been observed in both testis (Ergun et al., 2004) and oocytes (Georgiou et al., 2009). Cell culture experiments demonstrated high levels of endogenous L1 expression in human embryonic stem cells (hESCs) (Garcia-Perez et al., 2007b) and human embryonic carcinoma cells (hECs) (Garcia-Perez et al., 2010; Skowronski et al., 1988; Skowronski and Singer, 1985), along with high levels of engineered L1 retrotransposition (Garcia-Perez et al., 2007b; Garcia-Perez et al., 2010). Finally, *de novo* L1 retrotransposition events were identified in both hESCs (Wissing et al., 2011) and induced pluripotent stem cells (iPSCs) (Klawitter et al., 2016; Wissing et al., 2012). Thus, these data suggest that both endogenous (native) and engineered L1s can retrotranspose in germ cells and during early embryogenesis.

Somatic Retrotransposition

Somatic L1 retrotransposition events that occur after the formation of the germline cannot be transmitted to future generations. However, Muotri and colleagues found that engineered human L1 retrotransposition events could be detected in cultured rat neural precursor cells (NPCs) as well as neurons in the brains of transgenic mice (Muotri et al.,

2005). Subsequent studies revealed that cultured human neural precursor cells (NPCs) also supported engineered human L1 retrotransposition (Coufal et al., 2009). Furthermore, an increase in endogenous L1 DNA copy number was detected by qPCR in human hippocampal samples compared to matched heart or liver samples (Coufal et al., 2009). This increase in copy number could be due to a number of factors, including increases in neuronal aneuploidy, the generation of single-strand L1 cDNAs, and/or increases in L1 retrotransposition (Zhu et al., 2021). In addition to NPCs, nondividing human neurons also appear to support somatic engineered L1 retrotransposition (Macia et al., 2017). Thus, these studies suggest that L1 retrotransposition can generate considerable somatic mosaicism in the brain.

Somatic variation is difficult to detect given that the mutation is only present in a subset, or even a single, cell. Advancements in DNA sequencing technologies, such as whole genome single-cell sequencing, provides an avenue to identify rare somatic variants. Current estimates of endogenous L1 retrotransposition rates in human neuronal cells range from 0.32 to 13.7 L1 insertions per neuron (Faulkner and Garcia-Perez, 2017). A study using single-cell retrotransposon capture sequencing (RC-seq) estimated a rate of ~13.7 somatic L1 insertions per hippocampal neuron (Upton et al., 2015). A conflicting estimate of ~0.32 somatic L1 insertions per neuron was reported using single-cell whole genome sequencing (Evrony et al., 2015). Lastly, single-cell sequencing using a machine learning approach estimated that ~0.58 of neurons contained a somatic L1-associated variant (SLAV) (Erwin et al., 2016). Although L1 retrotransposition rates in the brain remain hotly debated, an extrapolation of these rates to the ~80 billion neurons in the human brain suggests that L1 retrotransposition

can perhaps generate millions of somatic mosaic L1 insertions in the brains of healthy individuals, leading to intra-individual genetic variation.

Neuropsychiatric disorders such as schizophrenia (Bundo et al., 2014; Doyle et al., 2017) and neurodevelopmental disorders like Rett syndrome (Muotri et al., 2010; Zhao et al., 2019) and autism (Shpyleva et al., 2018; Tangsuwansri et al., 2018) have been associated with L1 expression (Terry and Devine, 2019). The Brain Somatic Mosaicism Network (BSMN) was founded to investigate the role of somatic genomic variation in the development of mental illness (McConnell et al., 2017). Recent work by the BSMN has improved the methodology to detect rare somatic variation (Rodin et al., 2021; Wang et al., 2021; Zhou et al., 2020; Zhu et al., 2021). Future studies will develop and refine tools that will help to elucidate the consequences of neurological somatic mosaicism in human health and disease.

Cellular Restriction of L1 Retrotransposition

The vast majority of L1 retrotransposition events are inactive due to 5' truncation and/or structural rearrangements that are generated upon TPRT-mediated retrotransposition (Beck et al., 2011; Kazazian and Moran, 1998; Lander et al., 2001; Ostertag and Kazazian, 2001; Richardson et al., 2015). Most of the remaining insertions are eventually disabled due to mutations that accumulate during evolutionary time (Beck et al., 2011; Kazazian and Moran, 2017; Richardson et al., 2015). However, each human genome contains ~80-100 active L1 sequences capable of retrotransposition (Brouha et al., 2003; Moran et al., 1996; Sassaman et al., 1997). Given the mutagenic potential of L1 retrotransposition, it is evident that the host has evolved numerous restrictive mechanisms to inhibit endogenous L1 retrotransposition and that L1 has

evolved to evade these mechanisms (Figure 1.4) (Goodier, 2016; Levin and Moran, 2011).

Transcriptional Regulation

Transcriptional regulatory sequences (*i.e.,* promoters and enhancers) are necessary for L1 expression and its subsequent retrotransposition. Various epigenetic silencing mechanisms have evolved to restrain L1 expression. DNA methylation is a heritable epigenetic mark that is established when a methyl group is transferred to DNA (Robertson, 2005). DNA methylation at 5'-CpG residues present in gene promoters typically represses gene transcription (Jones, 2012). Typically, L1 sequences exhibit a high degree of methylated CpG residues in somatic cells, which likely represses their expression (Figure 1.4) (Goll and Bestor, 2005). However, many cancers exhibit decreased CpG methylation within endogenous L1 5'UTRs, which correlates with increased L1 expression (Alves et al., 1996; Bratthauer and Fanning, 1993; Iskow et al., 2010; Suter et al., 2004; Thayer et al., 1993; Tubio et al., 2014).

DNA methyltransferases are cellular proteins that mediate DNA methylation (Chen and Zhang, 2020). Loss of DNMT3L in mice decreased endogenous L1 methylation, increased endogenous L1 expression, and led to failures in spermatogenesis (Bourc'his and Bestor, 2004). Knockout of DNMT3C in mice resulted in a similar increase in L1 expression (Barau et al., 2016). The knockout of both *de novo* (DNMT3A and DNMT3B) and maintenance (DNMT1) methyltransferases in hESCs also increased expression of endogenous L1s (Castro-Diaz et al., 2014). Methyl CpG binding protein 2 (MECP2) recognizes and binds methylated DNA. Expression of MECP2 repressed L1 retrotransposition in HeLa cells (Yu et al., 2001), whereas knockout of MECP2 in

rodents promoted L1 retrotransposition in neurons (Muotri et al., 2010). Together, these studies have established that DNA methylation can actively repress L1 transcription (Figure 1.4).

Post-translational modifications of histone tails such as methylation, acetylation, phosphorylation, and ubiquitination can modulate gene activity (Kornberg and Lorch, 2020; Kouzarides, 2007). A study using a CRISPR/Cas9-based screening strategy to identify factors that restrict L1 retrotransposition in human somatic cancer cells identified MORC2 and components of the human silencing hub (HUSH) complex (Liu et al., 2018). The authors concluded that these proteins bind evolutionarily young (*i.e.*, primate specific) L1s to promote transcriptional silencing through deposition of histone methylation (H3K9me3) (Liu et al., 2018). Similarly, the HUSH complex and KAP1 were required to repress L1 expression in NTERA-2 human embryonic carcinoma cells (Robbez-Masson et al., 2018). Notably, HUSH and KAP1 were required for H3K9me3 maintenance, which mediated L1 repression (Robbez-Masson et al., 2018). Additionally, ChIP-seq experiments identified Suv39h-dependent H3K9me3 at transcriptionally repressed L1s in mouse ESCs (Bulut-Karslioglu et al., 2014). Together, these results suggest that histone modifications, particularly H3K9me3, are necessary to maintain L1 repression (Figure 1.4).

Kruppel-associated box zinc finger proteins (KRAB-ZFPs) can repress transcription by binding directly to DNA or through heterochromatin formation by associating with histone modifying proteins (Iyengar and Farnham, 2011). The KRAB-ZFPs KAP1 and ZNF93 were necessary to transcriptionally repress older L1 subfamilies (L1PA7-L1PA3) in human ESCs (Castro-Diaz et al., 2014; Jacobs et al., 2014). Furthermore, KRAB-

ZFPs are rapidly evolving in mammals, revealing a continuous evolutionary “arms race” between KRAB-ZFPs and L1s (Yang et al., 2017). In sum, the cell has evolved numerous ways to inhibit L1 transcription.

RNA Interference

Small RNA-based interference (RNAi) utilizes short RNA sequences that are loaded onto Argonaute proteins, forming the RNA-induced silencing complex (RISC) (Wilson and Doudna, 2013). The P-element induced wimpy testis (PIWI) proteins are germline-specific Argonaute family members (Wilson and Doudna, 2013). One study demonstrated that the piwi-interacting RNA pathway (piRNA) was required to maintain H3K9me3 and transcriptional repression at endogenous L1s in mouse germ cells (Pezic et al., 2014). Additionally, a similar study in mouse germ cells found that G9a-mediated H3K9me2 repressed endogenous L1s (Di Giacomo et al., 2014).

The microprocessor complex (Drosha/DGCR8), a regulator of microRNA (miRNA) biosynthesis, associates with L1 RNA in human cells and is thought to cleave L1 RNA (Heras et al., 2014; Heras et al., 2013). Subsequent studies demonstrated that miR128 associated with Argonaute to target and degrade L1 RNA (Hamdorf et al., 2015). Recently, the tumor suppressor miRNA let-7 was found to bind L1 mRNA and repress ORF2p translation (Tristan-Ramos et al., 2020). Thus, multiple steps within the miRNA pathway may regulate L1 retrotransposition.

Dicer-dependent small interfering RNAs (siRNAs) have been demonstrated to silence transposons (Castel and Martienssen, 2013). One report suggests that L1-derived siRNAs exist in mouse oocytes (Watanabe et al., 2008). In human cultured cells, expression of sense and antisense L1 transcripts generated siRNAs thought to

target and degrade L1 RNA (Yang and Kazazian, 2006). In sum, the above studies show that multiple small RNA-mediated mechanisms (*i.e.*, piRNA, miRNA, and siRNA) can restrict L1 retrotransposition (Figure 1.4).

Post-Transcriptional Regulation

L1 can undergo tissue-specific post-transcriptional splicing at splice donor and splice acceptor sites within human and mouse L1 RNA (Belancio et al., 2006; Belancio et al., 2008; Belancio et al., 2010). Recently, splicing was demonstrated to inhibit L1 retrotransposition in cultured human cells (Larson et al., 2018). Furthermore, L1 splicing facilitated spliced integrated retrotransposed element (SpIRE) formation in the human genome (Larson et al., 2018). Thus, even if full-length L1 RNA is expressed, tissue-specific splicing of L1 RNA may prevent retrotransposition of a functional L1 (Figure 1.4).

Northern blot analyses revealed that human and mouse L1s transfected into cultured mouse cells exhibited premature polyadenylation (Perepelitsa-Belancio and Deininger, 2003). Removal of the predicted poly(A) sites promoted full-length L1 RNA expression (Perepelitsa-Belancio and Deininger, 2003). The lack of prematurely polyadenylated L1s in the human genome suggests that premature polyadenylation inhibits L1 retrotransposition, possibly due to the absence of functional ORF2p in *cis*.

L1-RNP Formation

Several reports have identified cellular factors that associate with the L1 RNP (Goodier et al., 2013; Moldovan and Moran, 2015; Taylor et al., 2018; Taylor et al., 2016; Taylor et al., 2013). Components of the L1 RNP have been demonstrated to associate with stress granules (SGs), which are cytoplasmic aggregations of proteins

and mRNAs that influence mRNA decay and protein translation (Decker and Parker, 2012; Doucet et al., 2010). The zinc-finger antiviral protein (ZAP) colocalizes with the L1 RNP and stress granules (Goodier et al., 2015; Moldovan and Moran, 2015). Furthermore, overexpression of ZAP inhibits engineered L1 and Alu retrotransposition events in cultured human cells (Goodier et al., 2015; Moldovan and Moran, 2015). ZAP also inhibits Moloney and murine leukemia virus (MMLV) and human immunodeficiency virus type 1 (HIV-1) by initiating degradation of the viral mRNA (Gao et al., 2002; Zhu et al., 2011). Thus, similar mechanisms of action may restrict L1.

Moloney Leukemia Virus 10 protein (MOV10) is an ATP-dependent RNA helicase implicated in L1 restriction. Similar to ZAP, overexpressed Moloney Leukemia Virus 10 protein (MOV10) localizes to cytoplasmic L1 foci and inhibits non-LTR retrotransposition in cultured human cells (Goodier et al., 2012; Moldovan and Moran, 2015). Recently, MOV10 was demonstrated to cooperate with TUT4/TUT7 to facilitate 3' uridylation of L1 RNA, inhibiting L1 retrotransposition (Warkocki et al., 2018). The authors propose a multifaceted mechanism of inhibition, where TUT7-mediated uridylation inhibits the initiation of reverse transcription, whereas TUT4-mediated uridylation acts to destabilize L1 mRNA (Warkocki et al., 2018). Future studies will be important to determine whether cytoplasmic foci formation plays a role in restricting L1 retrotransposition.

SAM domain and HD domain 1 (SAMHD1) is a dNTPase that inhibits viral cDNA synthesis by diminishing the pool of available dNTPs (Lahouassa et al., 2012). SAMHD1 also restricts L1 retrotransposition in human and mouse cultured cells (Zhao et al., 2013). SAMHD1 mutants deficient in dNTPase activity also inhibited L1 retrotransposition, suggesting that SAMHD1 may utilize a different mechanism to

restrict L1 (Zhao et al., 2013). Later studies demonstrated that overexpression of SAMHD1 resulted in colocalization of the L1 RNP and SGs (Hu et al., 2015). Another confounding report implicated SAMHD1 RNase activity in restricting retroviruses (Choi et al., 2015). Thus, cellular antiviral proteins have likely evolved to inhibit L1 and viral proliferation through distinct mechanisms of action (Figure 1.4).

An accumulation of cytoplasmic L1 nucleic acid has been implicated in autoimmune disorders (Thomas et al., 2017). Aicardi-Goutières syndrome (AGS) is a rare disorder that affects the immune system. AGS genes such as TREX1 and SAMHD1 have been associated with restricting L1 retrotransposition. TREX1 is an exonuclease that degrades single-stranded DNA replication intermediates (Yang et al., 2007). Lack of TREX1 results in an accumulation of single-stranded L1 DNA in cultured human and mouse cells (Thomas et al., 2017). Recent reports suggest that the interferon response is activated by single-strand ORF2p-mediated cytoplasmic L1 cDNA (De Cecco et al., 2019) and Alu cDNA (Fukuda et al., 2021). Lastly, the interferon response was demonstrated to promote condensin II and GAIT mediated restriction of L1 retrotransposition in cultured human cells (Ward et al., 2017). Thus, current literature supports a potential role for L1 in autoimmunity.

TPRT Regulation

TPRT-mediated genomic integration is unique to non-LTR retrotransposons. TPRT exploits endonuclease and reverse transcriptase enzymatic activities that may suscept L1 to restriction by the host. APOBEC3 (apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like3, A3) proteins encode cytidine deaminase activity. There are seven APOBEC3 proteins in humans (A3A-D, A3F, A3G and A3H) that can deaminate

single-strand nucleic acid substrates to generate cytidine to uridine mutations (Chiu and Greene, 2008). A3 proteins may also inhibit reverse transcriptase activity (Koito and Ikeda, 2013). Transient expression of A3A and A3B inhibited both L1 and Alu retrotransposition in cultured human cells (Bogerd et al., 2006; Muckenfuss et al., 2006; Richardson et al., 2014; Stenglein and Harris, 2006). Catalytically inactive A3B and A3C mutants efficiently repressed L1 retrotransposition in cultured human cells, suggesting a deaminase-independent mechanism (Bogerd et al., 2006). Subsequent studies demonstrated that A3A can inhibit L1 retrotransposition by deaminating nascent L1 cDNA during TPRT, suggesting that exposed single-stranded cDNA becomes accessible to deamination during TPRT (Richardson et al., 2014).

A previous study suggests that TPRT-mediated retrotransposition can instigate an epigenetic silencing response in hECs (Garcia-Perez et al., 2010). TPRT-mediated delivery of an *EGFP* retrotransposition-indicator cassette (*mEFGPI)* by a human L1, a synthetic mouse L1, or a zebrafish L2 (a L1-like element) was efficiently silenced in PA-1 cells (Garcia-Perez et al., 2010). However, similar selectable or screenable reporter genes were not efficiently silenced when integrated into PA-1 genomic DNA using stable transfection or retroviral vectors. These results suggest that structures generated during TPRT may target L1 retrotransposition events for epigenetic silencing.

Intriguingly, class I histone deacetylase inhibitors (HDACi) reversed L1-delivered reporter gene silencing, whereas subsequent removal of HDACi resulted in re-establishment of reporter gene silencing over time. Thus, the mechanism of L1-delivered reporter gene silencing in PA-1 cells exhibits the characteristics of an epigenetic system: (1) silencing is mitotically stable; (2) silencing does not alter the

primary DNA sequence (3) silencing is reversible; and (4) silencing is reestablished after drug removal, suggesting an epigenetic mark, or memory, to the system. Together, these results established that L1-delivered reporter genes are efficiently and stably silenced in hECs upon TPRT-mediated genomic integration (Garcia-Perez et al., 2010). Future studies are necessary to elucidate the substrates that mediate this epigenetic silencing mechanism in hECs.

Thesis Overview

This thesis examines an epigenetic silencing mechanism that mediates transcriptional repression of L1 retrotransposition events during embryonic development. In Chapter two we harness CRISPR/Cas9 technology to perform a genome-wide forward genetic screen to identify cellular factors that mediate epigenetic repression of L1-delivered reporter genes in PA-1 human embryonic carcinoma cells. We also develop an adaptable system to test candidate genes in a cell culture-based assay. In Chapter 3 we demonstrate that the tumor suppressor gene neurofibromin 2 (*NF2*) is necessary for efficient transcriptional repression of L1-delivered reporter genes in PA-1 cells. We further demonstrate that exogenous expression of NF2/merlin is sufficient to reinstate efficient silencing of L1-delivered reporter genes in PA-1 cells lacking function NF2/merlin expression. In Chapter four we summarize our findings and consider possible implications for NF2/merlin in L1 biology and propose avenues for future studies.

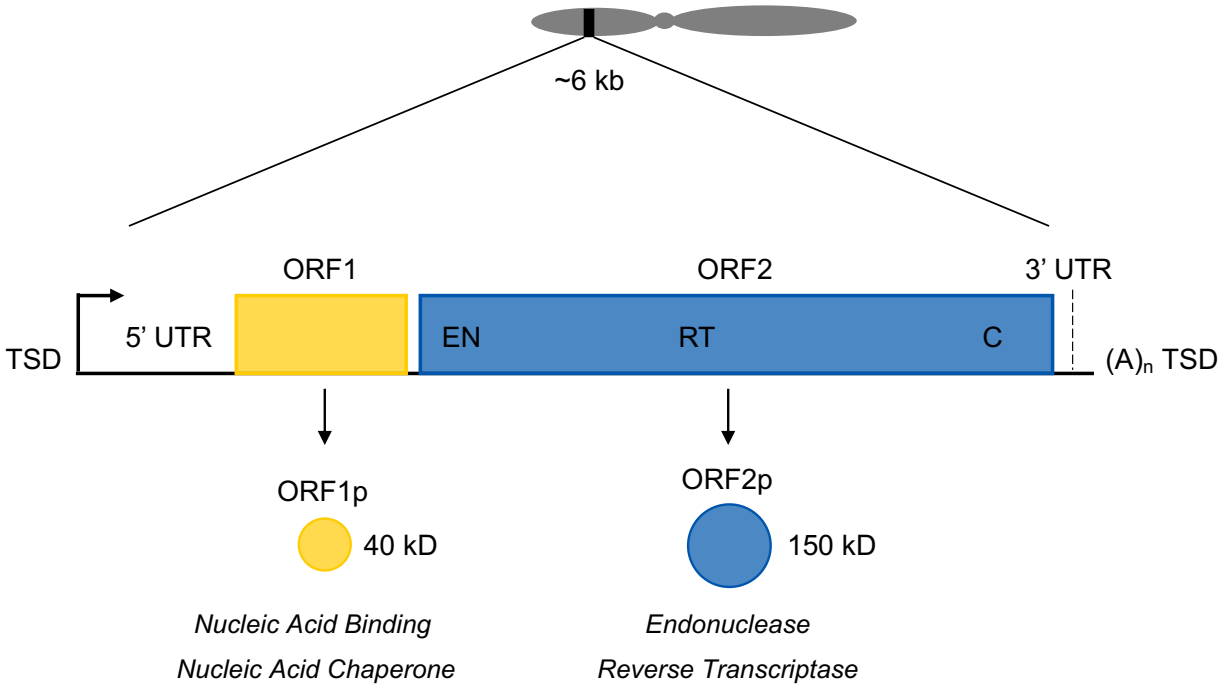


Figure 1.1: Human Long INterspersed Element-1 (LINE-1 or L1).

L1 is an autonomous non-LTR retrotransposon and is currently the only active element in the human genome. Full-length human L1s are ~6 kilobases (kb) in length and comprise a 5' untranslated region (UTR) containing an internal RNA polymerase II promoter, two ORFs separated by a 63 bp inter-ORF region, and a 3'UTR ending in a poly-(A) tract. Genomic L1s also exhibit characteristic ~7-20 bp target site duplications (TSDs). ORF1 encodes a ~40 kilodalton (kD) protein (ORF1p) with nucleic acid binding and chaperone activities. ORF2 encodes a ~150 kD protein (ORF2p) with endonuclease and reverse transcriptase activities.

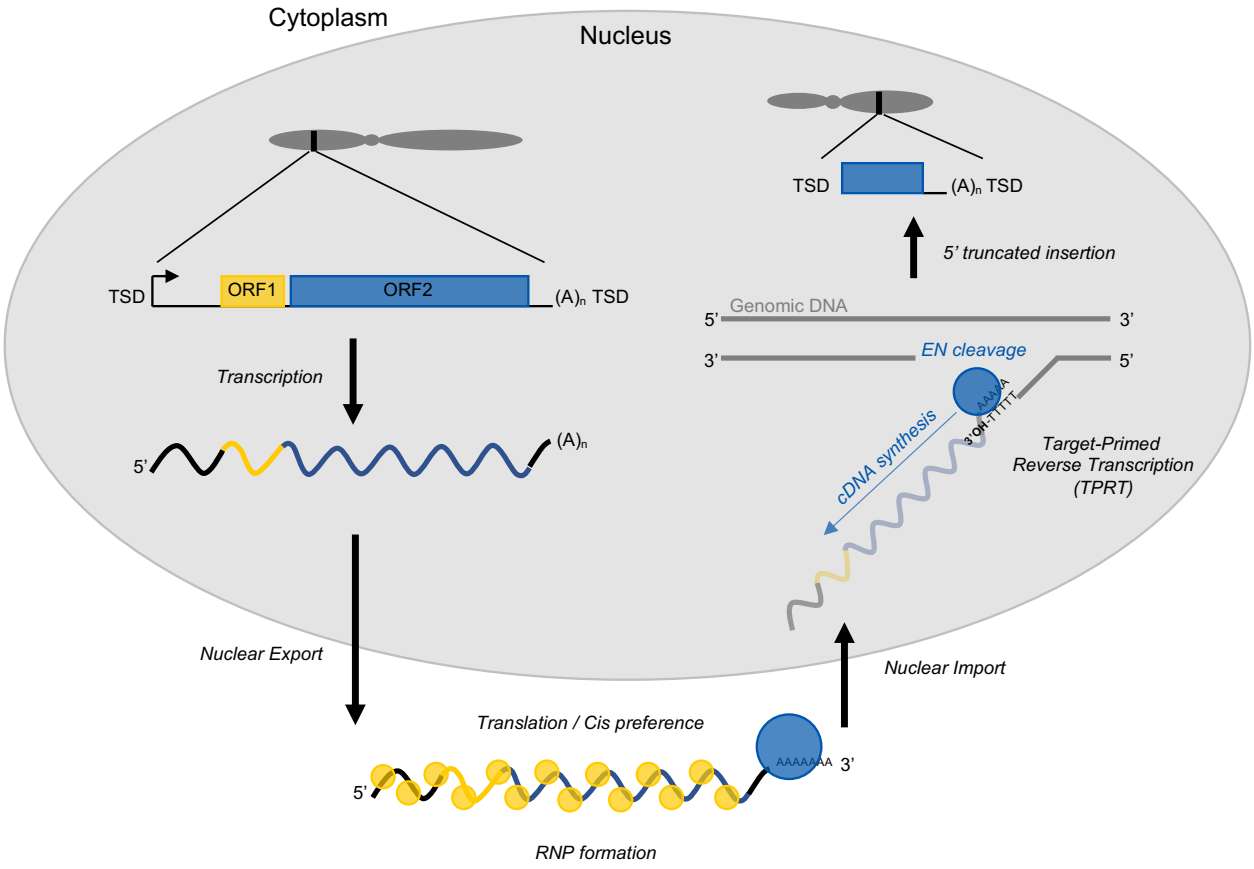


Figure 1.2: L1 retrotransposition cycle.

Schematic of the L1 retrotransposition cycle. L1 RNA is transcribed from an internal promoter located within the 5'UTR. The bicistronic polyadenylated mRNA is exported to the cytoplasm, where it undergoes translation of the L1-encoded proteins ORF1p and ORF2p, which bind their encoding RNA by *cis*-preference, forming L1-RNP. Components of the L1-RNP enter the nucleus and undergo genomic integration by target-primed reverse transcription (TPRT); the L1 EN makes a single-strand endonucleolytic nick at the EN consensus cleavage site 5'-TTTTT/AA-3', which liberates a 3' hydroxyl group (OH) that primes L1 RT cDNA synthesis. The completion of TPRT typically results in a 5' truncated insertion flanked by variable-length target site duplications (TSDs) at a new genomic location.

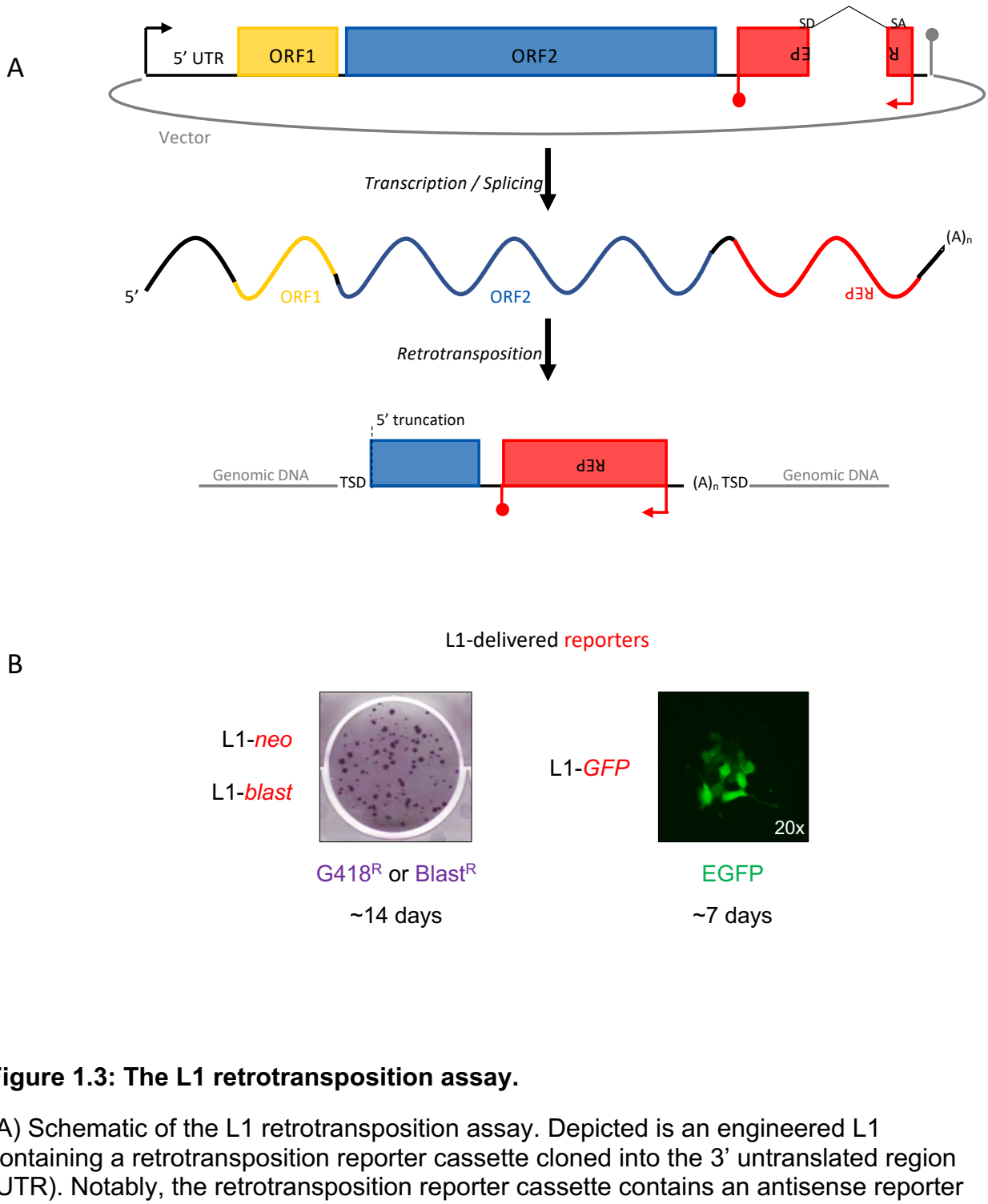


Figure 1.3: The L1 retrotransposition assay.

(A) Schematic of the L1 retrotransposition assay. Depicted is an engineered L1 containing a retrotransposition reporter cassette cloned into the 3' untranslated region (UTR). Notably, the retrotransposition reporter cassette contains an antisense reporter gene (REP) driven by its own promoter. The reporter gene is interrupted by an intron cloned in the same transcriptional orientation as the L1 sequence. This arrangement ensures that a functional reporter gene is only expressed after successful retrotransposition. The vector is transfected into cultured cells, the intron is spliced from sense strand L1 mRNA, and the L1 is integrated into genomic DNA by TPRT. Expression of the reporter gene indicates successful retrotransposition. (B) Results of

the indicated L1 retrotransposition assay using selectable (*neomycin* or *blasticidin*) or screenable (*EGFP*) reporter genes. Drug resistant foci were stained with crystal violet for visualization.

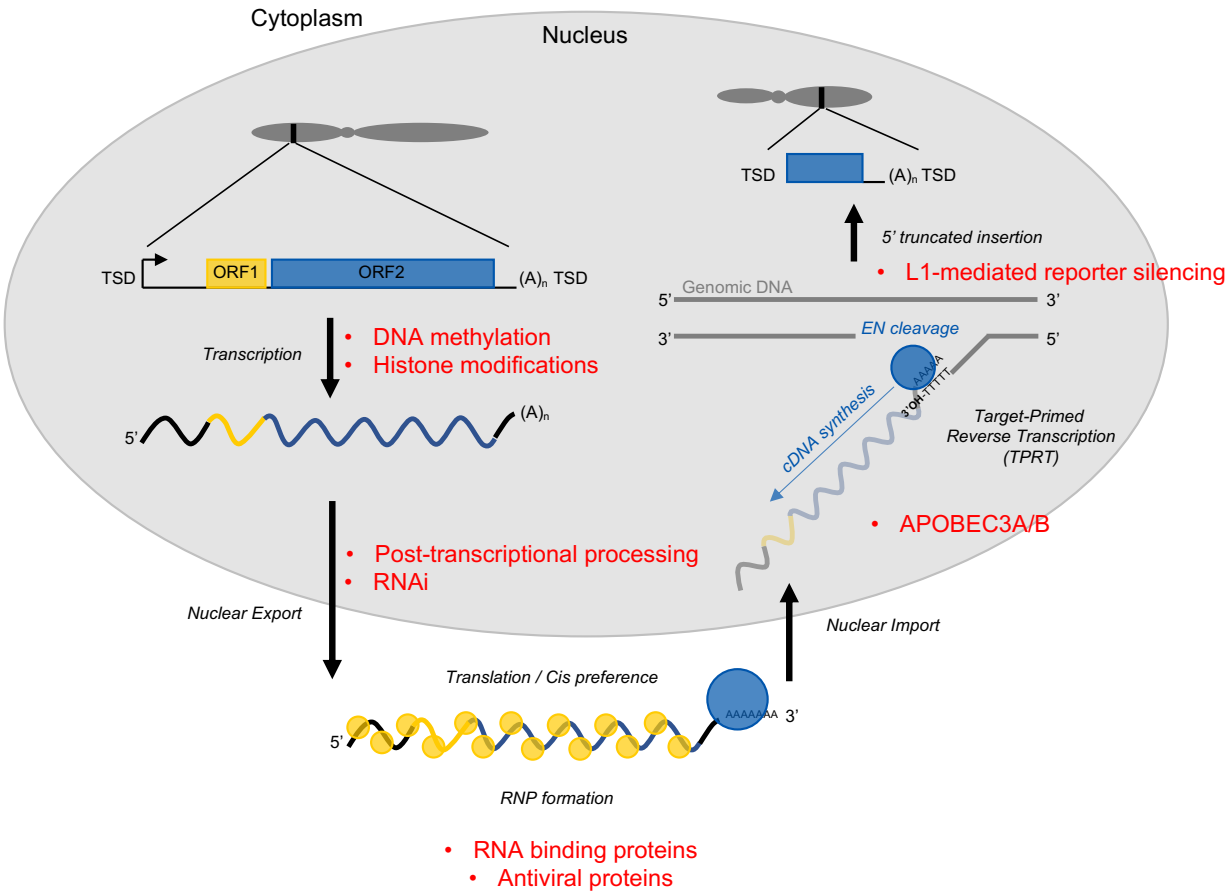


Figure 1.4: Cellular processes that restrict L1 retrotransposition.

Schematic showing the cellular processes that restrict L1 retrotransposition. The cell has evolved to restrict L1 at virtually every step of the L1 retrotransposition cycle. Restrictive processes are indicated in red.

References

Ahl, V., Keller, H., Schmidt, S., and Weichenrieder, O. (2015). Retrotransposition and Crystal Structure of an Alu RNP in the Ribosome-Stalling Conformation. Mol Cell *60*, 715-727.

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

Alves, G., Tatro, A., and Fanning, T. (1996). Differential methylation of human LINE-1 retrotransposons in malignant cells. Gene *176*, 39-44.

An, W., Dai, L., Niewiadomska, A.M., Yetil, A., O'Donnell, K.A., Han, J.S., and Boeke, J.D. (2011). Characterization of a synthetic human LINE-1 retrotransposon ORFeus-Hs. Mob DNA *2*, 2.

An, W., Han, J.S., Wheelan, S.J., Davis, E.S., Coombes, C.E., Ye, P., Triplett, C., and Boeke, J.D. (2006). Active retrotransposition by a synthetic L1 element in mice. Proc Natl Acad Sci U S A *103*, 18662-18667.

Arif, A., Yao, P., Terenzi, F., Jia, J., Ray, P.S., and Fox, P.L. (2018). The GAIT translational control system. Wiley Interdiscip Rev RNA *9*.

Athanikar, J.N., Badge, R.M., and Moran, J.V. (2004). A YY1-binding site is required for accurate human LINE-1 transcription initiation. Nucleic Acids Res *32*, 3846-3855.

Barau, J., Teissandier, A., Zamudio, N., Roy, S., Nalesso, V., Herault, Y., Guillou, F., and Bourc'his, D. (2016). The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. Science *354*, 909-912.

Basame, S., Wai-lun Li, P., Howard, G., Branciforte, D., Keller, D., and Martin, S.L. (2006). Spatial assembly and RNA binding stoichiometry of a LINE-1 protein essential for retrotransposition. J Mol Biol *357*, 351-357.

Beck, C.R., Collier, P., Macfarlane, C., Malig, M., Kidd, J.M., Eichler, E.E., Badge, R.M., and Moran, J.V. (2010). LINE-1 retrotransposition activity in human genomes. Cell *141*, 1159-1170.

Beck, C.R., Garcia-Perez, J.L., Badge, R.M., and Moran, J.V. (2011). LINE-1 elements in structural variation and disease. Annu Rev Genomics Hum Genet *12*, 187-215.

Becker, K.G., Swergold, G.D., Ozato, K., and Thayer, R.E. (1993). Binding of the ubiquitous nuclear transcription factor YY1 to a cis regulatory sequence in the human LINE-1 transposable element. Hum Mol Genet *2*, 1697-1702.

Belancio, V.P., Hedges, D.J., and Deininger, P. (2006). LINE-1 RNA splicing and influences on mammalian gene expression. Nucleic Acids Res *34*, 1512-1521.

Belancio, V.P., Roy-Engel, A.M., and Deininger, P. (2008). The impact of multiple splice sites in human L1 elements. Gene *411*, 38-45.

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

Belshaw, R., Dawson, A.L., Woolven-Allen, J., Redding, J., Burt, A., and Tristem, M. (2005). Genomewide screening reveals high levels of insertional polymorphism in the human endogenous retrovirus family HERV-K(HML2): implications for present-day activity. J Virol *79*, 12507-12514.

Bennett, E.A., Keller, H., Mills, R.E., Schmidt, S., Moran, J.V., Weichenrieder, O., and Devine, S.E. (2008). Active Alu retrotransposons in the human genome. Genome Res *18*, 1875-1883.

Boeke, J.D., Garfinkel, D.J., Styles, C.A., and Fink, G.R. (1985). Ty elements transpose through an RNA intermediate. Cell *40*, 491-500.

Boeke, J.D., and Stoye, J.P. (1997). Retrotransposons, Endogenous Retroviruses, and the Evolution of Retroelements. In Retroviruses, J.M. Coffin, S.H. Hughes, and H.E. Varmus, eds. (Cold Spring Harbor (NY)).

Bogerd, H.P., Wiegand, H.L., Doehle, B.P., Lueders, K.K., and Cullen, B.R. (2006). APOBEC3A and APOBEC3B are potent inhibitors of LTR-retrotransposon function in human cells. Nucleic Acids Res *34*, 89-95.

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

Bratthauer, G.L., and Fanning, T.G. (1993). LINE-1 retrotransposon expression in pediatric germ cell tumors. Cancer *71*, 2383-2386.

Brouha, B., Meischl, C., Ostertag, E., de Boer, M., Zhang, Y., Neijens, H., Roos, D., and Kazazian, H.H., Jr. (2002). Evidence consistent with human L1 retrotransposition in maternal meiosis I. Am J Hum Genet *71*, 327-336.

Brouha, B., Schustak, J., Badge, R.M., Lutz-Prigge, S., Farley, A.H., Moran, J.V., and Kazazian, H.H., Jr. (2003). Hot L1s account for the bulk of retrotransposition in the human population. Proc Natl Acad Sci U S A *100*, 5280-5285.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods *10*, 1213-1218.

Bulut-Karslioglu, A., De La Rosa-Velazquez, I.A., Ramirez, F., Barenboim, M., Onishi-Seebacher, M., Arand, J., Galan, C., Winter, G.E., Engist, B., Gerle, B.*, et al.* (2014). Suv39h-dependent H3K9me3 marks intact retrotransposons and silences LINE elements in mouse embryonic stem cells. Mol Cell *55*, 277-290.

Bundo, M., Toyoshima, M., Okada, Y., Akamatsu, W., Ueda, J., Nemoto-Miyauchi, T., Sunaga, F., Toritsuka, M., Ikawa, D., Kakita, A.*, et al.* (2014). Increased l1 retrotransposition in the neuronal genome in schizophrenia. Neuron *81*, 306-313.

Burns, K.H. (2017). Transposable elements in cancer. Nat Rev Cancer *17*, 415-424.

Burns, K.H. (2020). Our Conflict with Transposable Elements and Its Implications for Human Disease. Annu Rev Pathol *15*, 51-70.

Buzdin, A., Ustyugova, S., Gogvadze, E., Vinogradova, T., Lebedev, Y., and Sverdlov, E. (2002). A new family of chimeric retrotranscripts formed by a full copy of U6 small nuclear RNA fused to the 3' terminus of l1. Genomics *80*, 402-406.

Callahan, K.E., Hickman, A.B., Jones, C.E., Ghirlando, R., and Furano, A.V. (2012). Polymerization and nucleic acid-binding properties of human L1 ORF1 protein. Nucleic Acids Res *40*, 813-827.

Callinan, P.A., Wang, J., Herke, S.W., Garber, R.K., Liang, P., and Batzer, M.A. (2005). Alu retrotransposition-mediated deletion. J Mol Biol *348*, 791-800.

Cary, L.C., Goebel, M., Corsaro, B.G., Wang, H.G., Rosen, E., and Fraser, M.J. (1989). Transposon mutagenesis of baculoviruses: analysis of Trichoplusia ni transposon IFP2 insertions within the FP-locus of nuclear polyhedrosis viruses. Virology *172*, 156-169.

Castel, S.E., and Martienssen, R.A. (2013). RNA interference in the nucleus: roles for small RNAs in transcription, epigenetics and beyond. Nat Rev Genet *14*, 100-112.

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

Chen, Z., and Zhang, Y. (2020). Role of Mammalian DNA Methyltransferases in Development. Annu Rev Biochem *89*, 135-158.

Chiu, Y.L., and Greene, W.C. (2008). The APOBEC3 cytidine deaminases: an innate defensive network opposing exogenous retroviruses and endogenous retroelements. Annu Rev Immunol *26*, 317-353.

Choi, J., Ryoo, J., Oh, C., Hwang, S., and Ahn, K. (2015). SAMHD1 specifically restricts retroviruses through its RNase activity. Retrovirology *12*, 46.

Christensen, S.M., and Eickbush, T.H. (2005). R2 target-primed reverse transcription: ordered cleavage and polymerization steps by protein subunits asymmetrically bound to the target DNA. Mol Cell Biol *25*, 6617-6628.

Chu, W.M., Liu, W.M., and Schmid, C.W. (1995). RNA polymerase III promoter and terminator elements affect Alu RNA expression. Nucleic Acids Res *23*, 1750-1757.

Chuong, E.B., Elde, N.C., and Feschotte, C. (2016). Regulatory evolution of innate immunity through co-option of endogenous retroviruses. Science *351*, 1083-1087.

Chuong, E.B., Elde, N.C., and Feschotte, C. (2017). Regulatory activities of transposable elements: from conflicts to benefits. Nat Rev Genet *18*, 71-86.

Clements, A.P., and Singer, M.F. (1998). The human LINE-1 reverse transcriptase:effect of deletions outside the common reverse transcriptase domain. Nucleic Acids Res *26*, 3528-3535.

Cook, P.R., Jones, C.E., and Furano, A.V. (2015). Phosphorylation of ORF1p is required for L1 retrotransposition. Proc Natl Acad Sci U S A *112*, 4298-4303.

Cordaux, R., and Batzer, M.A. (2009). The impact of retrotransposons on human genome evolution. Nat Rev Genet *10*, 691-703.

Cost, G.J., and Boeke, J.D. (1998). Targeting of human retrotransposon integration is directed by the specificity of the L1 endonuclease for regions of unusual DNA structure. Biochemistry *37*, 18081-18093.

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

Cost, G.J., Golding, A., Schlissel, M.S., and Boeke, J.D. (2001). Target DNA chromatinization modulates nicking by L1 endonuclease. Nucleic Acids Res *29*, 573-577.

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

Davidson, E.H., and Britten, R.J. (1979). Regulation of gene expression: possible role of repetitive sequences. Science *204*, 1052-1059.

De Cecco, M., Ito, T., Petrashen, A.P., Elias, A.E., Skvir, N.J., Criscione, S.W., Caligiana, A., Brocculi, G., Adney, E.M., Boeke, J.D.*, et al.* (2019). L1 drives IFN in senescent cells and promotes age-associated inflammation. Nature *566*, 73-78.

de Koning, A.P., Gu, W., Castoe, T.A., Batzer, M.A., and Pollock, D.D. (2011). Repetitive elements may comprise over two-thirds of the human genome. PLoS Genet *7*, e1002384.

Decker, C.J., and Parker, R. (2012). P-bodies and stress granules: possible roles in the control of translation and mRNA degradation. Cold Spring Harb Perspect Biol *4*, a012286.

Deininger, P.L., Jolly, D.J., Rubin, C.M., Friedmann, T., and Schmid, C.W. (1981). Base sequence studies of 300 nucleotide renatured repeated human DNA clones. J Mol Biol *151*, 17-33.

Denli, A.M., Narvaiza, I., Kerman, B.E., Pena, M., Benner, C., Marchetto, M.C., Diedrich, J.K., Aslanian, A., Ma, J., Moresco, J.J.*, et al.* (2015). Primate-specific ORF0 contributes to retrotransposon-mediated diversity. Cell *163*, 583-593.

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

Dewannieux, M., and Heidmann, T. (2005). LINEs, SINEs and processed pseudogenes: parasitic strategies for genome modeling. Cytogenet Genome Res *110*, 35-48.

Di Giacomo, M., Comazzetto, S., Sampath, S.C., Sampath, S.C., and O'Carroll, D. (2014). G9a co-suppresses LINE1 elements in spermatogonia. Epigenetics Chromatin *7*, 24.

Ding, S., Wu, X., Li, G., Han, M., Zhuang, Y., and Xu, T. (2005). Efficient transposition of the piggyBac (PB) transposon in mammalian cells and mice. Cell *122*, 473-483.

Dombroski, B.A., Feng, Q., Mathias, S.L., Sassaman, D.M., Scott, A.F., Kazazian, H.H., Jr., and Boeke, J.D. (1994). An in vivo assay for the reverse transcriptase of human retrotransposon L1 in Saccharomyces cerevisiae. Mol Cell Biol *14*, 4485-4492.

Dombroski, B.A., Mathias, S.L., Nanthakumar, E., Scott, A.F., and Kazazian, H.H., Jr. (1991). Isolation of an active human transposable element. Science *254*, 1805-1808.

Doolittle, W.F., and Sapienza, C. (1980). Selfish genes, the phenotype paradigm and genome evolution. Nature *284*, 601-603.

Doucet, A.J., Hulme, A.E., Sahinovic, E., Kulpa, D.A., Moldovan, J.B., Kopera, H.C., Athanikar, J.N., Hasnaoui, M., Bucheton, A., Moran, J.V.*, et al.* (2010). Characterization of LINE-1 ribonucleoprotein particles. PLoS Genet *6*.

Doucet, A.J., Wilusz, J.E., Miyoshi, T., Liu, Y., and Moran, J.V. (2015). A 3' Poly(A) Tract Is Required for LINE-1 Retrotransposition. Mol Cell *60*, 728-741.

Doucet-O'Hare, T.T., Rodic, N., Sharma, R., Darbari, I., Abril, G., Choi, J.A., Young Ahn, J., Cheng, Y., Anders, R.A., Burns, K.H.*, et al.* (2015). LINE-1 expression and

retrotransposition in Barrett's esophagus and esophageal carcinoma. Proc Natl Acad Sci U S A *112*, E4894-4900.

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

Doudna, J.A., and Rath, V.L. (2002). Structure and function of the eukaryotic ribosome: the next frontier. Cell *109*, 153-156.

Doyle, G.A., Crist, R.C., Karatas, E.T., Hammond, M.J., Ewing, A.D., Ferraro, T.N., Hahn, C.G., and Berrettini, W.H. (2017). Analysis of LINE-1 Elements in DNA from Postmortem Brains of Individuals with Schizophrenia. Neuropsychopharmacology *42*, 2602-2611.

Dupressoir, A., Lavialle, C., and Heidmann, T. (2012). From ancestral infectious retroviruses to bona fide cellular genes: role of the captured syncytins in placentation. Placenta *33*, 663-671.

Eickbush, T.H. (1997). Telomerase and retrotransposons: which came first? Science *277*, 911-912.

Eickbush, T.H., and Jamburuthugoda, V.K. (2008). The diversity of retrotransposons and the properties of their reverse transcriptases. Virus Res *134*, 221-234.

Ergun, S., Buschmann, C., Heukeshoven, J., Dammann, K., Schnieders, F., Lauke, H., Chalajour, F., Kilic, N., Stratling, W.H., and Schumann, G.G. (2004). Cell type-specific expression of LINE-1 open reading frames 1 and 2 in fetal and adult human tissues. J Biol Chem *279*, 27753-27763.

Erwin, J.A., Paquola, A.C., Singer, T., Gallina, I., Novotny, M., Quayle, C., Bedrosian, T.A., Alves, F.I., Butcher, C.R., Herdy, J.R.*, et al.* (2016). L1-associated genomic regions are deleted in somatic cells of the healthy human brain. Nat Neurosci *19*, 1583-1591.

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

Evrony, G.D., Lee, E., Mehta, B.K., Benjamini, Y., Johnson, R.M., Cai, X., Yang, L., Haseley, P., Lehmann, H.S., Park, P.J.*, et al.* (2015). Cell lineage analysis in human brain using endogenous retroelements. Neuron *85*, 49-59.

Ewing, A.D., Gacita, A., Wood, L.D., Ma, F., Xing, D., Kim, M.S., Manda, S.S., Abril, G., Pereira, G., Makohon-Moore, A.*, et al.* (2015). Widespread somatic L1 retrotransposition occurs early during gastrointestinal cancer evolution. Genome Res *25*, 1536-1545.

Fanning, T.G., and Singer, M.F. (1987). LINE-1: a mammalian transposable element. Biochim Biophys Acta *910*, 203-212.

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

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

Feschotte, C., and Pritham, E.J. (2005). Non-mammalian c-integrases are encoded by giant transposable elements. Trends Genet *21*, 551-552.

Feschotte, C., and Pritham, E.J. (2007). DNA transposons and the evolution of eukaryotic genomes. Annu Rev Genet *41*, 331-368.

Feusier, J., Watkins, W.S., Thomas, J., Farrell, A., Witherspoon, D.J., Baird, L., Ha, H., Xing, J., and Jorde, L.B. (2019). Pedigree-based estimation of human mobile element retrotransposition rates. Genome Res *29*, 1567-1577.

Finnegan, D.J. (1989). Eukaryotic transposable elements and genome evolution. Trends Genet *5*, 103-107.

Flasch, D.A., Macia, A., Sanchez, L., Ljungman, M., Heras, S.R., Garcia-Perez, J.L., Wilson, T.E., and Moran, J.V. (2019). Genome-wide de novo L1 Retrotransposition Connects Endonuclease Activity with Replication. Cell *177*, 837-851 e828.

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

Fricker, A.D., and Peters, J.E. (2014). Vulnerabilities on the lagging-strand template: opportunities for mobile elements. Annu Rev Genet *48*, 167-186.

Fukuda, S., Varshney, A., Fowler, B.J., Wang, S.B., Narendran, S., Ambati, K., Yasuma, T., Magagnoli, J., Leung, H., Hirahara, S.*, et al.* (2021). Cytoplasmic synthesis of endogenous Alu complementary DNA via reverse transcription and implications in age-related macular degeneration. Proc Natl Acad Sci U S A *118*.

Gao, G., Guo, X., and Goff, S.P. (2002). Inhibition of retroviral RNA production by ZAP, a CCCH-type zinc finger protein. Science *297*, 1703-1706.

Garcia-Montojo, M., Doucet-O'Hare, T., Henderson, L., and Nath, A. (2018). Human endogenous retrovirus-K (HML-2): a comprehensive review. Crit Rev Microbiol *44*, 715-738.

Garcia-Perez, J.L., Doucet, A.J., Bucheton, A., Moran, J.V., and Gilbert, N. (2007a). Distinct mechanisms for trans-mediated mobilization of cellular RNAs by the LINE-1 reverse transcriptase. Genome Res *17*, 602-611.

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

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

Garfinkel, D.J., Curcio, M.J., Youngren, S.D., and Sanders, N.J. (1989). The biology and exploitation of the retrotransposon Ty in Saccharomyces cerevisiae. Genome *31*, 909-919.

Georgiou, I., Noutsopoulos, D., Dimitriadou, E., Markopoulos, G., Apergi, A., Lazaros, L., Vaxevanoglou, T., Pantos, K., Syrrou, M., and Tzavaras, T. (2009). Retrotransposon RNA expression and evidence for retrotransposition events in human oocytes. Hum Mol Genet *18*, 1221-1228.

Gilbert, N., Lutz, S., Morrish, T.A., and Moran, J.V. (2005). Multiple fates of L1 retrotransposition intermediates in cultured human cells. Mol Cell Biol *25*, 7780-7795.

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

Goll, M.G., and Bestor, T.H. (2005). Eukaryotic cytosine methyltransferases. Annu Rev Biochem *74*, 481-514.

Goodier, J.L. (2016). Restricting retrotransposons: a review. Mob DNA *7*, 16.

Goodier, J.L., Cheung, L.E., and Kazazian, H.H., Jr. (2012). MOV10 RNA helicase is a potent inhibitor of retrotransposition in cells. PLoS Genet *8*, e1002941.

Goodier, J.L., Cheung, L.E., and Kazazian, H.H., Jr. (2013). Mapping the LINE1 ORF1 protein interactome reveals associated inhibitors of human retrotransposition. Nucleic Acids Res *41*, 7401-7419.

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

Goodier, J.L., Pereira, G.C., Cheung, L.E., Rose, R.J., and Kazazian, H.H., Jr. (2015). The Broad-Spectrum Antiviral Protein ZAP Restricts Human Retrotransposition. PLoS Genet *11*, e1005252.

Grimaldi, G., Skowronski, J., and Singer, M.F. (1984). Defining the beginning and end of KpnI family segments. EMBO J *3*, 1753-1759.

Grow, E.J., Flynn, R.A., Chavez, S.L., Bayless, N.L., Wossidlo, M., Wesche, D.J., Martin, L., Ware, C.B., Blish, C.A., Chang, H.Y.*, et al.* (2015). Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. Nature *522*, 221-225.

Hamdorf, M., Idica, A., Zisoulis, D.G., Gamelin, L., Martin, C., Sanders, K.J., and Pedersen, I.M. (2015). miR-128 represses L1 retrotransposition by binding directly to L1 RNA. Nat Struct Mol Biol *22*, 824-831.

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

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

Han, K., Lee, J., Meyer, T.J., Remedios, P., Goodwin, L., and Batzer, M.A. (2008). L1 recombination-associated deletions generate human genomic variation. Proc Natl Acad Sci U S A *105*, 19366-19371.

Han, K., Sen, S.K., Wang, J., Callinan, P.A., Lee, J., Cordaux, R., Liang, P., and Batzer, M.A. (2005). Genomic rearrangements by LINE-1 insertion-mediated deletion in the human and chimpanzee lineages. Nucleic Acids Res *33*, 4040-4052.

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

Hancks, D.C., and Kazazian, H.H., Jr. (2016). Roles for retrotransposon insertions in human disease. Mob DNA *7*, 9.

Hattori, M., Kuhara, S., Takenaka, O., and Sakaki, Y. (1986). L1 family of repetitive DNA sequences in primates may be derived from a sequence encoding a reverse transcriptase-related protein. Nature *321*, 625-628.

Hayakawa, T., Satta, Y., Gagneux, P., Varki, A., and Takahata, N. (2001). Alu-mediated inactivation of the human CMP- N-acetylneuraminic acid hydroxylase gene. Proc Natl Acad Sci U S A *98*, 11399-11404.

Heidmann, O., and Heidmann, T. (1991). Retrotransposition of a mouse IAP sequence tagged with an indicator gene. Cell *64*, 159-170.

Heras, S.R., Macias, S., Caceres, J.F., and Garcia-Perez, J.L. (2014). Control of mammalian retrotransposons by cellular RNA processing activities. Mob Genet Elements *4*, e28439.

Heras, S.R., Macias, S., Plass, M., Fernandez, N., Cano, D., Eyras, E., Garcia-Perez, J.L., and Caceres, J.F. (2013). The Microprocessor controls the activity of mammalian retrotransposons. Nat Struct Mol Biol *20*, 1173-1181.

Hickman, A.B., and Dyda, F. (2016). DNA Transposition at Work. Chem Rev *116*, 12758-12784.

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

Hohjoh, H., and Singer, M.F. (1997). Sequence-specific single-strand RNA binding protein encoded by the human LINE-1 retrotransposon. EMBO J *16*, 6034-6043.

Holmes, S.E., Dombroski, B.A., Krebs, C.M., Boehm, C.D., and Kazazian, H.H., Jr. (1994). A new retrotransposable human L1 element from the LRE2 locus on chromosome 1q produces a chimaeric insertion. Nat Genet *7*, 143-148.

Holmes, S.E., Singer, M.F., and Swergold, G.D. (1992). Studies on p40, the leucine zipper motif-containing protein encoded by the first open reading frame of an active human LINE-1 transposable element. J Biol Chem *267*, 19765-19768.

Howell, R., and Usdin, K. (1997). The ability to form intrastrand tetraplexes is an evolutionarily conserved feature of the 3' end of L1 retrotransposons. Mol Biol Evol *14*, 144-155.

Hu, S., Li, J., Xu, F., Mei, S., Le Duff, Y., Yin, L., Pang, X., Cen, S., Jin, Q., Liang, C.*, et al.* (2015). SAMHD1 Inhibits LINE-1 Retrotransposition by Promoting Stress Granule Formation. PLoS Genet *11*, e1005367.

Huang, S., Tao, X., Yuan, S., Zhang, Y., Li, P., Beilinson, H.A., Zhang, Y., Yu, W., Pontarotti, P., Escriva, H.*, et al.* (2016). Discovery of an Active RAG Transposon Illuminates the Origins of V(D)J Recombination. Cell *166*, 102-114.

Iskow, R.C., McCabe, M.T., Mills, R.E., Torene, S., Pittard, W.S., Neuwald, A.F., Van Meir, E.G., Vertino, P.M., and Devine, S.E. (2010). Natural mutagenesis of human genomes by endogenous retrotransposons. Cell *141*, 1253-1261.

Ivics, Z., Hackett, P.B., Plasterk, R.H., and Izsvak, Z. (1997). Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells. Cell *91*, 501-510.

Ivics, Z., and Izsvak, Z. (2015). Sleeping Beauty Transposition. Microbiol Spectr *3*, MDNA3-0042-2014.

Iyengar, S., and Farnham, P.J. (2011). KAP1 protein: an enigmatic master regulator of the genome. J Biol Chem *286*, 26267-26276.

Jacobs, F.M., Greenberg, D., Nguyen, N., Haeussler, M., Ewing, A.D., Katzman, S., Paten, B., Salama, S.R., and Haussler, D. (2014). An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. Nature *516*, 242-245.

Januszyk, K., Li, P.W., Villareal, V., Branciforte, D., Wu, H., Xie, Y., Feigon, J., Loo, J.A., Martin, S.L., and Clubb, R.T. (2007). Identification and solution structure of a highly conserved C-terminal domain within ORF1p required for retrotransposition of long interspersed nuclear element-1. J Biol Chem *282*, 24893-24904.

Jones, J.M., and Gellert, M. (2004). The taming of a transposon: V(D)J recombination and the immune system. Immunol Rev *200*, 233-248.

Jones, P.A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet *13*, 484-492.

Jurka, J. (1997). Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons. Proc Natl Acad Sci U S A *94*, 1872-1877.

Jurka, J. (2008). Conserved eukaryotic transposable elements and the evolution of gene regulation. Cell Mol Life Sci *65*, 201-204.

Kajikawa, M., Sugano, T., Sakurai, R., and Okada, N. (2012). Low dependency of retrotransposition on the ORF1 protein of the zebrafish LINE, ZfL2-1. Gene *499*, 41-47.

Kapitonov, V.V., and Jurka, J. (2001). Rolling-circle transposons in eukaryotes. Proc Natl Acad Sci U S A *98*, 8714-8719.

Kapitonov, V.V., and Jurka, J. (2005). RAG1 core and V(D)J recombination signal sequences were derived from Transib transposons. PLoS Biol *3*, e181.

Kapitonov, V.V., and Jurka, J. (2006). Self-synthesizing DNA transposons in eukaryotes. Proc Natl Acad Sci U S A *103*, 4540-4545.

Kawakami, K. (2007). Tol2: a versatile gene transfer vector in vertebrates. Genome Biol *8 Suppl 1*, S7.

Kazazian, H.H., Jr. (1999). An estimated frequency of endogenous insertional mutations in humans. Nat Genet *22*, 130.

Kazazian, H.H., Jr., and Moran, J.V. (1998). The impact of L1 retrotransposons on the human genome. Nat Genet *19*, 19-24.

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

Kazazian, H.H., Jr., Wong, C., Youssoufian, H., Scott, A.F., Phillips, D.G., and Antonarakis, S.E. (1988). Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. Nature *332*, 164-166.

Kebriaei, P., Izsvak, Z., Narayanavari, S.A., Singh, H., and Ivics, Z. (2017). Gene Therapy with the Sleeping Beauty Transposon System. Trends Genet *33*, 852-870.

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

Khazina, E., and Weichenrieder, O. (2009). Non-LTR retrotransposons encode noncanonical RRM domains in their first open reading frame. Proc Natl Acad Sci U S A *106*, 731-736.

Khazina, E., and Weichenrieder, O. (2018). Human LINE-1 retrotransposition requires a metastable coiled coil and a positively charged N-terminus in L1ORF1p. Elife *7*.

Kidd, J.M., Graves, T., Newman, T.L., Fulton, R., Hayden, H.S., Malig, M., Kallicki, J., Kaul, R., Wilson, R.K., and Eichler, E.E. (2010). A human genome structural variation sequencing resource reveals insights into mutational mechanisms. Cell *143*, 837-847.

Kimberland, M.L., Divoky, V., Prchal, J., Schwahn, U., Berger, W., and Kazazian, H.H., Jr. (1999). Full-length human L1 insertions retain the capacity for high frequency retrotransposition in cultured cells. Hum Mol Genet *8*, 1557-1560.

Klawitter, S., Fuchs, N.V., Upton, K.R., Munoz-Lopez, M., Shukla, R., Wang, J., Garcia-Canadas, M., Lopez-Ruiz, C., Gerhardt, D.J., Sebe, A.*, et al.* (2016). Reprogramming triggers endogenous L1 and Alu retrotransposition in human induced pluripotent stem cells. Nat Commun *7*, 10286.

Koito, A., and Ikeda, T. (2013). Intrinsic immunity against retrotransposons by APOBEC cytidine deaminases. Front Microbiol *4*, 28.

Kojima, K.K., and Jurka, J. (2011). Crypton transposons: identification of new diverse families and ancient domestication events. Mob DNA *2*, 12.

Kopera, H.C., Flasch, D.A., Nakamura, M., Miyoshi, T., Doucet, A.J., and Moran, J.V. (2016a). LEAP: L1 Element Amplification Protocol. Methods Mol Biol *1400*, 339-355.

Kopera, H.C., Larson, P.A., Moldovan, J.B., Richardson, S.R., Liu, Y., and Moran, J.V. (2016b). LINE-1 Cultured Cell Retrotransposition Assay. Methods Mol Biol *1400*, 139-156.

Kopera, H.C., Moldovan, J.B., Morrish, T.A., Garcia-Perez, J.L., and Moran, J.V. (2011). Similarities between long interspersed element-1 (LINE-1) reverse transcriptase and telomerase. Proc Natl Acad Sci U S A *108*, 20345-20350.

Kornberg, R.D., and Lorch, Y. (2020). Primary Role of the Nucleosome. Mol Cell *79*, 371-375.

Kouzarides, T. (2007). Chromatin modifications and their function. Cell *128*, 693-705.

Kriegs, J.O., Churakov, G., Jurka, J., Brosius, J., and Schmitz, J. (2007). Evolutionary history of 7SL RNA-derived SINEs in Supraprimates. Trends Genet *23*, 158-161.

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

Kulpa, D.A., and Moran, J.V. (2005). Ribonucleoprotein particle formation is necessary but not sufficient for LINE-1 retrotransposition. Hum Mol Genet *14*, 3237-3248.

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

Lahouassa, H., Daddacha, W., Hofmann, H., Ayinde, D., Logue, E.C., Dragin, L., Bloch, N., Maudet, C., Bertrand, M., Gramberg, T.*, et al.* (2012). SAMHD1 restricts the replication of human immunodeficiency virus type 1 by depleting the intracellular pool of deoxynucleoside triphosphates. Nat Immunol *13*, 223-228.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W.*, et al.* (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860-921.

Larson, P.A., Moldovan, J.B., Jasti, N., Kidd, J.M., Beck, C.R., and Moran, J.V. (2018). Spliced integrated retrotransposed element (SpIRE) formation in the human genome. PLoS Biol *16*, e2003067.

Lee, E., Iskow, R., Yang, L., Gokcumen, O., Haseley, P., Luquette, L.J., 3rd, Lohr, J.G., Harris, C.C., Ding, L., Wilson, R.K.*, et al.* (2012). Landscape of somatic retrotransposition in human cancers. Science *337*, 967-971.

Levin, H.L., and Moran, J.V. (2011). Dynamic interactions between transposable elements and their hosts. Nat Rev Genet *12*, 615-627.

Liao, C.H., Kuang, Y.Q., Liu, H.L., Zheng, Y.T., and Su, B. (2007). A novel fusion gene, TRIM5-Cyclophilin A in the pig-tailed macaque determines its susceptibility to HIV-1 infection. AIDS *21 Suppl 8*, S19-26.

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

Luan, D.D., Korman, M.H., Jakubczak, J.L., and Eickbush, T.H. (1993). Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. Cell *72*, 595-605.

Macia, A., Munoz-Lopez, M., Cortes, J.L., Hastings, R.K., Morell, S., Lucena-Aguilar, G., Marchal, J.A., Badge, R.M., and Garcia-Perez, J.L. (2011). Epigenetic control of retrotransposon expression in human embryonic stem cells. Mol Cell Biol *31*, 300-316.

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

Mager, D.L., and Stoye, J.P. (2015). Mammalian Endogenous Retroviruses. Microbiol Spectr *3*, MDNA3-0009-2014.

Magiorkinis, G., Gifford, R.J., Katzourakis, A., De Ranter, J., and Belshaw, R. (2012). Env-less endogenous retroviruses are genomic superspreaders. Proc Natl Acad Sci U S A *109*, 7385-7390.

Malik, H.S., Burke, W.D., and Eickbush, T.H. (1999). The age and evolution of non-LTR retrotransposable elements. Mol Biol Evol *16*, 793-805.

Martin, F., Maranon, C., Olivares, M., Alonso, C., and Lopez, M.C. (1995). Characterization of a non-long terminal repeat retrotransposon cDNA (L1Tc) from Trypanosoma cruzi: homology of the first ORF with the ape family of DNA repair enzymes. J Mol Biol *247*, 49-59.

Martin, S.L. (1991). Ribonucleoprotein particles with LINE-1 RNA in mouse embryonal carcinoma cells. Mol Cell Biol *11*, 4804-4807.

Martin, S.L. (2006). The ORF1 protein encoded by LINE-1: structure and function during L1 retrotransposition. J Biomed Biotechnol *2006*, 45621.

Martin, S.L. (2010). Nucleic acid chaperone properties of ORF1p from the non-LTR retrotransposon, LINE-1. RNA Biol *7*, 706-711.

Martin, S.L., and Branciforte, D. (1993). Synchronous expression of LINE-1 RNA and protein in mouse embryonal carcinoma cells. Mol Cell Biol *13*, 5383-5392.

Martin, S.L., Branciforte, D., Keller, D., and Bain, D.L. (2003). Trimeric structure for an essential protein in L1 retrotransposition. Proc Natl Acad Sci U S A *100*, 13815-13820.

Martin, S.L., and Bushman, F.D. (2001). Nucleic acid chaperone activity of the ORF1 protein from the mouse LINE-1 retrotransposon. Mol Cell Biol *21*, 467-475.

Martin, S.L., Cruceanu, M., Branciforte, D., Wai-Lun Li, P., Kwok, S.C., Hodges, R.S., and Williams, M.C. (2005). LINE-1 retrotransposition requires the nucleic acid chaperone activity of the ORF1 protein. J Mol Biol *348*, 549-561.

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

Matlik, K., Redik, K., and Speek, M. (2006). L1 antisense promoter drives tissue-specific transcription of human genes. J Biomed Biotechnol *2006*, 71753.

McClintock, B. (1950). The origin and behavior of mutable loci in maize. Proc Natl Acad Sci U S A *36*, 344-355.

McClintock, B. (1951). Chromosome organization and genic expression. Cold Spring Harb Symp Quant Biol *16*, 13-47.

McConnell, M.J., Moran, J.V., Abyzov, A., Akbarian, S., Bae, T., Cortes-Ciriano, I., Erwin, J.A., Fasching, L., Flasch, D.A., Freed, D.*, et al.* (2017). Intersection of diverse neuronal genomes and neuropsychiatric disease: The Brain Somatic Mosaicism Network. Science *356*.

Medstrand, P., and Mager, D.L. (1998). Human-specific integrations of the HERV-K endogenous retrovirus family. J Virol *72*, 9782-9787.

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

Minakami, R., Kurose, K., Etoh, K., Furuhata, Y., Hattori, M., and Sakaki, Y. (1992). Identification of an internal cis-element essential for the human L1 transcription and a nuclear factor(s) binding to the element. Nucleic Acids Res *20*, 3139-3145.

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

Moldovan, J.B., and Moran, J.V. (2015). The Zinc-Finger Antiviral Protein ZAP Inhibits LINE and Alu Retrotransposition. PLoS Genet *11*, e1005121.

Moldovan, J.B., Wang, Y., Shuman, S., Mills, R.E., and Moran, J.V. (2019). RNA ligation precedes the retrotransposition of U6/LINE-1 chimeric RNA. Proc Natl Acad Sci U S A *116*, 20612-20622.

Moran, J.V. (1999). Human L1 retrotransposition: insights and peculiarities learned from a cultured cell retrotransposition assay. Genetica *107*, 39-51.

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

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

Moriarity, B.S., and Largaespada, D.A. (2015). Sleeping Beauty transposon insertional mutagenesis based mouse models for cancer gene discovery. Curr Opin Genet Dev *30*, 66-72.

Morrish, T.A., Garcia-Perez, J.L., Stamato, T.D., Taccioli, G.E., Sekiguchi, J., and Moran, J.V. (2007). Endonuclease-independent LINE-1 retrotransposition at mammalian telomeres. Nature *446*, 208-212.

Morrish, T.A., Gilbert, N., Myers, J.S., Vincent, B.J., Stamato, T.D., Taccioli, G.E., Batzer, M.A., and Moran, J.V. (2002). DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. Nat Genet *31*, 159-165.

Moyes, D., Griffiths, D.J., and Venables, P.J. (2007). Insertional polymorphisms: a new lease of life for endogenous retroviruses in human disease. Trends Genet *23*, 326-333.

Muckenfuss, H., Hamdorf, M., Held, U., Perkovic, M., Lower, J., Cichutek, K., Flory, E., Schumann, G.G., and Munk, C. (2006). APOBEC3 proteins inhibit human LINE-1 retrotransposition. J Biol Chem *281*, 22161-22172.

Munoz-Lopez, M., and Garcia-Perez, J.L. (2010). DNA transposons: nature and applications in genomics. Curr Genomics *11*, 115-128.

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

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

Nakamura, M., Okada, N., and Kajikawa, M. (2012). Self-interaction, nucleic acid binding, and nucleic acid chaperone activities are unexpectedly retained in the unique ORF1p of zebrafish LINE. Mol Cell Biol *32*, 458-469.

Naufer, M.N., Callahan, K.E., Cook, P.R., Perez-Gonzalez, C.E., Williams, M.C., and Furano, A.V. (2016). L1 retrotransposition requires rapid ORF1p oligomerization, a novel coiled coil-dependent property conserved despite extensive remodeling. Nucleic Acids Res *44*, 281-293.

Nigumann, P., Redik, K., Matlik, K., and Speek, M. (2002). Many human genes are transcribed from the antisense promoter of L1 retrotransposon. Genomics *79*, 628-634.

Oettinger, M.A., Schatz, D.G., Gorka, C., and Baltimore, D. (1990). RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination. Science *248*, 1517-1523.

Ohno, S. (1972). So much "junk" DNA in our genome. Brookhaven Symp Biol *23*, 366-370.

Orgel, L.E., and Crick, F.H. (1980). Selfish DNA: the ultimate parasite. Nature *284*, 604-607.

Ostertag, E.M., DeBerardinis, R.J., Goodier, J.L., Zhang, Y., Yang, N., Gerton, G.L., and Kazazian, H.H., Jr. (2002). A mouse model of human L1 retrotransposition. Nat Genet *32*, 655-660.

Ostertag, E.M., Goodier, J.L., Zhang, Y., and Kazazian, H.H., Jr. (2003). SVA elements are nonautonomous retrotransposons that cause disease in humans. Am J Hum Genet *73*, 1444-1451.

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

Ostertag, E.M., Prak, E.T., DeBerardinis, R.J., Moran, J.V., and Kazazian, H.H., Jr. (2000). Determination of L1 retrotransposition kinetics in cultured cells. Nucleic Acids Res *28*, 1418-1423.

Perepelitsa-Belancio, V., and Deininger, P. (2003). RNA truncation by premature polyadenylation attenuates human mobile element activity. Nat Genet *35*, 363-366.

Pezic, D., Manakov, S.A., Sachidanandam, R., and Aravin, A.A. (2014). piRNA pathway targets active LINE1 elements to establish the repressive H3K9me3 mark in germ cells. Genes Dev *28*, 1410-1428.

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

Piskareva, O., Denmukhametova, S., and Schmatchenko, V. (2003). Functional reverse transcriptase encoded by the human LINE-1 from baculovirus-infected insect cells. Protein Expr Purif *28*, 125-130.

Piskareva, O., and Schmatchenko, V. (2006). DNA polymerization by the reverse transcriptase of the human L1 retrotransposon on its own template in vitro. FEBS Lett *580*, 661-668.

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

Ribet, D., Harper, F., Dupressoir, A., Dewannieux, M., Pierron, G., and Heidmann, T. (2008). An infectious progenitor for the murine IAP retrotransposon: emergence of an intracellular genetic parasite from an ancient retrovirus. Genome Res *18*, 597-609.

Richardson, S.R., Doucet, A.J., Kopera, H.C., Moldovan, J.B., Garcia-Perez, J.L., and Moran, J.V. (2015). The Influence of LINE-1 and SINE Retrotransposons on Mammalian Genomes. Microbiol Spectr *3*, MDNA3-0061-2014.

Richardson, S.R., Gerdes, P., Gerhardt, D.J., Sanchez-Luque, F.J., Bodea, G.O., Munoz-Lopez, M., Jesuadian, J.S., Kempen, M.H.C., Carreira, P.E., Jeddeloh, J.A.*, et al.* (2017). Heritable L1 retrotransposition in the mouse primordial germline and early embryo. Genome Res *27*, 1395-1405.

Richardson, S.R., Narvaiza, I., Planegger, R.A., Weitzman, M.D., and Moran, J.V. (2014). APOBEC3A deaminates transiently exposed single-strand DNA during LINE-1 retrotransposition. Elife *3*, e02008.

Robbez-Masson, L., Tie, C.H.C., Conde, L., Tunbak, H., Husovsky, C., Tchasovnikarova, I.A., Timms, R.T., Herrero, J., Lehner, P.J., and Rowe, H.M. (2018). The HUSH complex cooperates with TRIM28 to repress young retrotransposons and new genes. Genome Res *28*, 836-845.

Robertson, K.D. (2005). DNA methylation and human disease. Nat Rev Genet *6*, 597-610.

Rodic, N., Steranka, J.P., Makohon-Moore, A., Moyer, A., Shen, P., Sharma, R., Kohutek, Z.A., Huang, C.R., Ahn, D., Mita, P.*, et al.* (2015). Retrotransposon insertions in the clonal evolution of pancreatic ductal adenocarcinoma. Nat Med *21*, 1060-1064.

Rodin, R.E., Dou, Y., Kwon, M., Sherman, M.A., D'Gama, A.M., Doan, R.N., Rento, L.M., Girskis, K.M., Bohrson, C.L., Kim, S.N.*, et al.* (2021). The landscape of somatic mutation in cerebral cortex of autistic and neurotypical individuals revealed by ultra-deep whole-genome sequencing. Nat Neurosci *24*, 176-185.

Sahakyan, A.B., Murat, P., Mayer, C., and Balasubramanian, S. (2017). G-quadruplex structures within the 3' UTR of LINE-1 elements stimulate retrotransposition. Nat Struct Mol Biol *24*, 243-247.

Sandoval-Villegas, N., Nurieva, W., Amberger, M., and Ivics, Z. (2021). Contemporary Transposon Tools: A Review and Guide through Mechanisms and Applications of Sleeping Beauty, piggyBac and Tol2 for Genome Engineering. Int J Mol Sci *22*.

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

Scott, A.F., Schmeckpeper, B.J., Abdelrazik, M., Comey, C.T., O'Hara, B., Rossiter, J.P., Cooley, T., Heath, P., Smith, K.D., and Margolet, L. (1987). Origin of the human L1 elements: proposed progenitor genes deduced from a consensus DNA sequence. Genomics *1*, 113-125.

Scott, E.C., and Devine, S.E. (2017). The Role of Somatic L1 Retrotransposition in Human Cancers. Viruses *9*.

Shpyleva, S., Melnyk, S., Pavliv, O., Pogribny, I., and Jill James, S. (2018). Overexpression of LINE-1 Retrotransposons in Autism Brain. Mol Neurobiol *55*, 1740-1749.

Sinnett, D., Richer, C., Deragon, J.M., and Labuda, D. (1991). Alu RNA secondary structure consists of two independent 7 SL RNA-like folding units. J Biol Chem *266*, 8675-8678.

Skowronski, J., Fanning, T.G., and Singer, M.F. (1988). Unit-length line-1 transcripts in human teratocarcinoma cells. Mol Cell Biol *8*, 1385-1397.

Skowronski, J., and Singer, M.F. (1985). Expression of a cytoplasmic LINE-1 transcript is regulated in a human teratocarcinoma cell line. Proc Natl Acad Sci U S A *82*, 6050-6054.

Slotkin, R.K., and Martienssen, R. (2007). Transposable elements and the epigenetic regulation of the genome. Nat Rev Genet *8*, 272-285.

Smit, A.F. (1996). The origin of interspersed repeats in the human genome. Curr Opin Genet Dev *6*, 743-748.

Smit, A.F., and Riggs, A.D. (1995). MIRs are classic, tRNA-derived SINEs that amplified before the mammalian radiation. Nucleic Acids Res *23*, 98-102.

Speek, M. (2001). Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes. Mol Cell Biol *21*, 1973-1985.

Spradling, A.C., Bellen, H.J., and Hoskins, R.A. (2011). Drosophila P elements preferentially transpose to replication origins. Proc Natl Acad Sci U S A *108*, 15948-15953.

Stenglein, M.D., and Harris, R.S. (2006). APOBEC3B and APOBEC3F inhibit L1 retrotransposition by a DNA deamination-independent mechanism. J Biol Chem *281*, 16837-16841.

Suter, C.M., Martin, D.I., and Ward, R.L. (2004). Hypomethylation of L1 retrotransposons in colorectal cancer and adjacent normal tissue. Int J Colorectal Dis *19*, 95-101.

Swergold, G.D. (1990). Identification, characterization, and cell specificity of a human LINE-1 promoter. Mol Cell Biol *10*, 6718-6729.

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

Szak, S.T., Pickeral, O.K., Makalowski, W., Boguski, M.S., Landsman, D., and Boeke, J.D. (2002). Molecular archeology of L1 insertions in the human genome. Genome Biol *3*, research0052.

Tangsuwansri, C., Saeliw, T., Thongkorn, S., Chonchaiya, W., Suphapeetiporn, K., Mutirangura, A., Tencomnao, T., Hu, V.W., and Sarachana, T. (2018). Investigation of epigenetic regulatory networks associated with autism spectrum disorder (ASD) by integrated global LINE-1 methylation and gene expression profiling analyses. PLoS One *13*, e0201071.

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

Taylor, M.S., LaCava, J., Dai, L., Mita, P., Burns, K.H., Rout, M.P., and Boeke, J.D. (2016). Characterization of L1-Ribonucleoprotein Particles. Methods Mol Biol *1400*, 311-338.

Taylor, M.S., LaCava, J., Mita, P., Molloy, K.R., Huang, C.R., Li, D., Adney, E.M., Jiang, H., Burns, K.H., Chait, B.T.*, et al.* (2013). Affinity proteomics reveals human host factors implicated in discrete stages of LINE-1 retrotransposition. Cell *155*, 1034-1048.

Tchenio, T., Casella, J.F., and Heidmann, T. (2000). Members of the SRY family regulate the human LINE retrotransposons. Nucleic Acids Res *28*, 411-415.

Telesnitsky, A., and Goff, S.P. (1997). Reverse Transcriptase and the Generation of Retroviral DNA. In Retroviruses, J.M. Coffin, S.H. Hughes, and H.E. Varmus, eds. (Cold Spring Harbor (NY)).

Terry, D.M., and Devine, S.E. (2019). Aberrantly High Levels of Somatic LINE-1 Expression and Retrotransposition in Human Neurological Disorders. Front Genet *10*, 1244.

Thayer, R.E., Singer, M.F., and Fanning, T.G. (1993). Undermethylation of specific LINE-1 sequences in human cells producing a LINE-1-encoded protein. Gene *133*, 273-277.

Thomas, C.A., Tejwani, L., Trujillo, C.A., Negraes, P.D., Herai, R.H., Mesci, P., Macia, A., Crow, Y.J., and Muotri, A.R. (2017). Modeling of TREX1-Dependent Autoimmune Disease using Human Stem Cells Highlights L1 Accumulation as a Source of Neuroinflammation. Cell Stem Cell *21*, 319-331 e318.

Thomas, J., and Pritham, E.J. (2015). Helitrons, the Eukaryotic Rolling-circle Transposable Elements. Microbiol Spectr *3*.

Thompson, C.B. (1995). New insights into V(D)J recombination and its role in the evolution of the immune system. Immunity *3*, 531-539.

Tiwari, B., Jones, A.E., Caillet, C.J., Das, S., Royer, S.K., and Abrams, J.M. (2020). p53 directly represses human LINE1 transposons. Genes Dev *34*, 1439-1451.

Tristan-Ramos, P., Rubio-Roldan, A., Peris, G., Sanchez, L., Amador-Cubero, S., Viollet, S., Cristofari, G., and Heras, S.R. (2020). The tumor suppressor microRNA let-7 inhibits human LINE-1 retrotransposition. Nat Commun *11*, 5712.

Tubio, J.M.C., Li, Y., Ju, Y.S., Martincorena, I., Cooke, S.L., Tojo, M., Gundem, G., Pipinikas, C.P., Zamora, J., Raine, K.*, et al.* (2014). Mobile DNA in cancer. Extensive transduction of nonrepetitive DNA mediated by L1 retrotransposition in cancer genomes. Science *345*, 1251343.

Ullu, E., and Tschudi, C. (1984). Alu sequences are processed 7SL RNA genes. Nature *312*, 171-172.

Upton, K.R., Gerhardt, D.J., Jesuadian, J.S., Richardson, S.R., Sanchez-Luque, F.J., Bodea, G.O., Ewing, A.D., Salvador-Palomeque, C., van der Knaap, M.S., Brennan, P.M.*, et al.* (2015). Ubiquitous L1 mosaicism in hippocampal neurons. Cell *161*, 228-239.

Usdin, K., and Furano, A.V. (1989). The structure of the guanine-rich polypurine:polypyrimidine sequence at the right end of the rat L1 (LINE) element. J Biol Chem *264*, 15681-15687.

van den Hurk, J.A., Meij, I.C., Seleme, M.C., Kano, H., Nikopoulos, K., Hoefsloot, L.H., Sistermans, E.A., de Wijs, I.J., Mukhopadhyay, A., Plomp, A.S.*, et al.* (2007). L1 retrotransposition can occur early in human embryonic development. Hum Mol Genet *16*, 1587-1592.

van den Hurk, J.A., van de Pol, D.J., Wissinger, B., van Driel, M.A., Hoefsloot, L.H., de Wijs, I.J., van den Born, L.I., Heckenlively, J.R., Brunner, H.G., Zrenner, E.*, et al.* (2003). Novel types of mutation in the choroideremia (CHM) gene: a full-length L1 insertion and an intronic mutation activating a cryptic exon. Hum Genet *113*, 268-275.

Virgen, C.A., Kratovac, Z., Bieniasz, P.D., and Hatziioannou, T. (2008). Independent genesis of chimeric TRIM5-cyclophilin proteins in two primate species. Proc Natl Acad Sci U S A *105*, 3563-3568.

Volkmann, B., Wittmann, S., Lagisquet, J., Deutschmann, J., Eissmann, K., Ross, J.J., Biesinger, B., and Gramberg, T. (2020). Human TRIM5alpha senses and restricts LINE-1 elements. Proc Natl Acad Sci U S A *117*, 17965-17976.

Wallace, N., Wagstaff, B.J., Deininger, P.L., and Roy-Engel, A.M. (2008). LINE-1 ORF1 protein enhances Alu SINE retrotransposition. Gene *419*, 1-6.

Walter, P., and Blobel, G. (1983). Subcellular distribution of signal recognition particle and 7SL-RNA determined with polypeptide-specific antibodies and complementary DNA probe. J Cell Biol *97*, 1693-1699.

Wang, Y., Bae, T., Thorpe, J., Sherman, M.A., Jones, A.G., Cho, S., Daily, K., Dou, Y., Ganz, J., Galor, A.*, et al.* (2021). Comprehensive identification of somatic nucleotide variants in human brain tissue. Genome Biol *22*, 92.

Ward, J.R., Vasu, K., Deutschman, E., Halawani, D., Larson, P.A., Zhang, D., Willard, B., Fox, P.L., Moran, J.V., and Longworth, M.S. (2017). Condensin II and GAIT complexes cooperate to restrict LINE-1 retrotransposition in epithelial cells. PLoS Genet *13*, e1007051.

Warkocki, Z., Krawczyk, P.S., Adamska, D., Bijata, K., Garcia-Perez, J.L., and Dziembowski, A. (2018). Uridylation by TUT4/7 Restricts Retrotransposition of Human LINE-1s. Cell *174*, 1537-1548 e1529.

Watanabe, T., Totoki, Y., Toyoda, A., Kaneda, M., Kuramochi-Miyagawa, S., Obata, Y., Chiba, H., Kohara, Y., Kono, T., Nakano, T.*, et al.* (2008). Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. Nature *453*, 539-543.

Weber, M.J. (2006). Mammalian small nucleolar RNAs are mobile genetic elements. PLoS Genet *2*, e205.

Wei, W., Gilbert, N., Ooi, S.L., Lawler, J.F., Ostertag, E.M., Kazazian, H.H., Boeke, J.D., and Moran, J.V. (2001). Human L1 retrotransposition: cis preference versus trans complementation. Mol Cell Biol *21*, 1429-1439.

Weichenrieder, O., Stehlin, C., Kapp, U., Birse, D.E., Timmins, P.A., Strub, K., and Cusack, S. (2001). Hierarchical assembly of the Alu domain of the mammalian signal recognition particle. RNA *7*, 731-740.

Wells, J.N., and Feschotte, C. (2020). A Field Guide to Eukaryotic Transposable Elements. Annu Rev Genet *54*, 539-561.

Wilson, R.C., and Doudna, J.A. (2013). Molecular mechanisms of RNA interference. Annu Rev Biophys *42*, 217-239.

Wimmer, K., Callens, T., Wernstedt, A., and Messiaen, L. (2011). The NF1 gene contains hotspots for L1 endonuclease-dependent de novo insertion. PLoS Genet *7*, e1002371.

Wissing, S., Montano, M., Garcia-Perez, J.L., Moran, J.V., and Greene, W.C. (2011). Endogenous APOBEC3B restricts LINE-1 retrotransposition in transformed cells and human embryonic stem cells. J Biol Chem *286*, 36427-36437.

Wissing, S., Munoz-Lopez, M., Macia, A., Yang, Z., Montano, M., Collins, W., Garcia-Perez, J.L., Moran, J.V., and Greene, W.C. (2012). Reprogramming somatic cells into iPS cells activates LINE-1 retroelement mobility. Hum Mol Genet *21*, 208-218.

Xing, J., Wang, H., Belancio, V.P., Cordaux, R., Deininger, P.L., and Batzer, M.A. (2006). Emergence of primate genes by retrotransposon-mediated sequence transduction. Proc Natl Acad Sci U S A *103*, 17608-17613.

Xing, J., Zhang, Y., Han, K., Salem, A.H., Sen, S.K., Huff, C.D., Zhou, Q., Kirkness, E.F., Levy, S., Batzer, M.A.*, et al.* (2009). Mobile elements create structural variation: analysis of a complete human genome. Genome Res *19*, 1516-1526.

Xiong, Y., and Eickbush, T.H. (1988). Similarity of reverse transcriptase-like sequences of viruses, transposable elements, and mitochondrial introns. Mol Biol Evol *5*, 675-690.

Xiong, Y., and Eickbush, T.H. (1990). Origin and evolution of retroelements based upon their reverse transcriptase sequences. EMBO J *9*, 3353-3362.

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

Yang, N., Zhang, L., Zhang, Y., and Kazazian, H.H., Jr. (2003). An important role for RUNX3 in human L1 transcription and retrotransposition. Nucleic Acids Res *31*, 4929-4940.

Yang, P., Wang, Y., and Macfarlan, T.S. (2017). The Role of KRAB-ZFPs in Transposable Element Repression and Mammalian Evolution. Trends Genet *33*, 871-881.

Yang, Y.G., Lindahl, T., and Barnes, D.E. (2007). Trex1 exonuclease degrades ssDNA to prevent chronic checkpoint activation and autoimmune disease. Cell *131*, 873-886.

Yu, F., Zingler, N., Schumann, G., and Stratling, W.H. (2001). Methyl-CpG-binding protein 2 represses LINE-1 expression and retrotransposition but not Alu transcription. Nucleic Acids Res *29*, 4493-4501.

Yuan, Y.W., and Wessler, S.R. (2011). The catalytic domain of all eukaryotic cut-and-paste transposase superfamilies. Proc Natl Acad Sci U S A *108*, 7884-7889.

Zhao, B., Wu, Q., Ye, A.Y., Guo, J., Zheng, X., Yang, X., Yan, L., Liu, Q.R., Hyde, T.M., Wei, L.*, et al.* (2019). Somatic LINE-1 retrotransposition in cortical neurons and non-brain tissues of Rett patients and healthy individuals. PLoS Genet *15*, e1008043.

Zhao, K., Du, J., Han, X., Goodier, J.L., Li, P., Zhou, X., Wei, W., Evans, S.L., Li, L., Zhang, W.*, et al.* (2013). Modulation of LINE-1 and Alu/SVA retrotransposition by Aicardi-Goutieres syndrome-related SAMHD1. Cell Rep *4*, 1108-1115.

Zhou, W., Emery, S.B., Flasch, D.A., Wang, Y., Kwan, K.Y., Kidd, J.M., Moran, J.V., and Mills, R.E. (2020). Identification and characterization of occult human-specific LINE-1 insertions using long-read sequencing technology. Nucleic Acids Res *48*, 1146-1163.

Zhu, X., Zhou, B., Pattni, R., Gleason, K., Tan, C., Kalinowski, A., Sloan, S., Fiston-Lavier, A.S., Mariani, J., Petrov, D.*, et al.* (2021). Machine learning reveals bilateral distribution of somatic L1 insertions in human neurons and glia. Nat Neurosci *24*, 186-196.

Zhu, Y., Chen, G., Lv, F., Wang, X., Ji, X., Xu, Y., Sun, J., Wu, L., Zheng, Y.T., and Gao, G. (2011). Zinc-finger antiviral protein inhibits HIV-1 infection by selectively targeting multiply spliced viral mRNAs for degradation. Proc Natl Acad Sci U S A *108*, 15834-15839.

Chapter 2

A Genome-Wide Screen Identifies Cellular Factors that Mediate Silencing of L1-Delivered Reporter Genes in PA-1 Human Embryonic Carcinoma Cells

The initial experiments and analyses for GeCKO screen 1 were performed in collaboration with Dr. Peter Larson and Dr. Jacob Kitzman. Illumina sequencing was performed in collaboration with the Kitzman laboratory and the University of Michigan Sequencing CORE. Viral packaging was performed by the University of Michigan Vector CORE. I performed all other experiments and analyses discussed in this chapter.

Abstract

Long Interspersed Element-1 (LINE-1 or L1) comprises approximately 17% of human genomic DNA. L1 can mobilize (*i.e.*, retrotranspose) in the germline, during early development, and in select somatic cells. The resultant insertions can alter gene expression, generate structural variation, and create pathogenic mutations. The average human genome contains ~100 retrotransposition-competent L1s that can “copy and paste” themselves into new genomic locations by a mechanism termed target-primed reverse transcription (TPRT). Given the mutagenic potential of L1 retrotransposition, it is no surprise that a variety of cellular mechanisms have evolved to restrict each stage of the L1 retrotransposition cycle.

Previous studies established that human embryonic carcinoma-derived cell lines (hECs) differ from many somatic-derived cell lines in their ability to epigenetically silence reporter genes integrated into the genome by L1 retrotransposition. Specifically, L1 retrotransposition-delivered reporter genes were efficiently and stably silenced in hECs upon TPRT-mediated genomic integration, a process we have termed L1-REPEL (L1-delivered REPorter gEne siLencing). Interestingly, histone deacetylase inhibitors (HDACi) reversed L1-REPEL in hECs, suggesting an epigenetic silencing mechanism dependent upon changes in chromatin status at the L1 integration site. Here, we designed and implemented an unbiased genome-wide CRISPR/Cas9-based genetic knockout screen to identify host factors that may facilitate L1-REPEL in hECs. We identify 489 candidate L1-REPEL factors for subsequent investigation, including the top candidates neurofibromin 2 (*NF2)* and Exportin 7 (*XPO7).* Our results provide insight into L1 biology and the L1-REPEL silencing mechanism in human embryonic cells.

Introduction

Long Interspersed Element-1 (LINE-1 or L1) is an endogenous non-LTR retrotransposon that has proliferated throughout mammalian evolution to comprise approximately 17% of human genomic DNA (Lander et al., 2001). Human L1s are ~6 kb in length and contain a 5'-untranslated region (UTR) with RNA polymerase II activity (Athanikar et al., 2004; Olovnikov et al., 2007; Swergold, 1990). A retrotransposition-competent L1 encodes two open reading frames (ORF1 and ORF2) separated by a 63 bp inter-ORF spacer (Dombroski et al., 1991; Scott et al., 1987). Following ORF2 is a 3'-UTR and poly (A) tail (Dombroski et al., 1991; Scott et al., 1987). L1 insertions are typically flanked by short (4-16 bp) target-site duplications within genomic DNA. The

vast majority L1 sequences in the human genome are inactive due to 5' truncations or inversion deletion structures generated during genomic integration and/or the accumulation of deleterious point mutations over time (Grimaldi et al., 1984; Ostertag and Kazazian, 2001).

L1s can mobilize (*i.e.*, retrotranspose) in the germline, during early development, and in select somatic cells (Coufal et al., 2009; Faulkner and Billon, 2018; Faulkner and Garcia-Perez, 2017; Garcia-Perez et al., 2007b; Garcia-Perez et al., 2010; Kano et al., 2009; Kazazian, 2004; Kubo et al., 2006; Muotri et al., 2005; Richardson et al., 2017; Scott and Devine, 2017). The L1-encoded proteins (ORF1p and ORF2p) are necessary for L1 retrotransposition and preferentially associate with their encoding RNA by a process termed *cis* preference (Doucet et al., 2015; Feng et al., 1996; Kulpa and Moran, 2005, 2006; Martin et al., 2005; Moran et al., 1996; Wei et al., 2001). ORF2p can be "hijacked" by other cellular RNAs to mediate their mobility in *trans* (Dewannieux et al., 2003; Garcia-Perez et al., 2007a; Gilbert et al., 2005; Hancks et al., 2011; Moldovan et al., 2019; Weber, 2006; Wei et al., 2001). L1 mediated retrotransposition events can alter gene expression, generate structural variation, and create pathogenic mutations (Beck et al., 2011; Gilbert et al., 2005; Gilbert et al., 2002; Kazazian and Moran, 2017; Kazazian et al., 1988; Richardson et al., 2015; Scott and Devine, 2017; Solyom et al., 2012; Symer et al., 2002).

The average human genome contains ~100 retrotransposition-competent L1s that can "copy and paste" themselves into new genomic locations by a mechanism termed target-primed reverse transcription (TPRT) (Beck et al., 2010; Brouha et al., 2003; Cost et al., 2002; Feng et al., 1996; Kulpa and Moran, 2006; Luan et al., 1993; Sassaman et

al., 1997). TPRT is unique to non-LTR retrotransposons and differs from integration mechanisms used by LTR retrotransposons, DNA transposons, and retroviruses (Beck et al., 2011; Feschotte and Pritham, 2007; Kazazian and Moran, 2017; Lewinski and Bushman, 2005; Luan et al., 1993; Schorn et al., 2017). Although the mechanism of L1 retrotransposition requires elucidation, it is clear that at least two ORF2p enzymatic activities are utilized for TPRT. One is an endonuclease activity that generates a single-strand nick at the consensus 5'-TTTTT/AA-3' site of genomic integration (Cost et al., 2002; Feng et al., 1996; Flasch et al., 2019; Jurka, 1997; Morrish et al., 2002). The other is an RNA-dependent DNA polymerase activity that utilizes the 3'OH of the nicked genomic DNA as a primer to initiate reverse transcription of the L1 RNA (Cost et al., 2002; Dombroski et al., 1994; Doucet et al., 2015; Kulpa and Moran, 2006; Luan et al., 1993; Mathias et al., 1991; Monot et al., 2013; Piskareva et al., 2003; Piskareva and Schmatchenko, 2006).

Given the mutagenic potential of L1 retrotransposition, it is no surprise that a variety of cellular mechanisms have evolved to restrict L1 mobilization (Goodier, 2016; Levin and Moran, 2011). Previous studies established that human embryonic carcinoma-derived cell lines (hECs) differ from many somatic cancer cell lines in their ability to epigenetically silence reporter genes delivered by L1 retrotransposition (Garcia-Perez et al., 2010). For example, PA-1 hECs cells are a euploid human ovarian teratocarcinoma-derived cell line containing a single reciprocal translocation between chromosomes 15 and 20 (Sarraf et al., 2005; Zeuthen et al., 1980). PA-1 cells, like many hEC cell lines, express early developmental gene profiles similar to human embryonic stem cells (Abu Dawud et al., 2012; Sperger et al., 2003). Additionally, endogenous L1s are efficiently

expressed in human embryonic cell lines (Garcia-Perez et al., 2007b; Garcia-Perez et al., 2010; Hohjoh and Singer, 1996; Leibold et al., 1990). However, despite high levels of expression, engineered L1 elements are efficiently and stably silenced upon retrotransposition in hEC cells, which has uncovered an additional mechanism that restricts L1 retrotransposition in cell lines that serve as proxies for early human development (Garcia-Perez et al., 2010).

Previous studies demonstrated that an *EGFP* retrotransposition-indicator cassette (*mEGFPI)* delivered by a human L1 synthetic mouse L1, and zebrafish L2 (a L1 like element) are efficiently silenced in PA-1 cells (Garcia-Perez et al., 2010). However, similar selectable or screenable reporter genes were not efficiently silenced when integrated into PA-1 genomic DNA using stable transfection or retroviral vectors, suggesting that structures generated during TPRT may play a role in LINE-reporter gene mediated silencing (Garcia-Perez et al., 2010). These results established that L1-delivered reporter genes are efficiently and stably silenced in hECs upon TPRT-mediated genomic integration, a process we have termed L1-REPEL (L1-delivered REPorter gEne siLencing).

Interestingly, class I histone deacetylase inhibitors (HDACi) reversed L1-REPEL in a clonal PA-1-derived cell line (*i.e.*, pk5 cells) harboring a single, silenced L1-delivered reporter gene (*EGFP*), suggesting that reporter gene reactivation was dependent upon changes in chromatin status at the L1 integration site. Subsequent removal of HDACi resulted in re-establishment of L1-REPEL over time. Thus, the mechanism of L1-REPEL in PA-1 cells exhibits the characteristics of an epigenetic system: (1) silencing is mitotically stable; (2) silencing does not alter the primary DNA sequence (3) silencing is

reversible; and (4) silencing is reestablished after drug treatment, suggesting an epigenetic mark, or memory, to the system. Thus, L1-delivered reporter genes are efficiently and stably silenced in hECs upon TPRT-mediated genomic integration (Garcia-Perez et al., 2010).

To investigate L1-REPEL in PA-1 cells, we took advantage of recent advancements in CRISPR/Cas9-based gene knockout technology (Cong et al., 2013; Hsu et al., 2014; Jinek et al., 2012; Mali et al., 2013; Pyzocha et al., 2014). We hypothesized that cellular factors are necessary to initiate and/or maintain L1-REPEL in PA-1 cells. Thus, we employed a genome-wide scale (GeCKO) CRISPR/Cas9-based genome editing platform to identify genes that may play a role in L1-REPEL (Sanjana et al., 2014; Shalem et al., 2014).

The GeCKO platform is based on lentiviral delivery of Cas9 and a single-guide RNA (sgRNA). The sgRNA-guided Cas9 nuclease is targeted to genic coding regions, where it introduces a double-strand DNA break, which upon DNA repair, can result in a knockout (*i.e.*, the generation of a null mutation) in either one or both alleles of a target gene. We reasoned that knockout of genes involved in L1-REPEL would permit expression of reporter genes integrated into PA-1 genomic DNA via L1 retrotransposition.

We implemented the GeCKO system in PA-1 cells to knockout individual genes in a high-throughput manner. We then subjected the population of GeCKO-treated cells to L1 retrotransposition assays, using various retrotransposition indicator cassettes (Moran et al., 1996; Morrish et al., 2002; Ostertag et al., 2000; Wei et al., 2000), to identify cells expressing the L1 delivered reporter gene. Identification of the sgRNA-targeted genes

within cells that expressed the L1 delivered reporter gene allowed us to infer cellular factors that may mediate L1-REPEL. Through this screen, we identified 489 factors that may be necessary to establish and/or maintain L1-REPEL. My preliminary results suggest that knockout of our top candidate genes, *NF2*, *XPO7*, *TADA2B* and *ZC3H12B*, may be necessary for efficient L1-REPEL in PA-1 cells. Together, our data demonstrate that a genome-wide CRISPR/Cas9 gene knockout system was able to identify human factors necessary for efficient L1-REPEL in PA-1 cells, providing insight into L1 biology and the L1-REPEL mechanism in human embryonic cells.

Results

PA-1 human embryonic carcinoma cells exhibit L1-REPEL

To confirm that L1 delivered reporter genes are subject to L1-REPEL in PA-1 hECs (Garcia-Perez et al., 2010), we performed L1 retrotransposition assays using engineered L1 expression vectors containing three different reporter gene retrotransposition indicator cassettes (*mneoI*, *mblastI*, and *mEGFPI*) (Moran et al., 1996; Morrish et al., 2002; Ostertag et al., 2000; Wei et al., 2000). Each retrotransposition indicator cassette is cloned into the L1 3' UTR to not disrupt ORF1p and ORF2p expression. The retrotransposition indicator cassettes consist of an antisense selectable (*NEO* or *blasticidin*) or screenable (*EGFP*) reporter gene whose transcription is driven by an exogenous promoter and terminates at a heterologous polyadenylation signal (Moran et al., 1996; Morrish et al., 2002; Ostertag et al., 2000; Wei et al., 2000). The reporter gene also is interrupted by an intron cloned in the same transcriptional orientation as the L1 sequence. This arrangement ensures that a functional reporter gene only is expressed after a successful round of retrotransposition,

where the L1 is transcribed from an episomal expression vector (pCEP4), the intron is spliced from sense strand L1 mRNA containing the “backward” reporter gene, and L1 is integrated into genomic DNA by TPRT (Kopera et al., 2016b; Moran et al., 1996; Morrish et al., 2002; Ostertag et al., 2000; Wei et al., 2001). Thus, reporter gene expression (*i.e.*, the resultant number of G418- or blasticidin-resistant foci or EGFP-positive cells) provides a quantitative readout of L1 retrotransposition.

We first transiently transfected PA-1 hEC cells with pJM101/L1.3, a pCEP4 episomal expression vector that contains a retrotransposition-competent L1 (L1.3) (Sassaman et al., 1997) marked with the *mneoI* retrotransposition indicator cassette, to assay for retrotransposition (Kopera et al., 2016b; Moran et al., 1996; Wei et al., 2001). In agreement with previous reports (Garcia-Perez et al., 2010), we did not observe G418-resistant foci in PA-1 hECs (Figure 2.1A: top panel 1). However, G418-resistant foci readily were observed when PA-1 cells were transfected with a pCDNA3 expression vector that constitutively expresses a neomycin phosphotransferase resistance gene (Figure 2.1A: top panel 2).

Similar results were found for PA-1 cells transiently transfected with pJJ101/L1.3, a pCEP4 episomal expression vector that contains a retrotransposition-competent L1 (L1.3) (Sassaman et al., 1997) marked with the *mblastI* retrotransposition indicator cassette (Figure 2.1A: top panel 3) (Goodier et al., 2007; Morrish et al., 2002). Again, control experiments revealed blasticidin-resistant foci when PA-1 cells were transfected with a pCDNA6 expression vector that constitutively expresses a blasticidin resistance gene (Figure 2.1A: top panel 4).

Finally, L1-REPEL was observed in PA-1 cells transfected with p99EGFP/LRE3, a pCEP4-based episomal expression vector that contains a retrotransposition-competent L1 (LRE3) (Brouha et al., 2002) marked with the *mEGFPI* retrotransposition indicator cassette (Ostertag et al., 2000). We only observed low levels of EGFP expressing cells (~0.3% by flow cytometry) 7-days post-transfection (Figure 2.1A: top panel 5).

Notably, for in experiments performed with each of the above retrotransposition indicator cassettes, we performed parallel experiments with somatic cervical carcinoma-derived HeLa JVM cells. As expected, we observed robust reporter gene expression in the L1-*neo*, L1-*blast* and L1-*GFP* retrotransposition assays (Figure 2.1A: bottom panels) (Garcia-Perez et al., 2010).

To confirm that L1 delivered reporter genes are epigenetically silenced in hECs we treated a clonal PA-1-derived cell line, harboring a full-length silenced L1-*GFP* retrotransposition event (pk5 cells), with a pan histone deacetylase inhibitor, trichostatin A (TSA), and then assessed whether this treatment led to reactivation of EGFP expression (Garcia-Perez et al., 2010). As expected, TSA treatment induced EGFP expression in ~80% of pk5 cells after 18 hours (Figure 2.1 B), demonstrating that TSA treatment could efficiently reverse L1-REPEL in pk5 cells. Subsequent removal of TSA from the cell culture media resulted in complete re-establishment of *EGFP* silencing after 5 days (Figure 2.1 C), suggesting epigenetic memory to the mitotically stable silencing mechanism. Together, these results suggest that L1-delivered reporter genes are epigenetically silenced upon retrotransposition in hEC cells, a mechanism we have labeled L1-REPEL (Figure 2.1 D).

Overview of a genome-wide CRISPR/Cas9 strategy to identify L1-REPEL factors

We hypothesized that cellular proteins are necessary for L1-REPEL in hEC cells. To investigate this hypothesis, we utilized an unbiased genome-wide CRISPR/Cas9-based gene knockout system (Sanjana et al., 2014; Shalem et al., 2014), in conjunction with the L1 retrotransposition assay, to identify host factors that may mediate L1-REPEL (Figure 2.2 B). Briefly, hECs we transduced with lentiviral vectors that express the Cas9 nuclease, a single-guide RNA (sgRNA), and puromycin resistance. The sgRNA-guided Cas9 nuclease is targeted to genic coding regions, where it introduces a double-strand DNA break, which upon repair, can generate mutations that may result in a loss-of-function allele. We then performed the L1 retrotransposition assay, where the puromycin-resistant hECs expressing Cas9 and a sgRNA were transfected with an engineered human L1 vector to identify cells that expressed the L1-delivered reporter gene. Genomic DNAs from the population of drug-resistant cells then were harvested to identify sgRNAs within cells that escaped L1-REPEL via next generation sequencing. We then use Model-based Analysis of Genome-wide CRISPR-Cas9 Knockout (MAGeCK) and sgRNA counts to infer statistically enriched genes that may facilitate L1-REPEL.

Implementation of the GeCKO system

The GeCKO system is based on lentiviral delivery of Cas9 and a single-guide RNA (sgRNA). The GeCKO system can be delivered by a single lentiviral vector (lentiCRISPRv2) or though serial infection of two independent lentiviral vectors (lentiCas9-Blast and lentiGuide-Puro) (Figure 2.2 A). Notably, lentiCRISPRv2 and lentiGuide-Puro each contain the puromycin N-acetyltransferase (pac) gene that confers

resistance to the antibiotic puromycin (Sanjana et al., 2014; Shalem et al., 2014). The sgRNA libraries are cloned into lentiCRISPRv2 or lentiGuide-Puro to generate a pool of lentiviral vectors, where each virus contains a single sgRNA sequence from the sgRNA library.

We utilized two different sgRNA libraries. First, the GeCKOv2 sgRNA library contains 123,411 guide sequences (sgRNAs) divided equally between two libraries (A and B), which are designed to target 19,050 human genes and 1,864 miRNAs; 2000 non-targeting control sgRNAs are included in the library as controls. Importantly, the GeCKOv2 library has 6 sgRNAs per gene and 4 sgRNAs per miRNA (Sanjana et al., 2014). Second, the Brunello sgRNA library contains 76,441 sgRNAs targeting 19,114 human genes, along with 1000 non-targeting control sgRNAs. Notably, the Brunello library is more optimized and has 4 sgRNAs per gene (Doench et al., 2016).

To ensure efficient cloning and coverage of sgRNAs into lentiviral vectors, we sequenced the sgRNA-containing lentiCRISPRv2 vectors on the Illumina Hi-Seq 2000 platform (performed by the University of Michigan Sequencing Core). We verified that 99.13% of the GeCKOv2 “A” sgRNAs were represented in the final vector pool (performed in collaboration with the Kitzman Laboratory), suggesting efficient amplification and cloning of sgRNAs into the lentiviral vectors. LentiCRISPRv2 “A” and “B” vectors, along with lentiCa9-Blast and lentiGuide-Puro “Brunello” vectors, were separately packaged into a self-inactivating lentivirus expressing the vesicular stomatitis virus glycoprotein (VSV-G) by the University of Michigan Vector Core.

We first aimed to determine a timeframe for efficient gene editing in transduced PA-1 cells. Previous studies suggest that efficient CRISPR/Cas9 editing of human genes

requires ~12 cell doublings (Cross et al., 2016; Sanjana et al., 2014; Shalem et al., 2014). We reasoned that sgRNAs targeting critical cell survival genes would be lost over time due to cell death (Blomen et al., 2015; Hart et al., 2014; Hart et al., 2015) and that we could infer overall editing efficiency within our population of transduced PA-1 cells by identifying the dropout of sgRNAs targeting essential genes.

We first transduced lentiCRISPRv2 containing the GeCKOv2 sgRNA library into PA-1 cells. We infected $1.2x10^8$ PA-1 cells at a multiplicity of infection (MOI) of 0.35 to ensure that each cell was infected by only one viral particle. This strategy ensured that each sgRNA had sufficient representation (~340x) in the transduced population of cells. Cultures of cells transduced with the "A" and "B" GeCKOv2 sgRNA libraries were maintained separately. Transduced PA-1 cells were cultured in puromycin-supplemented media for 31 days. Transduced PA-1 cells were collected at 7-, 14-, 21-, 24-, 28- and 31-days post-transduction. At each timepoint, cells were re-plated at high-density to maintain representation of sgRNAs.

Cas9 is expressed in LentiCRISPRv2 transduced PA-1 cells

To determine whether Cas9 was efficiently expressed in transduced PA-1 cells, we performed western blot analyses using whole cell lysates derived from puromycin-resistant lentiCRISPRv2 transduced PA-1 cells at each timepoint post-transduction (Day 7, 14, 21, 24, 28, 31). We used an anti-FLAG mouse monoclonal antibody (Sigma-Aldrich, F1804) to detect FLAG-tagged Cas9 protein expression from lentiCRISPRv2. An ~160 kD band representing Cas9 protein was detected in whole cell lysates at each timepoint (Figure 2.2 C). These data demonstrate efficient expression of lentiviral-delivered Cas9 protein for at least 31 days post-transduction in PA-1 cells. Furthermore,

in agreement with previous studies (Garcia-Perez et al., 2010), these data demonstrate that both Cas9 and puromycin resistance are continuously expressed over 31 days, indicating that lentiviral-delivered sequences are not silenced in PA-1 cells.

Efficient CRISPR/Cas9 editing occurs 21 days post-transduction in PA-1 cells

To determine editing efficiency in PA-1 cells, we analyzed sgRNA representation at multiple timepoints post-transduction. Genomic DNA was collected and sgRNA sequences were PCR amplified for day 7, day 21, and day 31 timepoints post-transduction. The first round of PCR uses primers that flank the sgRNA within lentiCRISPRv2. The second round of PCR adds Illumina barcodes that allows simultaneous multiplex sequencing. PCR products from each timepoint were sequenced using the Illumina MiSeq platform. We identified 45,599 unique sgRNAs at timepoint day 7, 49,017 unique sgRNAs at timepoint day 21, and 51,124 unique sgRNAs at timepoint day 31.

To determine dropout of sgRNAs, we compared sgRNA representation at each timepoint to the initial sgRNA representation within the lentiCRISPRv2 “A” vector pool. We used a pipeline called model-based analysis of genome-wide CRISPR/Cas9 knockout (MAGeCK) (Li et al., 2014) to analyze sgRNA dropout and to identify depleted signaling pathways at each timepoint. We looked at several essential KEGG signaling pathways to determine dropout of essential genes (Figure 2.2 D). At timepoint day 7, we did not find significant ($p<1X10^{-4}$) depletion of genes associated with the KEGG ribosome, proteasome, spliceosome, or RNA polymerase pathways. At timepoint day 21, we found significant depletion of genes associated with the KEGG_RIBSOME (42/88), KEGG_PROTEASOME (17/46), KEGG_RNA_POLYMERASE (9/29), and

KEGG_SPLICESOME (44/127) pathways (Figure 2.2 D). Similarly, at day 31, we found significant depletion of genes associated with the KEGG_RIBSOME (38/88), KEGG_PROTEASOME (18/46), KEGG_RNA_POLYMERASE (10/29), and KEGG_SPLICESOME (54/127) pathways (Figure 2.2 D). These data demonstrate that 21 days after lentiCRISPRv2 transduction, many genes within pathways critical for cell survival effectively contain null mutations.

To help identify true candidate genes, we implemented the GeCKO system using an alternative sgRNA library (Brunello). We reasoned that candidate genes identified using both GeCKOv2 and Brunello sgRNA libraries were more likely to be true candidate genes. We delivered the Brunello sgRNA library into PA-1 cells using the two-vector lentiviral system. First, we transduced lentiCas9-Blast into $8x10^6$ PA-1 cells at a multiplicity of infection (MOI) of 0.35. Transduced cells were cultured in blasticidin-supplemented media for 5 days. Subsequently, we transduced lentiGuide-Puro, containing the Brunello sgRNA library, into $7.6x10^7$ blasticidin-resistant PA-1 cells at a multiplicity of infection (MOI) of 0.35. Our strategy ensured that each sgRNA had sufficient representation (~500x) in the transduced population of cells. Cells were cultured in puromycin-supplemented media for 21 days. Blasticidin/puromycin-resistant PA-1 cells were collected and cryo-frozen 21 days post-transduction for future analyses. We found significant depletion of genes associated with the KEGG_RIBSOME (28/88), KEGG_PROTEASOME (15/46), KEGG_RNA_POLYMERASE (14/29), and KEGG_SPLICESOME (47/127) pathways, suggesting efficient gene editing 21 days post-transduction (Figure 2.2 D).

L1-REPEL screen in PA-1 cells

To screen for factors necessary to initiate and/or maintain L1-REPEL in PA-1 cells, we performed the L1 retrotransposition assay in day 21 transduced cells containing either the GeCKOv2 (21G) or Brunello (21B) sgRNA libraries. Although both day-21 and day-31 cells exhibited efficient editing, using the earlier timepoint reduces the potential for off-target effects caused by prolonged sgRNA/Cas9 expression (Sanjana et al., 2014; Shalem et al., 2014).

For GeCKOv2-based experiments, we combined 21G cells from both the “A” and “B” libraries. Approximately $8.4x10^6$ d21-GeCKO cells were plated in 15 cm plates and subjected to the L1-*neo* retrotransposition assay (see methods). We performed two independent GeCKOv2-based screens: GeCKO screen 1; containing 10 biological replicates of cells transfected with pJM101/L1.3; and GeCKO screen 2; containing 7 biological replicates of cells transfected with pJM101/L1.3. As a transfection control, one 15 cm plate was transfected with pCDNA3, a vector that constitutively expresses a neomycin resistance expression cassette. As a negative control, we transfected an additional 15 cm plate with pJM105/L1.3, which contains an inactivating missense mutation in the ORF2p reverse transcriptase domain that abolishes L1 retrotransposition in *cis*, to ensure the efficacy of G418 selection. Notably, if the knockout of a particular gene results in G418-resistance independent of L1-delivered reporter gene expression, we would expect to see G418-resistant foci formation in cells transfected with pJM105/L1.3.

For Brunello-based experiments, approximately $8x10^6$ 21B cells were plated and subjected to the L1-*neo* retrotransposition assay. The Brunello screen contained 5

biological replicates of cells transfected with pJM101/L1.3. To generate our control population for enrichment analysis, 5 additional 15 cm plates were transfected with pCDNA3, a vector that constitutively expresses a neomycin resistance expression cassette. To determine if lentiviral transduction alone influenced L1-REPEL, transduced PA-1 cells lacking a sgRNA library (lenti-Cas9-Blast only) were transfected with pJM101/L1.3 and subjected to the L1-*neo* retrotransposition assay in parallel.

Approximately 3 days post-transfection, 200 µg/ml G418 was added to PA-1 culture media to select for cells expressing the L1-delivered reporter gene. After 14 days in G418-media, all of the untransfected and pJM105/L1.3 transfected cells had died, indicating efficient G418 selection. By comparison, d21-GeCKO and d21-Brunello cells transfected with pJM101/L1.3 contained ~300-500 visible G418-resistant foci per 15 cm plate (Figure 2.3 A: top). These colonies represent possible L1-*neo* retrotransposition events that escaped L1-REPEL due to the knockout of a cellular gene. In contrast, transduced cells lacking sgRNAs that were transfected with pJM101/L1.3 presented ~40 G418-resistant foci (Figure 2.3 A: bottom). These results indicate that there is a low level of background G418-resistant foci that apparently escape L1-REPEL independent of gene knockout. This level of background is expected given that previous experiments demonstrated that a low-level of PA-1 cells escape L1-REPEL in the L1-*neo* based (*e.g.*, ~ 1-2 G418-resistant colonies per well of 6-well plate) and L1-*GFP* based (~0.3% EGFP positive cells) cultured cell retrotransposition assays (Garcia-Perez et al., 2010).

For GeCKO screens 1 and 2, the resultant G418-resistant foci were pooled from each 15 cm plate, genomic DNA was prepared from the pooled foci, and sgRNAs were PCR amplified and sequenced as described above. Surprisingly, sequencing and

analysis of sgRNA profiles revealed that each biological replicate contained ~3,000-5,000 unique sgRNA sequences. Interestingly, the top ~500 sgRNAs represented the majority (~50-70%) of the total reads per replicate (discussed below). Thus, we suspect that sgRNAs within large visible colonies comprise the majority of our reads and that small colonies or even single cells contribute low sgRNA read counts.

For the Brunello screen, G418-resistant foci were pooled from each replicate and replated in 15 cm plates. Based on results from the GeCKO experiments, we reasoned that replating the G418-resistant foci would help reduce the background of sgRNAs that could be attributed to non-colony forming G418-resistant cells. Three days after replating, the resultant G418-resistant cells were collected, genomic DNA was prepared from the cells, and sgRNAs were PCR amplified and sequenced as described above. Sequencing and analysis of sgRNA profiles revealed that each biological replicate contained ~1,500 unique sgRNA sequences, with the top ~500 sgRNAs representing the vast majority (~95%) of total reads per replicate. Thus, replating G418-resistant cells likely reduced the number of small colonies or even single cells that were no longer growing, diminishing low read count sgRNAs.

Identification of L1-REPEL candidate genes

To identify enriched genes within each L1-REPEL screen, we first employed model-based analysis of genome-wide CRISPR/Cas9 knockout (MAGeCK) (Li et al., 2014). Briefly, sgRNA read counts were averaged across replicates (each 15cm plate subjected to the L1-*neo* retrotransposition assay) for each L1-REPEL screen. The average sgRNA read count for pJM101/L1.3 transfected replicates represented the "treatment" population (see Figure 2.2 B). For GeCKO screens 1 and 2, we tested the

average sgRNA read counts in the “treatment” population *vs*. the average sgRNA read counts in the unselected day 21 transduced “control” population. For the Brunello screen, we tested sgRNA read counts in the pJM101/L1.3 transfected “treatment” population *vs*. sgRNA read counts in pCDNA3 transfected “control” population.

MAGeCK ranks sgRNAs based on p-values calculated from a negative binomial distribution; a modified robust ranking aggregation (RRA) algorithm then is used to rank candidate genes (Li et al., 2014). We used MAGeCK to plot the distribution of gene RRA scores from the combined GeCKO screens (Figure 2.3 B). We then used MAGeCK to generate a ranked list of candidate genes for each L1-REPEL screen: GeCKO 1, GeCKO 2, and Brunello (Figure 2.3 C, top 10 genes). Based on the RRA distribution, we then searched for common genes that were ranked within the top 100 of GeCKO 1 and GeCKO 2. Neurofibromin 2 *(NF2)* and Exportin 7 *(XPO7)* were the only common candidate genes ranked within the top 100 of each GeCKO screen.

The MAGeCK algorithm is not optimized to analyze data from a selection-based screen with low numbers of sgRNAs. Thus, we implemented an alternative method to better distinguish true candidate genes from false negatives. Briefly, we reasoned that if the knockout of a gene reduced L1-REPEL, then the candidate gene sgRNA profile would contain: (1) individual sgRNAs in multiple biological replicates (Figure 2.3 D: gene *B*) and/or (2) biological replicates with multiple unique sgRNAs (Figure 2.3 D: gene *B*).

We applied a binary readout for each sgRNA (based on its presence or absence) that was independent of sgRNA read depth. To reduce background, we limited our analysis to the top ~500 sgRNAs identified within each pJM101/L1.3 transfected replicate; this cutoff value ~500 was chosen based on the number of visible G418-

resistant colonies in the L1-REPEL screens (see above). Using these criteria, we identified 228 individual sgRNAs that were present in two or more biological replicates within GeCKO screen 1. The only gene with multiple sgRNAs was *NF2*: sgRNA NF2_19 was present in 8/10 replicates and sgRNA NF2_60 was present in 6/10 replicates (Figure 2.3 E). In GeCKO screen 2, we identified 113 individual sgRNAs that were present in two or more biological replicates. The only genes with multiple sgRNAs were *NF2* and *XPO7* (Figure 2.3 E). *NF2* was targeted by 4 sgRNAs: NF2_19 (7/7 biological replicates); NF2_60 (3/7 biological replicates); NF2_17 (2/7 biological replicates); and NF2_18 (2/7 biological replicates) (Figure 2.3 E). *XPO7* was targeted by 3 sgRNAs: XPO7_06 (3/7 biological replicates); XPO7_07 (2/7 biological replicates); and XPO7_08 (2/7 biological replicates) (Figure 2.3 E). Of the 59 sgRNAs identified within the Brunello screen, none targeted the same gene. Thus, this alternative method of analysis similarly identified *NF2* and *XPO7* as our top candidate genes.

Finally, using a less stringent but similar strategy, we next aimed to identify candidates that had at least two unique sgRNAs targeting a gene across all biological replicates. We identified a list of 400 genes in GeCKO screen 1,181 genes in GeCKO screen 2, and 251 genes in the Brunello screen. Only 14 genes were common between GeCKO screen 1 and GeCKO screen 2 (Figure 2.3 E). Of the 14 genes found in both GeCKO screens, only *NF2* had multiple sgRNAs in the Brunello screen. Thus, like the previous analyses, these results clearly identify *NF2* as our top candidate L1-REPEL factor.

Validation of L1-REPEL candidate genes

We next sought to perform proof-of-principle experiments to test whether the top candidate genes identified in our screens are necessary for L1-REPEL in PA-1 cells. Briefly, we used an orthogonal plasmid-based approach to knockout single candidate genes in a wild-type population of PA-1 cells and then performed L1 retrotransposition assays to test whether the sgRNA-mediated knockouts affect L1 REPEL (Figure 2.4 A).

We used the PX459 expression plasmid, which allows the cloning of individual candidate sgRNAs into a vector that expresses the Cas9 protein and a puromycin resistance selectable marker (Ran et al., 2013a; Ran et al., 2013b). We designed and subcloned sgRNA sequences into PX459 targeting our top three L1 REPEL genes as well as a cohort of six “hand-picked” candidate genes that we hypothesize may affect L1 REPEL in PA-1 cells. These genes include: *NF2*, *XPO7*, *TADA2B*, *ZC3H12B*, *EZH1*, *HES5*, *TRIM46*, *WBP5*, and *TRIM49C* (Figure 2.4 B). Notably, the sgRNA sequence used in these experiments are identical to the recovered sgRNA sequences from the L1-REPEL GeCKO screens.

Wild-type PA-1 cells were plated and subsequently transfected with the candidate knockout vectors. As a control, we also transfected an empty PX459 vector into PA-1 cells in parallel experiments. Approximately 24 hours post-transfection, the cell growth media was supplemented with puromycin (2 μg/mL) to select for the PX459 derivative vectors. After 48 hours, when all of the untransfected cells had died, we replaced the puromycin-containing media with fresh PA-1 cell growth media. The resultant cells were passaged upon reaching high density. At 12 days post-transfection, the resultant population of PA-1 cells transfected with the relevant PX459 derivative vector was

subjected to the L1-*neo* assay (Figure 2.4). Herein we refer to this assay as the "population knockout assay." As a negative control, we transfected cells with pJM105/L1.3, which contains an inactivating missense mutation in the ORF2p reverse transcriptase domain. When G418-selection was complete (after ~14 days in G418-containing media), the resultant foci were fixed and stained for visualization (Figure 2.4 C). G418-resistant foci were evident in wells transfected with candidate knockout vectors targeting *NF2*, *XPO7*, *TADA2B* and *ZC3H12B* (Figure 2.5 C), but were not observed in wild-type PA-1 cells or cells transfected with an empty PX459 vector.

To further validate our top candidate gene, *NF2*, we repeated this assay using multiple sgRNAs (NF2_19 and NF2_60) (Figure 2.4 D). Importantly, cells transfected with pJM105/L1.3 did not yield G418 resistant foci, suggesting that *NF2* knockout does not simply confer G418-resistance in PA-1 cells by activating a multi-drug resistant response or another mechanism (Figure 2.4 D). Together, these results suggest that "population knockout" of *NF2*, *XPO7*, *TADA2B* and *ZC3H12B* allows a subset of cells to escape L1-REPEL. Additional experiments to further characterize how *NF2* affects L1 REPEL are presented in Chapter 3.

In sum, we implemented a genome-wide CRISPR/Cas9 gene knockout screen in PA-1 human embryonic carcinoma cells to identify cellular factors mediating L1-REPEL. We identified *NF2* and *XPO7* as top candidate genes using multiple method analyses. Preliminary validation experiments in PA-1 cells suggest that CRISPR-Cas9-based knockout of *NF2*, *XPO7*, *TADA2B* and *ZC3H12B* may alleviate L1-REPEL. Together, our results suggest that a genome-wide CRISPR/Cas9 gene knockout system can identify putative factors that mediate L1-REPEL in PA-1 cells.

Discussion

PA-1 cells exhibit L1-REPEL

Previous studies demonstrated that human L1s are expressed in PA-1 human embryonic carcinoma cells lines (Garcia-Perez et al., 2007b). However, despite the high levels of expression, engineered L1s are efficiently and stably silenced upon retrotransposition in hEC cells, thereby uncovering a mechanism that may restrict the expression of L1 retrotransposition events during early human development (Garcia-Perez et al., 2010).

We verified previous reports that L1-delivered reporter genes are efficiently silenced in PA-1 cells by a process that we have termed L1-REPEL (Figure 2.1). We demonstrated that PA-1 cells efficiently silence three different L1-delivered reporter genes, including a previously untested blasticidin resistance gene (Figure 2.1 A). Notably, each of the reporter genes were efficiently expressed in somatic HeLa cells upon L1 retrotransposition, agreeing with previous reports that L1-REPEL is peculiar to hEC cells (Garcia-Perez et al., 2010).

We also replicated previously reported data and demonstrated that TSA treated pk5 cells efficiently reactivated a silenced L1-*GFP* retrotransposition event (Figure 2.1 B) and that the subsequent removal of TSA resulted in re-establishment of L1-REPEL (Figure 2.1 C) (Garcia-Perez et al., 2010). These data are consistent with our working model that L1-REPEL is an epigenetic mechanism that leads to stable, and reversible, silencing of the L1-delivered *EGFP* reporter gene either during or immediately after TPRT-mediated integration into the genome of PA-1 cells. Notably, the current L1-REPEL model posits that the silencing of L1-delivered reporter genes involves both

initiation and maintenance steps to establish and maintain reporter gene silencing (Figure 2.1 D).

A genome-wide CRISPR/Cas9 screen to identify L1-REPEL factors

To gain insight into the L1-REPEL mechanism, we sought to identify cellular factors that mediate L1-REPEL. We designed and implemented a forward genetic screen, utilizing a lentiviral-delivered CRISPR/Cas9-based system (GeCKO), to knockout genes on a genome-wide scale (Shalem et al., 2014). The GeCKO system relies on lentiviral-delivered expression of Cas9, a single-guide RNA (sgRNA), and a puromycin selectable marker (Figure 2.2 A) (Sanjana et al., 2014; Shalem et al., 2014). The sgRNA-guided Cas9 nuclease then is targeted to genic coding regions, where it induces a double-strand DNA break, which upon repair, can result in the generation of a loss-of-function allele(s). We then performed L1 retrotransposition assays, where puromycin-resistant PA-1 cells are transfected with an engineered human L1 vector and subjected to drug selection to identify cells that express the L1-delivered reporter gene (Figure 2.2 B). We hypothesized that knockout of genes involved in the initiation and/or maintenance of L1-REPEL would result in L1-delivered reporter gene expression.

The GeCKO system efficiently edits genes within 21 days

To implement the GeCKO system in PA-1 cells, we utilized two different sgRNA libraries, GeCKOv2 and Brunello, to target Cas9 to cellular genes. We demonstrated that the Cas9 protein is continuously expressed after implementing the GeCKO system (Figure 2.2 C). Thus, in agreement with previous reports, our results confirm that lentiviral-delivered sequences are not efficiently silenced in PA-1 cells (Garcia-Perez et

al., 2010), consistent with the model that L1 REPEL may recognize structural features generated during TPRT that are particular to non-LTR retrotransposition.

To determine the kinetics of gene knockout in PA-1 cells, we used MAGeCK (Li et al., 2014) to identify sgRNA within essential cellular pathways that were depleted in transduced PA-1 cells as a function of time. We demonstrated that sgRNAs targeting genes within the KEGG-ribosome, -proteasome, -spliceosome, and -RNA polymerase pathways were significantly depleted by 21 days post-transduction (Figure 2.2 D). Similar sgRNA depletion was observed 31 days post-transduction. These results demonstrate that efficient gene knockout occurs by day 21 post-transduction in PA-1 cells; this data is consistent with previous reports that demonstrated GeCKOv2-mediated drop out of “gold standard” essential genes by ~20 cell doublings (Hart et al., 2014; Hart et al., 2015).

L1-REPEL screen in PA-1 cells

We performed the L1 retrotransposition assay in day-21 transduced PA-1 cells to identify genes involved in either the initiation and/or maintenance of L1-REPEL. We performed three independent L1-REPEL screens: GeCKO screens 1 and 2 utilized the GeCKOv2 sgRNA library delivered by lentiCRISPRv2: GeCKO screen 1 consisted of 10 biological replicates, whereas GeCKO screen 2 consisted of 7 biological replicates. A third L1-REPEL screen, using the Brunello sgRNA library, delivered by a two-vector system (lentiCas9-Blast and lentiGuide-Puro), consisted of 5 biological replicates.

Genome-scale CRISPR/Cas9 gene knockout resulted in a ~10-fold increase in G418-resistant colonies per 15 cm plate transfected with pJM101/L1.3 when compared to wild-type PA-1 cells or Cas9-transduced PA-1 cells lacking the sgRNA library (Figure

2.3A). This increase in G418-resistant cells, presumably due to gene knockout, then allowed an unbiased way to identify sgRNAs and, in turn, genes that may be necessary for L1 REPEL.

We used various strategies to identify candidate sgRNAs that knockout genes required for L1 REPEL. First, we used MAGeCK to plot the distribution of candidate gene RRA scores for the GeCKO screens (Figure 2.3 B) and to generate ranked lists of candidate genes based on sgRNA enrichment within each L1-REPEL screen (Figure 2.3 C: top 10 genes shown). We identified two genes, *NF2* and *XPO7*, that were ranked in the top 100 of GeCKO screen 1 and GeCKO screen 2. Although MAGeCK calculates p-values and false discovery rates (FDR) for each gene, it was difficult to set a threshold for significance. For example, 232 genes within GeCKO screen 1 had a p-value <0.01; however, only 4 genes had an FDR <30%, and 0 genes had an FDR <10%. It is noteworthy that MAGeCK was not optimized to analyze data from a selection-based screen with low numbers of sgRNAs.

The analysis of sgRNA profiles revealed that each biological replicate contained an unexpectedly high number (~3,000-5,000) of unique sgRNA sequences when considering the ~300-500 G418-resistant colonies we obtained per biological replicate. Clearly, there are more sgRNAs in our population than visible colonies, but what could account for such a result? Intriguingly, in our initial screens, we consistently observed several small colonies (<10 cells) and many single cells throughout the plates upon microscopic analysis. Thus, we posit these cells likely contributed to the high number of unique sgRNAs per biological replicate.

To further prioritize our list of candidate genes, we implemented an alternative strategy to identify L1-REPEL candidate genes that utilized the power of biological replicates. We reasoned that the knockout of a gene involved in L1 REPEL should consistently be present in more than one biological replicate with perhaps a greater representation of different sgRNAs targeting that gene (Figure 2.3 D, see gene *B*). To further increase stringency, we limited our analysis to the top ~500 sgRNAs within each GeCKO biological replicate. A cutoff of ~500 was chosen based on the maximum number of visible G418-resistant colonies, reasoning that sgRNAs derived from G418-resistant foci would comprise the majority of total sgRNA reads, as opposed to microscopic foci and single cells, where sgRNAs would be predicted to have low read counts. Indeed, the top ~500 sgRNAs represented the majority (~50-70%) of the total reads per replicate.

We next conducted a Brunello screen based on what we learned from the GeCKO screens. We hypothesized that some single G418-resistant cells arising in the GeCKO screens might be post-mitotic. Thus, after the completion of G418 selection, we trypsinized and replated cells in the presence of G418. The top 500 sgRNAs within the Brunello screen represented >95% of the total reads, suggesting that replating G418-resistant cells may help reduce background of low read count sgRNAs.

We consistently identified *NF2* and *XPO7* as the only genes with two or more sgRNAs in multiple biological replicates. We also identified genes that had at least two different sgRNAs targeting a gene across all biological replicates. Only 14 genes were found in both GeCKO screen 1 and GeCKO screen 2 (Figure 2.3 E). Of the 14 genes,

only *NF2* had multiple sgRNAs identified within the Brunello screen (Figure 2.3 E). These results clearly identified NF2 and XPO7 as our top candidate L1-REPEL factors.

Validation of L1-REPEL candidate genes

To establish a candidate gene list, we started with the top 50 genes ranked by MAGeCK in each L1-REPEL screen, then looked for common genes with sgRNA profiles exhibiting either: (1) 2 or more sgRNAs within 2 or more screens or (2) a single sgRNA within 3 or more biological replicates within a single screen. These analyses resulted in a list of 20 highly enriched candidate genes.

To identify additional genes with moderate enrichment, we started with the top 100 genes ranked by MAGeCK in each L1-REPEL screen, then looked for common genes with sgRNA profiles exhibiting either: (1) 2 or more sgRNAs in a single screen or (2) a single sgRNA within 3 or more biological replicates across all L1-REPEL screens. These analyses resulted in the addition of 384 moderately enriched genes to our candidate gene list.

Finally, we then screened for candidate genes targeted by 2 or more sgRNAs with annotated functions that may be relevant for the epigenetic regulation or other cellular processes related to TPRT and L1 biology. An additional 85 enriched candidate genes were added to our final candidate gene list based on biological relevance. Together, our final candidate gene list includes 489 genes: 20 are highly enriched, 384 are moderately enriched, and 85 “hand-picked” genes (Table 2.1).

To attempt to validate a subset of our top candidate genes, we employed a transient “population knockout assay” using the PX459 vector in conjunction with the L1 retrotransposition assay (Figure 2.4 A) (Ran et al., 2013a; Ran et al., 2013b). Cas9

nuclease specificity is determined by the 20-nucleotide guide sequence within the sgRNA. For *S. pyogenes* Cas9, the target sequence must precede a 5′-NGG PAM sequence. We used CRISPick software (https://portals.broadinstitute.org/gppx/crispick/public) to identify sgRNAs targeted to our candidate genes and then designed oligonucleotides containing the sgRNA sequence flanked by *Bbs*I restriction sites (Figure 2.4 B) targeting the following genes: *NF2, XPO7, TADA2B, HES5, TRIM46, TRIM49C, ZC3H12B, EZH1,* and *WBP5*.

The resultant PX459-based knockout vectors then were transfected into PA-1 cells to generate "population knockout" cells that were subjected to the L1 retrotransposition assay. We observed G418-resistant foci in wells transfected with knockout vectors targeting *NF2*, *XPO7*, *TADA2B* and *ZC3H12B* (Figure 2.4 C). To further validate our top candidate gene, *NF2*, we replicated this experiment using two *NF2*-targeting sgRNAs (NF2_19 and NF2_60) (Figure 2.4 D). Importantly, cells transfected with the retrotransposition-defective pJM105/L1.3 mutant L1 expression construct were susceptible to G418 selection, suggesting that knockout of NF2 does not simply confer G418-resistance in PA-1 cells (Figure 2.4 C). These data suggest that "population knockout" of *NF2*, *XPO7*, *TADA2B* and *ZC3H12B* allows a subset of cells to escape L1-REPEL. A brief description of the documented functions of each of the genes/proteins is provided below:

> Neurofibromin 2 (*NF2*): is a tumor suppressor gene that encodes the membrane-cytoskeletal scaffold protein NF2/merlin (moesin-ezrin-radixin-like protein). Loss of function mutations in human *NF2* cause neurofibromatosis type 2 (NF2) (Cooper and Giancotti, 2014; Rouleau et al., 1993; Trofatter et al., 1993). NF2/merlin is an upstream regulator of the Hippo signaling pathway and has been implicated in a variety of cellular processes, including embryonic

development and the cellular stress response (Laulajainen et al., 2008; Perrimon et al., 2012; Stamenkovic and Yu, 2010). Furthermore, NF2/merlin can localize to the nucleus where it binds DCAF1 (VPRBP) and modulates $CRL4^{DCAF1}$ E3 ubiquitin ligase activity (Li et al., 2010). Possible roles for NF2/merlin in L1-REPEL are further discussed below.

Exportin 7 (*XPO7*): encodes a protein that mediates the bidirectional nuclear transport of cargo molecules in a Ran GTP-dependent manner (Aksu et al., 2018). XPO7 is thought to export histones H2A, H3, and H4 and is necessary for nuclear condensation during erythroid differentiation (Cantu et al., 2019; Hattangadi et al., 2014). Thus, XPO7 may facilitate import of histone variants associated with transcriptional repression. Intriguingly, one of our highly enriched candidate genes was *H2AFY*, a H2A variant associated with transcriptional repression and heterochromatin formation (Costanzi and Pehrson, 1998; Douet et al., 2017). Moreover, H2AFY was identified as an L1 ORF2p-interacting protein by IP-coupled mass spectrometry (Miyoshi et al., 2019). Thus, loss of XPO7 may prevent H2AFY from entering the nucleus, preventing heterochromatin formation at the site of L1 integration. Recently, XPO7 was classified as a tumor suppressor protein that regulates cellular senescence (Innes et al., 2021). Senescence is a type of cellular stress response that alters transcription, metabolism, and chromatin organization, ultimately causing cell cycle arrest (Gorgoulis et al., 2019). Activation of the DNA damage response and the production of reactive oxygen species also can induce cellular senescence (Correia-Melo et al., 2016; Takahashi et al., 2006; Victorelli and Passos, 2017). Thus, the loss of XPO7 may hinder the DNA damage response during target-primed reverse transcription (TPRT), preventing DDR-mediated cellular senescence.

Transcriptional adapter 2B (*TADA2B*): encodes a protein that is part of the chromatin-modifying SAGA (Spt-Ada-Gcn5 acetyltransferase) complex. Recent work suggests that TADA2B and the SAGA complex regulate pluripotency, growth, and differentiation (Naxerova et al., 2021). Interestingly, the SAGA

complex contains both histone acetyltransferase (HAT) and histone deubiquitinase (DUB) activities. Histone ubiquitination plays critical roles in transcription, DNA replication, and DNA repair (Cheon et al., 2020; Fleming et al., 2008; Shilatifard, 2006). Thus, the loss of TADA2B may attenuate L1-REPEL by disrupting SAGA complex activity.

Zinc Finger CCCH-Type Containing 12B (*ZC3H12B*): encodes a cytoplasmic RNA-binding protein. ZC3H12B binds motifs similar to the splice donor sequence, suggesting it may function in the recognition and degradation of unspliced cytoplasmic cellular or viral RNAs (Jolma et al., 2020). Intriguingly, the ZC3H12 family of proteins have been linked to viral immunity (Fu and Blackshear, 2017). Furthermore, the ZC3H protein ZAP has been implicated in restricting L1 retrotransposition (Goodier et al., 2015; Moldovan and Moran, 2015). Thus, ZC3H12B may influence post-transcriptional regulation of L1 RNA in the cytoplasm.

Additional experiments are required to determine CRISPR-Cas9 editing efficiency in "population knockout" cells. It also will be important to confirm whether genomic edits for each candidate gene generate loss-of-function alleles and affect protein expression and/or establish "clonal" knockout cell lines to rigorously evaluate each candidate gene in L1-REPEL. We further assess the role of our top candidate gene, *NF2*, in Chapter 3.

Interestingly, both *NF2* and *TADA2B* were top hits within the original published GeCKO screen (Shalem et al., 2014). Here, the authors sought to identify genes whose loss resulted in resistance to Vemurafenib (PLX). PLX is a BRAF inhibitor used to treat melanoma (Bollag et al., 2010). Notably, NRAS activation confers resistance to PLX (Nazarian et al., 2010; Romano et al., 2013) and PA-1 cells typically contain activated NRAS (Tainsky et al., 1984), providing a possible link as to why we also identified *NF2* in our screen.

TADA2B is a chromatin-modifying protein involved in transcription and is a member of the STAGA (SPT3-TAF9-GCN5-acetylase) complex (Barlev et al., 2003), which recruits Mediator complex proteins to the c-MYC oncogene to activate proliferation (Liu et al., 2008). An increase in cell proliferation could explain why *NF2* and *TADA2B* were found in the PLX screen. However, in our L1-REPEL screen, if knockout of a particular gene results in G418-resistance independent of L1-delivered reporter gene expression, we would predict to see G418-resistant foci formation in plates transfected with the retrotransposition-defective pJM105/L1.3 mutant L1 expression construct. Because we only see G418-resistant foci in plates transfected with the WT pJM101/L1.3 L1 expression construct, it is unlikely that knockouts of *NF2* or *TADA2B* confer a proliferative advantage that allows cells to survive G418 selection.

Conclusion

In sum, we designed and executed a genome-wide CRISPR/Cas9 gene knockout screen in PA-1 human embryonic carcinoma cells to identify putative cellular factors mediating L1-REPEL. We identified 489 candidate genes for further investigation, which includes our top candidate genes *NF2* and *XPO7*. Preliminary validation experiments in PA-1 cells suggest that CRIPR-Cas9-based knockout of *NF2*, *XPO7*, *TADA2B* and *ZC3H12B* may allow a subset of cells to escape L1-REPEL. Together, our results suggest that a genome-wide CRISPR/Cas9 gene knockout system can identify putative factors necessary for efficient L1-REPEL in PA-1 cells. Future studies are required to assess the role of the identified candidate genes in L1-REPEL.

Materials and Methods

Cell lines and cell culture conditions

PA-1 cells were purchased from American Type Culture Collection and cultured as previously described (Garcia-Perez et al., 2010). Briefly, cells were cultured in Minimum Essential Media (MEM) supplemented with 10% heat-inactivated fetal bovine serum, 2 mM Glutamax, 100 U/ml penicillin, 0.1 mg/ml streptomycin and 0.1 mM non-essential amino acids. We found that heat inactivating the fetal bovine serum was critical for conducting assays in PA-1 cells. HeLa-JVM cells (Moran et al., 1996) were cultured in Dulbecco's Modified Eagle Medium (DMEM) high glucose supplemented with 10% fetal bovine serum (FBS), 2 mM Glutamax, 100 U/ml penicillin and 0.1mg/ml streptomycin. All cell lines were grown at 37ºC in a humidified 7% CO2 incubator.

Expression vectors

All vectors were propagated in *Escherichia coli* strain DH5a (F-f80lacZDM15D[lacZYA-argF] U169 recA1 endA1 hsdR17 [rk-, mk+] phoA supE44 l- thi-1 gyrA96 relA1) (Invitrogen). Competent *E. coli* were prepared and transformed using previously described methods (Inoue et al., 1990; Moran et al., 1996). Plasmids were prepared using the Qiagen Plasmid Midi Kit according to the manufacturer's instructions.

pJM101/L1.3: contains a full-length L1 (L1.3, GenBank: L19088) that includes the *neomycin* retrotransposition indicator cassette within its 3'UTR (Sassaman et al., 1997; Wei et al., 2001). A CMV promoter and SV40 polyadenylation signal in the pCEP4 plasmid backbone facilitate L1.3 expression. This vector was used in our L1 *neo*-based retrotransposition assays.

pJM105/L1.3: is identical to pJM101/L1.3 except for the presence of a missense mutation (D702A) in the L1.3 ORF2p reverse transcriptase (RT) domain, which renders L1.3 retrotransposition-defective (Wei et al., 2001). This vector was used as a negative control in L1 *neo*-based retrotransposition assays.

pCDNA3: expresses a neomycin resistance gene from the vector backbone (Invitrogen). This vector was used as a positive control for transfection and G418 drug selection in the L1 *neo*-based retrotransposition assays.

pJJ101/L1.3: is similar to pJM101/L1.3, but contains an *mblastI* retrotransposition indicator cassette within its 3'UTR (Goodier et al., 2007; Kopera et al., 2011; Morrish et al., 2002). A CMV promoter and SV40 polyadenylation signal in the pCEP4 plasmid backbone facilitate L1.3 expression. This vector was used in our L1-*blast* retrotransposition assays.

pJJ105/L1.3: is identical to pJJ101/L1.3 except for the presence of a missense mutation (D702A) in the L1.3 ORF2p reverse transcriptase (RT) domain, which renders L1.3 retrotransposition-defective (Goodier et al., 2007; Kopera et al., 2011; Morrish et al., 2002). This vector was used as a negative control in our L1-*blast* retrotransposition assays.

pCDNA6: expresses a blasticidin resistance gene from the vector backbone (Invitrogen). This vector was used as a positive control for transfection and blasticidin drug selection in the L1-*blast* retrotransposition assays.

p99EGFP/LRE3: contains a full-length RC-L1 (LRE3) with an *mEGFPI* retrotransposition indicator cassette within its 3'UTR. LRE3 expression is driven from its

native 5'UTR (Ostertag et al., 2000). The LRE3 expression construct was cloned into a version of pCEP4 that lacks the CMV promoter. A puromycin-resistance selectable marker replaced the hygromycin-resistance selectable marker in pCEP4 (Garcia-Perez et al., 2010). This vector was used in our L1-*GFP* retrotransposition assays.

p99EGFP/LRE3-111: is identical to p99EGFP/LRE3 except that it contains two missense mutations in LRE3 ORF1p (RR261-262AA), which renders LRE3 retrotransposition-defective (Zhang et al., 2014). This vector was used as a negative control in our L1-*GFP* retrotransposition assays.

pCEP4/GFP: contains the coding sequence of the humanized Renilla reniformis green fluorescent protein (hrGFP) from phrGFP-C (Stratagene). GFP expression is driven by a cytomegalovirus (CMV) immediate early promoter and terminated at a simian virus 40 (SV40) late polyadenylation signal present in the pCEP4 plasmid backbone (Alisch et al., 2006). This vector was used to calculate transfection efficiencies.

sgRNA libraries and lentiviral vectors

GeCKOv2 library: The human GeCKOv2 sgRNA library has been previously described (Sanjana et al., 2014). Briefly, the GeCKOv2 library contains 123,411 single-guide sequences (sgRNAs) divided equally between two libraries (A and B). The GeCKOv2 library targets 19,050 genes (six sgRNAs per gene), 1,864 miRNAs (four sgRNAs per miRNA), and has 1000 control non-targeting sgRNAs. The GeCKOv2 library was purchased from Addgene.

Brunello library: The human Brunello sgRNA library has been previously described (Doench et al., 2016). Briefly, the improved Brunello library contains 76,441 single-guide

sequences (sgRNAs). The Brunello library targets 19,114 genes (four sgRNAs per gene) and has 1000 non-targeting control sgRNAs.

LentiCRISPRv2: LentiCRISPRv2 is a single vector lentiviral system containing both Cas9 and candidate sgRNA (Sanjana et al., 2014; Shalem et al., 2014). The plasmid contains a psi+ packaging signal, a rev response element (RRE), central polypurine tract (cPPT), which are necessary for packaging RNA into lentiviral virus-like particles, and a human U6 snRNA promoter driving expression of the sgRNA sequence. The elongation factor 1α (EFS) short promoter drives expression of *Streptococcus pyogenes* Cas9-FLAG as well as a puromycin resistance marker. A P2A self-cleaving peptide separates Cas9-FLAG from the puromycin resistance selectable marker. A woodchuck hepatitis virus posttranscriptional regulatory element (WPRE) is present at the 3' end of the EFS transcriptional unit to enhance expression.

LentiCas9-Blast: LentiCas9-Blast is part of a two-vector lentiviral system. LentiCas9-Blast delivers the Cas9 coding sequence (Sanjana et al., 2014; Shalem et al., 2014). The plasmid contains a psi+ packaging signal, a rev response element (RRE), central polypurine tract (cPPT), which are necessary for packaging RNA into lentiviral virus-like particles. The elongation factor 1α (EFS) short promoter drives expression of *Streptococcus pyogenes* Cas9-FLAG as well as a blasticidin resistance marker. A P2A self-cleaving peptide separates Cas9-FLAG from the blasticidin resistance selectable marker. A woodchuck hepatitis virus posttranscriptional regulatory element (WPRE) is at the 3' end of the EFS transcriptional unit to enhance expression.

LentiGuide-Puro: LentiGuide-Puro is part of a two-vector lentiviral system. LentiGuide-Puro delivers the sgRNA (Sanjana et al., 2014; Shalem et al., 2014). The plasmid

contains a psi+ packaging signal, a rev response element (RRE), a central polypurine tract (cPPT), which are necessary for packaging RNA into lentiviral virus-like particles, and a human U6 snRNA promoter driving expression of the sgRNA. The elongation factor 1α promoter drives expression of a puromycin resistance marker. A woodchuck hepatitis virus posttranscriptional regulatory element (WPRE) is at the 3' end of the transcript to enhance expression.

Cloning of sgRNA libraries into lentiviral vectors: the plasmid libraries (GeCKOv2 and Brunello) were digested with *Bsm*BI to excise the sgRNA sequences. The sgRNA sequences were PCR amplified, digested with *Bsm*BI, and ligated into *Bsm*BI digested lentiviral vectors (lentiCRISPRv2 and lentiGuide-Puro) (Doench et al., 2016; Sanjana et al., 2014; Shalem et al., 2014).

Lentiviral production and packaging

Lentivirus packaging was performed at the University of Michigan Vector Core, which is directed by Dr. Thomas Lannigan. For each preparation, 650 µg of the purified proviral vector was packaged into lentiviral particles using the psPAX2 vector. Human embryonic kidney A293T cells were used for lentiviral production. Supernatants containing mature viral particles were collected and frozen at -80ºC.

GeCKOv2 lentiviral infection of PA-1 and pk5 cells

A single viral vector (lentiCRISPRv2) delivered Cas9 and the GeCKOv2 sgRNA library. Wild-type PA-1 or pk5 cells were plated in 15 cm dishes (BD Biosciences) at a density of $6X10^6$ cells/plate. Twenty-four hours post-plating 15 cm dishes were transduced with GeCKOv2-lentiCRISPRv2 viral supernatant at a multiplicity of infection (MOI) of 0.3 with

8 µg/ml Polybrene (Sigma). Transduced cells from each sub-library, A and B, were maintained separately. Twenty-four hours post-transduction, the media was replaced with fresh PA-1 media. Forty-eight hours post-transduction, puromycin (2 ug/ml) was added to the media. Transduced PA-1 cells were collected and reseeded 7-, 14-, 21-, 24-, 28-, and 31-days post-transduction. Transduced pk5 cells were collected and reseeded 7-,14- and 21-days post-transduction. At each time point we collected ~$1X10^9$ cells; $1X10^8$ cells from each sub-library were reseeded into 15 cm dishes, $1X10^8$ cells were cryopreserved as cell stocks and $1X10^8$ cells were frozen at -30ºC for gDNA collection.

Brunello lentiviral infection of PA-1 cells

Separate viral vectors delivered Cas9 (lentiCas9-Blast) and the Brunello sgRNA library (lentiGuide-Puro). Wild-type PA-1 cells were plated in 15 cm dishes (BD Biosciences) at a density of $6X10^6$ cells/plate. Twenty-four hours after plating, 15 cm dishes were transduced with media containing lentiCas9-Blast viral supernatant at an MOI of 0.3 with 8 µg/ml Polybrene (Sigma). Twenty-four hours post-transduction, media was replaced with fresh PA-1 media. Forty-eight hours post transduction, blasticidin (10 µg/ml) was supplemented to the media. 5 days post-transduction blast-media was replaced with fresh PA-1 media. 7 days post-transduction, cells were collected and re-plated in 15 cm dishes (BD Biosciences) at a density of $6X10^6$ cells/plate. Twenty-four hours later, 15 cm dishes were transduced with viral supernatant containing lentiGuide-Puro with the Brunello sgRNA library. Cells were transduced at an MOI of 0.3 with 8 µg/ml Polybrene (Sigma). Twenty-four hours post-transduction, media was replaced with fresh PA-1 media. Forty-eight hours post-transduction, puromycin (2 µg/ml) was

added to the media. Cells were collected 7- and 14-days after lentiGuide-Puro transduction. At each time point we collected ~1X10^9 cells; 1X10^8 cells were reseeded into 15 cm dishes and 1X10^8 cells were cryopreserved as cell stocks.

Genomic DNA collection

For drug selection-based experiments, genomic DNA was extracted using the QIAGEN DNeasy Blood and Tissue Kit (QIAGEN, 12145) per the manufacturer recommendations. For the pk5 L1-REPEL maintenance screen, genomic DNA was extracted using the Wizard Genomic DNA Purification Kit (Promega, A1120).

PCR amplification of genomic sgRNA sequences

Genomic DNA was subjected to two rounds of PCR amplification as previously described (Sanjana et al., 2014; Shalem et al., 2014) using the KAPA HiFi PCR kit (KapaBiosystems). To ensure complete coverage of sgRNA sequences, we performed eight PCR reactions using 2.5 µg gDNA per timepoint (7-, 21- and 31-days). A single PCR reaction using 0.5 µg gDNA was performed for each L1-*neo* biological replicate. An initial denaturation time of 5 minutes at 95 °C was followed by 27 cycles of amplification (20 second annealing at 60°C, 30 second elongation at 72°C). PCR_1 added sequences necessary for the MiSEQ platform (primers available upon request). For PCR_ 2, we used 5 ul of PCR_1 product, an initial denaturation of 2 minutes, and only 7 amplification cycles. P5 and P7 adapters (Illumina) were used to add 8 bp barcodes onto the PCR_1 product allowing multiplexing and sequencing of numerous samples on a single MiSEQ run. PCR products were size selected for products greater than 300 bp using the SPRI-cleanup (Beckman Coulter Agencourt AMPure XP purification system) following the manufacturers recommendations.

MiSEQ sequencing

Sequencing of sgRNA sequences was performed on the MiSEQ (Illumina) platform using MiSEQ Reagent Kits V3 with a 2 X 75 output, following the manufacturers recommendations. The final concentration of the multiplexed PCR products was 12 pmol per MiSeq run.

MAGeCK data analysis

Raw FASTQ files were sorted, trimmed, and mapped to an indexed reference of all sgRNA sequences. The MAGeCK -count command was used to generate normalized sgRNA read counts from FASTQ files. The MAGeCK -test command was used to rank sgRNAs and genes from read count tables. The MAGeCK -plot command was used to generate RRA score distribution graphs. MAGeCK version 0.5.8 was downloaded from (https://sourceforge.net/p/mageck/wiki/Home/). Analysis was performed as previously described (Li et al., 2014).

Western blotting

The following protocol contains minor changes from the original protocol developed by Dr. John Moldovan and Dr. Peter Larson (Moldovan and Moran, 2015). Cells were washed with PBS, trypsinized (0.25% Trypsin-EDTA) and pelleted at 500×g for 2 minutes. Whole cell lysates were prepared by incubating cell pellets in ~500 µL (1 mL lysis buffer per 100 mg of cell pellet) of lysis buffer (20 mM Tris-HCl (pH 7.5), 150 mM NaCl, 10% glycerol, 1 mM EDTA, 0.1% NP-40 (Sigma), 1X complete EDTA-free protease inhibitor cocktail (Roche) on ice for 30 minutes. Lysates were cleared at 15,000×g for 10 minutes at 4°C and supernatants were transferred to a clean tube.

Protein concentration was determined using the Bradford reagent assay (BioRad). For SDS-PAGE, samples were diluted in 4x Laemmli buffer (BioRad) containing 10% BME and incubated at 95°C for 5 minutes. 30µg of protein was loaded per well of a 4–15% precast mini-PROTEAN TGX gel (BioRad) run at 200V for 35 minutes in 1X Tris/Glycine/SDS (25 mM Tris-HCL, 192 mM glycine, 0.1% SDS, pH 8.3) buffer (Bio-Rad Laboratories). Protein was transferred using the Trans-Blot Turbo Mini PVDF Transfer Packs (BioRad) with the Trans-Blot Turbo Transfer System (BioRad Laboratories) at 2.5A and 25V for 3 minutes. The membranes then were incubated at room temperature in Intercept blocking buffer (LI-COR) for 1 hour. Following blocking, fresh blocking buffer containing 0.1% Tween 20 (Sigma) was added with the following primary antibodies: FLAGM2 (Agilent, 200472) at 1:2000; b-actin (ThermoFisher, MA1-744) at 1:2000. The membrane was incubated overnight at 4°C in a sealed container. The next day, the membrane was washed 5X with 1X PBS and fresh blocking solution containing 0.1% Tween 20 (Sigma) and 0.02% SDS was added with the following secondary antibodies: Anti-Mouse IRDye 680LT (LICOR, 925-68022) at 1:15,000, and anti-Rabbit IRDye 800CW (LI-COR, 925-32213) at 1:15,000. The membrane was incubated for 1 hour at room temperature. Then the membrane was washed 5X with 1X PBS and scanned using the Odyssey CLx scanner (LI-COR).

PA-1 GeCKOv2 L1-REPEL screen

Day 21 lentiCRISPRv2-GeCKOv2 transduced PA-1 cells (21G) were thawed and cultured in 15 cm dishes (BD Biosciences). When cells reached 90% confluency, cells were collected and re-plated in 15 cm dishes at $8x10^6$ cells/dish in 25 mL of PA-1 media. Eighteen hours later, each 15 cm dish was transfected with a FuGENE 6 mix

containing 16.8 µg JM101/L1.3 plasmid DNA and 50.4 µl FuGENE 6 (Promega, E2692) up to a final volume of 1.68 mL Opti-MEM (Life Technologies). Twenty-four hours post-transfection, the media was replaced with fresh PA-1 media. Three days post-transfection, media was supplemented with 200 µg/mL of geneticin (G418) (Gibco, 10131-035) to select for L1 retrotransposition events expressing the L1-delivered reporter gene. After 14 days of drug selection, G418-resistant foci were washed with ice cold 1X Phosphate-Buffered Saline (PBS), trypsinized (0.25% Trypsin-EDTA, Gibco, 25200-056) and pelleted at 500×g for gDNA extraction. Each plate was collected independently as a biological replicate.

PA-1 Brunello L1-REPEL screen

Day 21 lentiGuide-Brunello transduced PA-1 cells (21B) were thawed and cultured in 15 cm dishes (BD Biosciences). When cells reached 90% confluency, cells were collected and re-plated in 15 cm dishes at $8x10^6$ cells/dish in 25 mL of PA-1 media. Eighteen hours later, each 15 cm dish was transfected with a FuGENE HD mix containing 20 µg JM101/L1.3 or pCDNA3 plasmid DNA and 50 µl FuGENE HD (Promega, E2312) up to a final volume of 1.68 mL Opti-MEM (Life Technologies). Twenty-four hours post-transfection, the media was replaced with fresh PA-1 media. Three days post-transfection, media was supplemented with 200 µg/mL of geneticin (G418) (Gibco, 10131-035) to select for L1 retrotransposition events expressing the reporter gene. After 14 days of drug selection, G418-resistant cells were collected and re-plated in 25 mL G418-supplemented PA-1 media. After 18 days of drug selection, G418-resistant cells were washed with ice cold 1X Phosphate-Buffered Saline (PBS), trypsinized (0.25%

Trypsin-EDTA, Gibco, 25200-056) and pelleted at 500×g for gDNA extraction. Each plate was collected independently as a biological replicate.

Candidate gene sgRNA oligonucleotides

sgRNAs were identified using CRISPick software (Broad Institute) with the following parameters: Human GRCh38 reference genome, CRISPRko mechanism, SpyoCas9 enzyme. (https://portals.broadinstitute.org/gppx/crispick/public). sgRNA-containing oligonucleotides were designed so they could be easily cloned into BbsI-digested PX459 plasmid (Ran et al., 2013a; Ran et al., 2013b). Sense (forward) and antisense (reverse) oligos were ordered from Integrated DNA Technologies (IDT) (Figure 2.5 B).

Cloning of candidate gene sgRNA vectors

Candidate gene sgRNAs were cloned into the PX459 vector for plasmid-based expression of Cas9 and the sgRNA. Cloning of oligos into the PX459 vector was performed as previously described (Ran et al., 2013a; Ran et al., 2013b). Sense and antisense oligos were phosphorylated using T4 PNK (NEB) and annealed. PX459 was digested by BbsI (NEB), then the phosphorylated-and-annealed oligos were ligated into the BbsI-digested PX459 vector. Digestion-ligation reactions were treated with PlasmidSafe exonuclease (Lucigen) and transformed into competent bacteria. Bacterial transformations were plated on LB-ampicillin plates and individual clones were picked and cultured overnight in LB-ampicillin broth at 37°C. Plasmid DNA was extracted, and sanger sequenced to confirm the sgRNA insert.

Candidate gene "population knockout" in PA-1 cells

PA-1 cells were plated at $3x10^5$ cells/well of a six-well dish. Twenty-four hours later, each well was transfected with a FuGENE 6 mix containing 1 µg candidate knockout vector DNA and 3 µl FuGENE 6 (Promega, E2692) up to a final volume of 105 µL Opti-MEM (Life Technologies). One day later, the transfection-media was replaced with PA-1 media containing 2 µg/mL puromycin to select for vector expression. Three days post-selection, puro-media was replaced with fresh PA-1 media. Once cells reached 90% confluency, ~6 days post-transfection, cells were passaged until reaching day 12 post-transfection. "Population knockout" cells were collected for cryopreservation and re-plated for the L1 retrotransposition assay.

L1 retrotransposition assay in "population" knockout PA-1 cells

The L1 retrotransposition assay was conducted as previously described (Garcia-Perez et al., 2010). Briefly, "population knockout" cells, were plated at $3x10^5$ cells/well of a six-well dish. Eighteen hours after plating, each well was transfected with a FuGENE 6 mix containing 1 µg pJM101/L1.3 or pJM105/L1.3 plasmid DNA and 3 µl FuGENE 6 (Promega, E2692) up to a final volume of 105 µL Opti-MEM (Life Technologies). One day post-transfection, the media was replaced with fresh PA-1 media. Three days post-transfection, media was supplemented with 200 µg/mL of geneticin (G418) (Gibco, 10131-035) to select for L1 retrotransposition events expressing the neomycin resistance gene. After ~14 days of drug selection, G418-resistant foci were washed with PBS, then fixed for 30 minutes at room temperature in a 1X PBS solution containing 2% paraformaldehyde (Sigma Aldrich) and 0.4% glutaraldehyde (Sigma Aldrich). Fixed foci

were then stained with a 0.1% crystal violet solution for 1 hours with gentle rotation at room temperature to visualize G418-resistant foci.

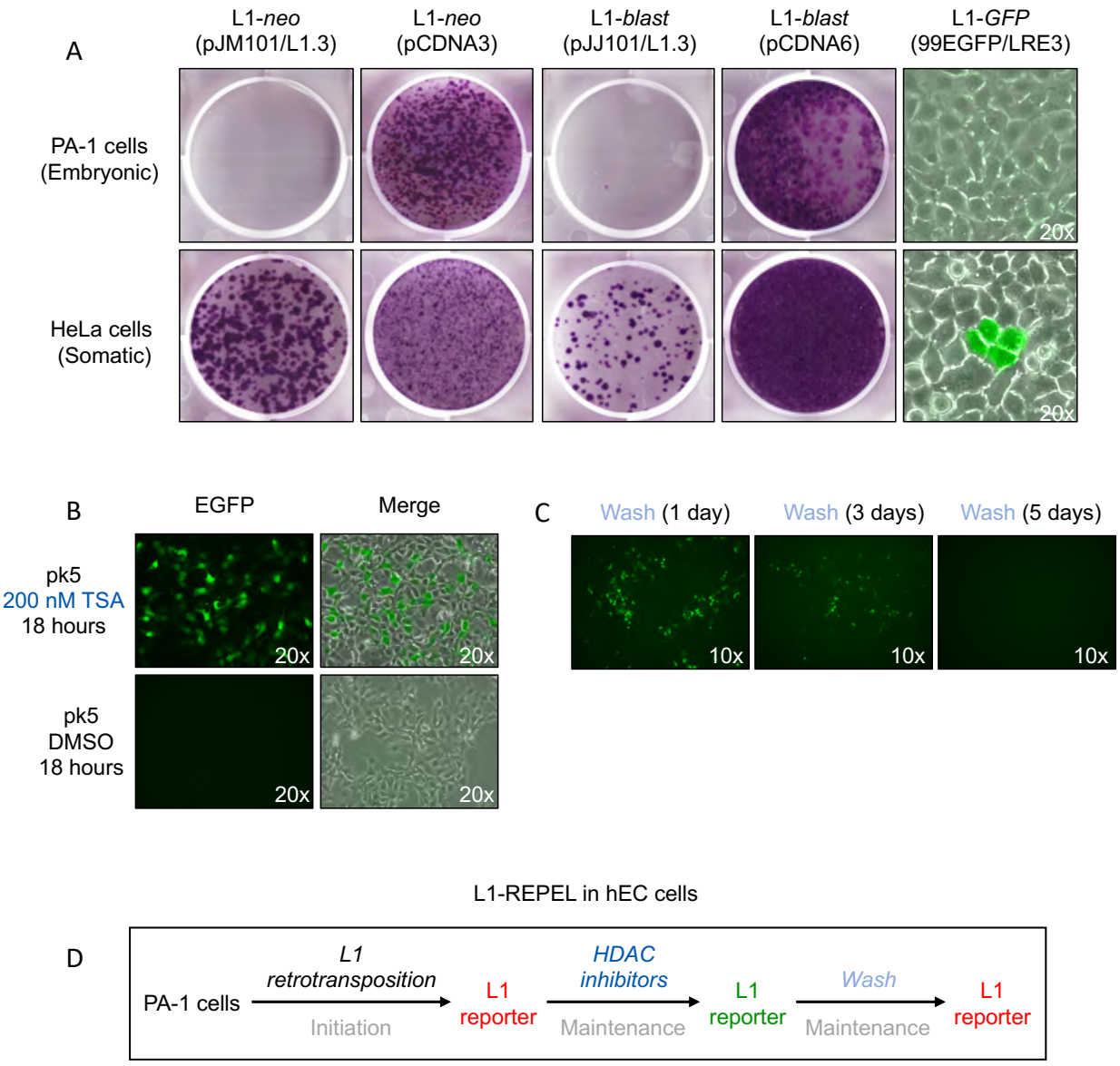
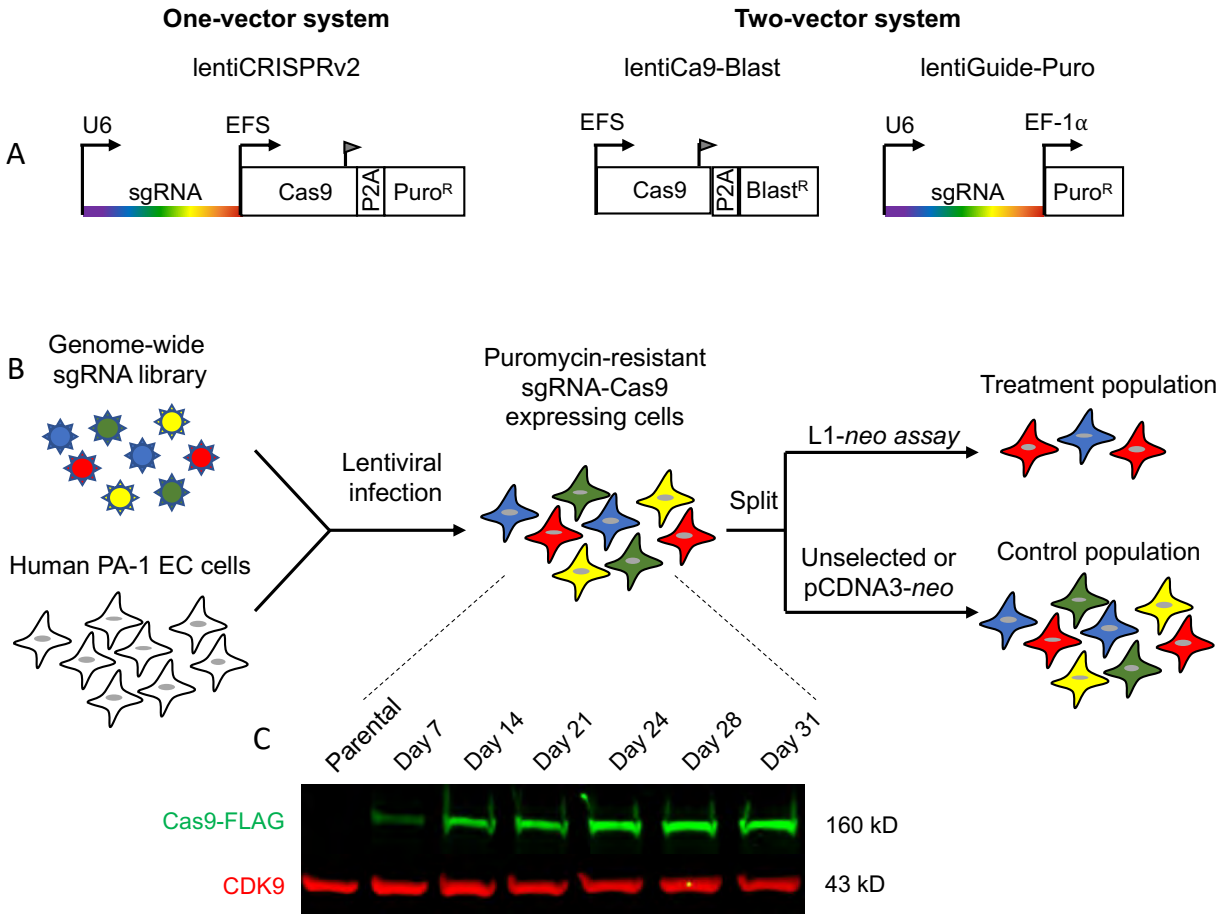


Figure 2.1: L1 reporter gene silencing in human embryonic carcinoma cells.

(A) Results of the L1 retrotransposition assay in PA-1 human embryonic carcinoma cells (top) and HeLa JVM cells (bottom) transfected with engineered L1 reporter constructs. G418-resistant (L1-*neo* assay) or blasticidin-resistant (L1-*blast* assay) colonies that expressed the L1-delivered reporter gene were fixed and stained with crystal violet for visualization. Cells expressing the L1-delivered *EGFP* (L1-*GFP assay*) reporter gene were imaged at 20x. Shown is a merged GFP/bright-field image. (B) pk5 cells were treated with a HDAC inhibitor (200nM TSA) or vehicle (DMSO) for eighteen hours. EGFP-expressing cells were imaged at 20x. (C) EGFP-expressing cells 1 day (left panel), 3 days

(middle panel) and 5 days (right panel) after removing TSA from the media (wash). EGFP-expressing cells were imaged at 10x. (D) L1-REPEL working model. Please see text for details.



D

Gene depletion within essential cellular pathways

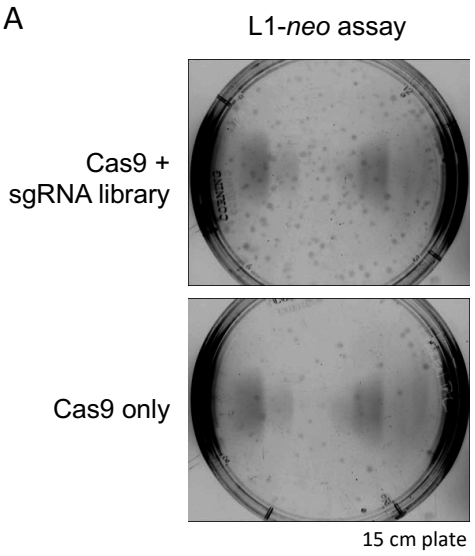
	# genes in pathway	Day 7 depleted	Day 21 depleted	Day 31 depleted	Brunello depleted
KEGG_RIBOSOME	88	2	42	38	28
KEGG_PROTEASOME	46	1	17	18	15
KEGG_RNA_POLYMERASE	29	3	9	10	14
KEGG_SPLICEOSOME	127	16	44	54	47

Figure 2.2: L1-REPEL screen in PA-1 hECs.

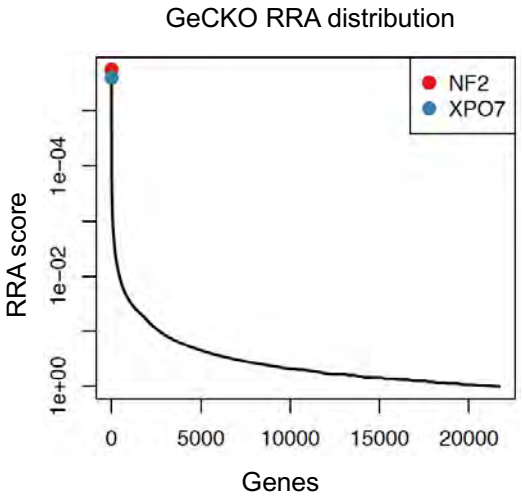
(A) Schematic of the lentiviral vectors used to deliver the genome-wide sgRNA libraries. LentiCRISPRv2 delivers both Cas9-FLAG and the sgRNA in one-vector. LentiCas9-Blast and lentiGuide-Puro are used in a two-vector system to deliver Cas9-FLAG and the sgRNA. P2A is a self-cleaving peptide. (B) Schematic of the L1-REPEL screen in PA-1 human embryonic carcinoma cells. (C) Western blot showing Cas9-FLAG protein expression at 7-, 14-, 24-, 28- and 31-days post lentiCRISPRv2 infection. CDK9 is

shown as a protein loading control. (D) MAGeCK-based identification of depleted genes with essential cellular pathways. The number of depleted genes, based on sgRNA read counts, is shown for 7-, 21- and 31-days post lentiCRISPRv2 infection containing the GeCKOv2 sgRNA library.

A



B



C

GeCKO screen 1

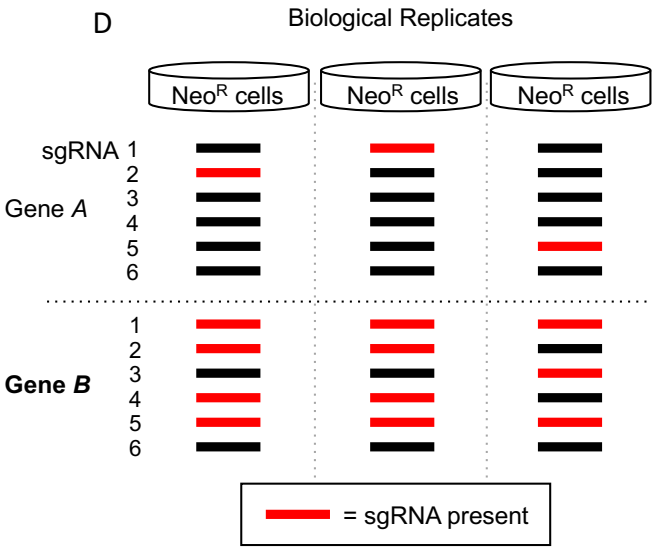
Rank	Gene	# sgRNAs
1	NF2	4
2	NIPSNAP3A	5
3	GGH	3
4	RAC3	4
5	SEMG2	2
6	mir-223	2
7	XPO7	5
8	LAMB4	2
9	TMOD3	3
10	WWP2	4

GeCKO screen 2

Rank	Gene	# sgRNAs
1	XPO7	4
2	NF2	4
3	UBE3B	3
4	LMO7	3
5	GMNN	2
6	GPATCH8	4
7	TADA2B	4
8	SLC1A3	3
9	TBC1D4	2
10	PI4KB	4

Brunello screen

Rank	Gene	# sgRNAs
1	NCAPG	2
2	RGL3	1
3	CT45A7	2
4	FAM46A	2
5	CKAP2	1
6	ISYNA1	1
7	HELB	1
8	MRPL4	2
9	PARD6G	2
10	MRPS11	1



E Genes with 2+ sgRNAs in 2+ replicates

sgRNA	GeCKO 1 reps	GeCKO 2 reps
NF2_19	8/10	7/7
NF2_60	6/10	3/7
NF2_18	1/10	2/7
NF2_17	0/10	2/7
XPO7_06	1/10	3/7
XPO7_07	2/10	2/7
XPO7_08	0/10	2/7

F Genes with 2+ sgRNAs in 2+ screens

Gene	GeCKO 1	GeCKO 2	Brunello
NF2	4	4	2
XPO7	3	3	0
RAC2	3	2	0
TADA2B	2	3	0
DCBLD2	2	2	0
DDX53	2	2	0
DGKK	2	2	0
H2AFY	2	2	0
HELT	2	2	0
LOC100130357	2	2	0
OR10V1	2	2	0
OR52I2	2	2	0
B9D2	2	2	1
HNF1A	2	2	1
SLC17A6	2	1	2
THEM5	2	0	2
HOXB1	1	2	2
SNX13	0	2	2
PRDM13	0	2	2
FASTKD1	0	2	2

Figure 2.3: Identification of L1-REPEL candidate genes.

(A) Results of the L1-*neo* assay in Cas9 + sgRNA expressing cells (top) or Cas9-only expressing cells (bottom). Unfixed 15 cm plates were imaged at 1x before genomic DNA collection. (B) MAGeCK-based Robust Rank Aggregation (RRA) distribution for GeCKO screens 1 and 2. The x-axis indicates all genes targeted within the GeCKOv2 sgRNA library. The y-axis indicates RRA score. (C) MAGeCK-based candidate gene rankings

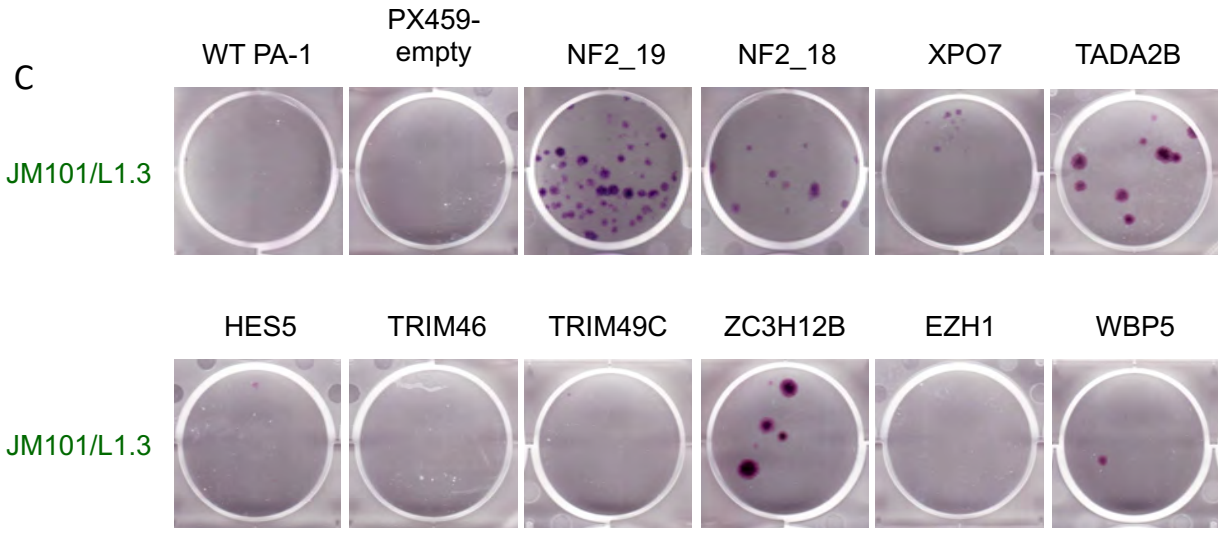
from three independent L1-REPEL screens. GeCKO screen 1 (left) was based on 10 biological replicates. GeCKO screen 2 (middle) was based on 7 biological replicates. The Brunello screen (right) was based on 5 biological replicates. The top 10 genes and the indicated number of enriched sgRNAs are shown. The GeCKO sgRNA library has 6 sgRNAs per gene. The Brunello sgRNA library has 4 sgRNAs per gene. (D) Schematic of the simplified method for L1-REPEL candidate gene ranking. G418-resistant cells are collected and sequenced to determine sgRNA read counts per biological replicate. Only the top ~500 sgRNAs per replicate were included in the analysis. Genes were ranked based on sgRNA presence or absence across biological replicates. Gene *B* demonstrates a more enriched and higher ranked candidate compared to gene *A.* (E) Manual identification of genes with 2 or more sgRNAs represented in two or more biological replicates. The number of biological replicates is indicated (F) Manual identification of genes with two or more sgRNAs in two or more screens. The number of identified sgRNAs is indicated.

A

B

sgRNA target	Forward oligo (5'-3')	Reverse oligo (5'-3')
NF2_19	CACCGTGAGCCTACCTTGGCCTGGA	AAACTCCAGGCCAAGGTAGGCTCAC
NF2_18	CACCGATTCCACGGGAAGGAGATCT	AAACAGATCTCCTTCCCGTGGAATC
NF2_60	CACCGCCTGGCTTCTTACGCCGTCC	AAACGGACGGCGTAAGAAGCCAGGC
XPO7	CACCGAGACACAACCACTCGACTCC	AAACGGAGTCGAGTGGTTGTGTCTC
TADA2B	CACCGCGAGCTGAAGCGCGCCCACG	AAACCGTGGGCGCGCTTCAGCTCGC
HES5	CACCGCGGCGCATCTTCTCCACCAC	AAACGTGGTGGAGAAGATGCGCCGC
TRIM46	CACCGGTGTGATCTTGTGCCCGCTG	AAACCAGCGGGCACAAGATCACACC
TRIM49C	CACCGAGGACTAACCTTCCAGCATC	AAACGATGCTGGAAGGTTAGTCCTC
ZC3H12B	CACCGCTTGGGCCGTGGCGTCCTAA	AAACTTAGGACGCCACGGCCCAAGC
EZH1	CACCGACAGGCTTCATTGACTGAAC	AAACGTTCAGTCAATGAAGCCTGTC
WBP5	CACCGTTTGGCTCTTCCTCATGCTT	AAACAAGCATGAGGAAGAGCCAAAC

C



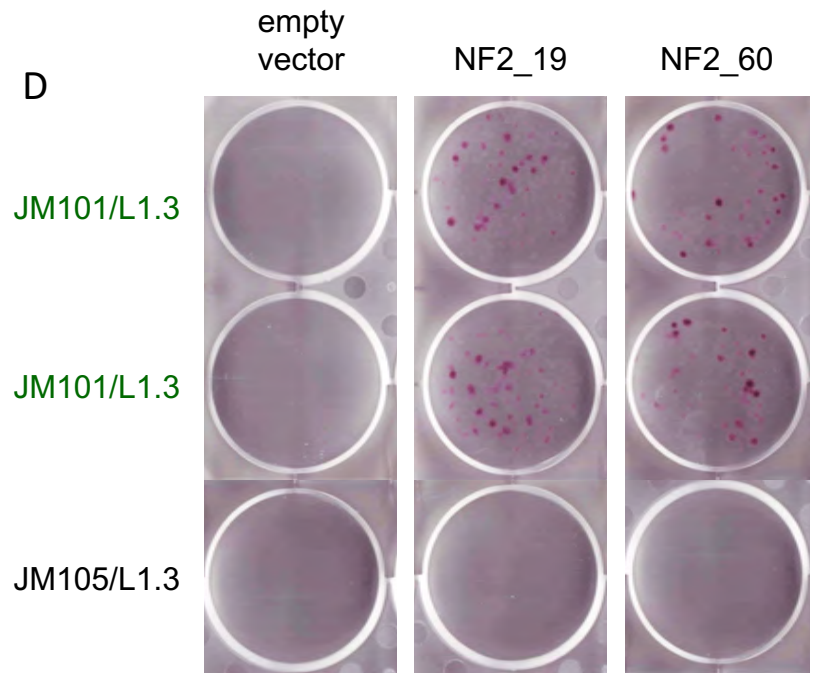


Figure 2.4: L1-REPEL candidate gene validation.

(A) Candidate gene validation strategy. PA-1 cells were transfected with a Cas9 + sgRNA expressing vector (PX459) targeting a single candidate gene. Then, puromycin-resistant "population knockout" cells were subjected to the L1-*neo* retrotransposition assay to determine L1-REPEL efficiency. (B) Summary table showing the oligos used to generate the candidate knockout vectors. Red nucleotides indicate the sgRNA-flanking restriction sites for PX459 cloning. C) Results of the candidate gene validation L1-*neo* retrotransposition assays. NF2_19 and NF2_18 are different sgRNAs targeting *NF2*. G418-resistant cells, expressing the L1-delivered reporter gene, were fixed and stained with crystal violet for visualization. (D) Results of the secondary *NF2* validation L1-*neo* retrotransposition assays. NF2_19 and NF2_60 are different sgRNAs targeting *NF2*. G418-resistant cells, expressing the L1-delivered reporter gene, were fixed and stained with crystal violet for visualization. Cells were transfected with wild-type L1 (pJM101/L1.3) or a retrotransposition deficient L1 (pJM105/L1.3) as indicated.

Table 2.1: Candidate gene list

Annotations provided by https://string-db.org (Szklarczyk et al., 2021).

Link: L1-REPEL candidate gene annotations

		Gene	Description
Highly enriched	1	NF2	Merlin; Probable regulator of the Hippo/SWH (Sav/Wts/Hpo) signaling pathway, a signaling
	2	XPO7	Exportin-7; Mediates the nuclear export of proteins (cargos) with broad substrate specificity.
	3	TADA2B	Transcriptional adapter 2-beta; Coactivates PAX5-dependent transcription together with eith
	4	ST6GALNAC1	ST6 N-acetylgalactosaminide alpha-2,6-sialyltransferase 1; Sialyltransferases
	5	H2AFY	Core histone macro-H2A.1; Variant histone H2A which replaces conventional H2A in a subset
	6	DYNC1H1	Cytoplasmic dynein 1 heavy chain 1; Cytoplasmic dynein 1 acts as a motor for the intracellula
	7	RTBDN	Retbindin; Riboflavin-binding protein which might have a role in retinal flavin transport
	8	SCN7A	Sodium channel protein type 7 subunit alpha; Mediates the voltage-dependent sodium ion pe
	9	MMAA	Methylmalonic aciduria type A protein, mitochondrial; GTPase, binds and hydrolyzes GTP. Inv
	10	TAOK1	Serine/threonine-protein kinase TAO1; Serine/threonine-protein kinase involved in various pr
	11	GGH	Gamma-glutamyl hydrolase; Hydrolyzes the polyglutamate sidechains of pteroylpolyglutama
	12	OR10V1	Olfactory receptor 10V1; Odorant receptor; Olfactory receptors, family 10
	13	TMEM86B	Lysoplasmalogenase; Enzyme catalyzing the degradation of lysoplasmalogen. Lysoplasmalog
	14	SORCS1	Sortilin related VPS10 domain containing receptor 1; Belongs to the VPS10-related sortilin fa
	15	TBC1D4	TBC1 domain family member 4; May act as a GTPase-activating protein for RAB2A, RAB8A,
	16	BMF	Bcl-2-modifying factor; May play a role in apoptosis. Isoform 1 seems to be the main initiato
	17	LMAN2	Vesicular integral-membrane protein VIP36; Plays a role as an intracellular lectin in the early
	18	NMBR	Neuromedin-B receptor; Receptor for neuromedin-B; Bombesin receptors
	19	CHSY3	Chondroitin sulfate synthase 3; Has both beta-1,3-glucuronic acid and beta-1,4-N- acetylgala
	20	LPAR2	Lysophosphatidic acid receptor 2; Receptor for lysophosphatidic acid (LPA), a mediator of div
Moderately enriched	21	APOL1	Apolipoprotein L1; May play a role in lipid exchange and transport throughout the body. May
	22	C11orf68	UPF0696 protein C11orf68; Chromosome 11 open reading frame 68; Belongs to the UPF0696
	23	SEPTIN6	Septin-6; Filament-forming cytoskeletal GTPase. Required for normal organization of the act
	24	ABL2	Abelson tyrosine-protein kinase 2; Non-receptor tyrosine-protein kinase that plays an ABL1- c
	25	LAMB4	Laminin subunit beta-4; Binding to cells via a high affinity receptor, laminin is thought to me
	26	OR3A1	Olfactory receptor 3A1; Odorant receptor; Olfactory receptors, family 3
	27	TRIM46	Tripartite motif-containing protein 46; Microtubule-associated protein that is involved in the
	28	PMPCA	Mitochondrial-processing peptidase subunit alpha; Cleaves presequences (transit peptides) f
	29	FAM115C	TRPM8 channel-associated factor 2; Isoform 2: Negatively regulates the plasma membrane
	30	GPATCH8	G-patch domain containing 8
	31	LRRC37A	Leucine-rich repeat-containing protein 37A; Leucine rich repeat containing 37A
	32	OST4	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 4; Acts as component
	33	SEMA4D	Semaphorin-4D; Cell surface receptor for PLXN1B and PLXNB2 that plays an important role in
	34	GAGE12C	G antigen 12C; Belongs to the GAGE family
	35	PEX13	Peroxisomal membrane protein PEX13; Component of the peroxisomal translocation machine
	36	NCK2	Cytoplasmic protein NCK2; Adapter protein which associates with tyrosine- phosphorylated g
	37	AJAP1	Adherens junction-associated protein 1; Plays a role in cell adhesion and cell migration
	38	POU6F2	POU domain, class 6, transcription factor 2; Probable transcription factor likely to be involved
	39	CCL19	G1/S-specific cyclin-D2; Regulatory component of the cyclin D2-CDK4 (DC) complex that phos
	40	GALNT12	Polypeptide N-acetylgalactosaminyltransferase 12; Catalyzes the initial reaction in O-linked c
	41	COPG2	Coatomer subunit gamma-2; The coatomer is a cytosolic protein complex that binds to dilysi
	42	CASP2	Caspase-2; Involved in the activation cascade of caspases responsible for apoptosis executior
	43	ACOT8	Acyl-coenzyme A thioesterase 8; Acyl-coenzyme A (acyl-CoA) thioesterases are a group of en
	44	CAMK2N2	Calcium/calmodulin-dependent protein kinase II inhibitor 2; Potent and specific cellular inhib
	45	MBL2	Mannose-binding protein C; Calcium-dependent lectin involved in innate immune defense. Bi
	46	DDOST	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase 48 kDa subunit; Essential sub
	47	NDUFB5	NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 5, mitochondrial; Accessory su
	48	VAV2	Guanine nucleotide exchange factor VAV2; Guanine nucleotide exchange factor for the Rho f
	49	PCDHGB5	Protocadherin gamma-B5; Potential calcium-dependent cell-adhesion protein. May be involv
	50	USP28	Ubiquitin carboxyl-terminal hydrolase 28; Deubiquitinase involved in DNA damage response

	51	ZNF771	Zinc finger protein 771; May be involved in transcriptional regulation; Zinc fingers C2H2-type
	52	MLL4	Histone-lysine N-methyltransferase 2D; Histone methyltransferase. Methylates 'Lys-4' of hist
	53	DCBLD2	Discoidin, CUB and LCCL domain containing 2
	54	KLK8	Kallikrein-8; Kallikrein related peptidase 8; Kallikreins
	55	NUCB1	Nucleobindin-1; Major calcium-binding protein of the Golgi. May have a role in calcium home
	56	PIP4K2B	Phosphatidylinositol 5-phosphate 4-kinase type-2 beta; Participates in the biosynthesis of pho
	57	ARIH2	E3 ubiquitin-protein ligase ARIH2; E3 ubiquitin-protein ligase, which catalyzes ubiquitination
	58	KRTAP4-2	Keratin-associated protein 4-2; In the hair cortex, hair keratin intermediate filaments are em
	59	TMC6	Transmembrane channel-like protein 6; Probable ion channel; Transmembrane channel likes
	60	B9D2	B9 domain-containing protein 2; Component of the tectonic-like complex, a complex localized
	61	PTAR1	Protein prenyltransferase alpha subunit repeat containing 1
	62	THSD7B	Thrombospondin type-1 domain-containing protein 7B; Thrombospondin type 1 domain conta
	63	HOXB1	Homeobox protein Hox-B1; Sequence-specific transcription factor which is part of a developm
	64	ZFP62	Zinc finger protein 62 homolog; May play a role in differentiating skeletal muscle; Zinc finger
	65	HNF1A	Hepatocyte nuclear factor 1-alpha; Transcriptional activator that regulates the tissue specific
	66	THEM5	Acyl-coenzyme A thioesterase THEM5; Has acyl-CoA thioesterase activity towards long-chain
	67	SLC17A6	Vesicular glutamate transporter 2; Mediates the uptake of glutamate into synaptic vesicles a
	68	THBS1	Thrombospondin-1; Adhesive glycoprotein that mediates cell-to-cell and cell-to-matrix intera
	69	PKDCC	Extracellular tyrosine-protein kinase PKDCC; Secreted tyrosine-protein kinase that mediates p
	70	RD3L	Protein RD3-like; Retinal degeneration 3-like
	71	SAP25	Histone deacetylase complex subunit SAP25; Involved in the transcriptional repression media
	72	ANKRD32	SMC5-SMC6 complex localization factor protein 1; Plays a role in the DNA damage response
	73	HBA1	Hemoglobin subunit alpha; Involved in oxygen transport from the lung to the various peripher
	74	FAT4	Protocadherin Fat 4; Cadherins are calcium-dependent cell adhesion proteins. FAT4 plays a ro
	75	HRH4	Histamine H4 receptor; The H4 subclass of histamine receptors could mediate the histamine
	76	RPS12	Ribosomal protein S12
	77	KMT2E	Histone-lysine N-methyltransferase 2E; Histone methyltransferase that specifically mono- ar
	78	AMICA1	Junctional adhesion molecule-like; Transmembrane protein of the plasma membrane of leuk
	79	TMEM184A	Transmembrane protein 184A; Acts as a heparin receptor in vascular cells (By similarity). Ma
	80	PPP3CA	Serine/threonine-protein phosphatase 2B catalytic subunit alpha isoform; Calcium-dependen
	81	TNFRSF9	Tumor necrosis factor receptor superfamily member 9; Receptor for TNFSF9/4-1BBL. Possibl
	82	LRRC8D	Volume-regulated anion channel subunit LRRC8D; Non-essential component of the volume-re
	83	hsa-mir-4436a	microRNA
	84	ZNF648	Zinc finger protein 648; May be involved in transcriptional regulation; Zinc fingers C2H2-type
	85	hsa-mir-223	microRNA ZC3H12B, TAOK1, SEPTIN6, SIAH1, PARP1
	86	SLCO3A1	Solute carrier organic anion transporter family member 3A1; Mediates the Na(+)-independer
	87	C10orf99	Protein GPR15L; Chemotactic factor that mediates lymphocytes recruitement to epithelia thr
	88	PAWR	PRKC apoptosis WT1 regulator protein; Pro-apoptopic protein capable of selectively inducing
	89	ENTPD2	Ectonucleoside triphosphate diphosphohydrolase 2; In the nervous system, could hydrolyze AT
	90	RABGAP1	Rab GTPase-activating protein 1; May act as a GTPase-activating protein of RAB6A. May play
	91	hsa-mir-1307	microRNA
	92	KCNQ1	Potassium voltage-gated channel subfamily KQT member 1; Potassium channel that plays ar
	93	RPS6KA2	Ribosomal protein S6 kinase alpha-2; Serine/threonine-protein kinase that acts downstream
	94	MAGEF1	Melanoma-associated antigen F1; May enhance ubiquitin ligase activity of RING-type zinc fir
	95	hsa-mir-4268	microRNA
	96	ALDH1L1	Cytosolic 10-formyltetrahydrofolate dehydrogenase; Aldehyde dehydrogenase 1 family memb
	97	NIPSNAP3A	Nipsnap homolog 3A; Belongs to the NipSnap family
	98	OR6B2	Olfactory receptor 6B2; Odorant receptor; Olfactory receptors, family 6
	99	TTC31	Tetratricopeptide repeat domain containing
	100	DRC1	Dynein regulatory complex protein 1; Key component of the nexin-dynein regulatory complex

	101	LIMCH1	LIM and calponin homology domains 1
	102	C15orf57	Coiled-coil domain-containing protein 32; Chromosome 15 open reading frame 57
	103	RAC3	Ras-related C3 botulinum toxin substrate 3; Plasma membrane-associated small GTPase wh
	104	DET1	DET1 homolog; Component of the E3 ubiquitin ligase DCX DET1-COP1 complex, which is requ
	105	C3orf18	Uncharacterized protein C3orf18; Chromosome 3 open reading frame 18
	106	STS	Steryl-sulfatase; Conversion of sulfated steroid precursors to estrogens during pregnancy; Sul
	107	RNF186	E3 ubiquitin protein ligase that is part of an apoptotic signaling pathway activated by endopl
	108	ZNF570	Zinc finger protein 570; May be involved in transcriptional regulation; Zinc fingers C2H2-type
	109	GSK3A	Glycogen synthase kinase-3 alpha; Constitutively active protein kinase that acts as a negative
	110	WWP2	NEDD4-like E3 ubiquitin-protein ligase WWP2; E3 ubiquitin-protein ligase which accepts ubic
	111	TMOD3	Tropomodulin-3; Blocks the elongation and depolymerization of the actin filaments at the po
	112	TAF5L	TAF5-like RNA polymerase II p300/CBP-associated factor-associated factor 65 kDa subunit 5l
	113	GJA1	Gap junction alpha-1 protein; Gap junction protein that acts as a regulator of bladder capacit
	114	C4A	Complement C4-A; Non-enzymatic component of C3 and C5 convertases and thus essential f
	115	CHRNB2	Neuronal acetylcholine receptor subunit beta-2; After binding acetylcholine, the AChR respon
	116	FAM195A	MAPK regulated corepressor interacting protein 2; Family with sequence similarity 195, mem
	117	KCNK7	Potassium channel subfamily K member 7; Probable potassium channel subunit. No channel
	118	GRM1	Metabotropic glutamate receptor 1; G-protein coupled receptor for glutamate. Ligand bindin
	119	MTMR2	Myotubularin-related protein 2; Phosphatase that acts on lipids with a phosphoinositol headg
	120	GPR101	Probable G-protein coupled receptor 101; Orphan receptor; G protein-coupled receptors, Clas
	121	CMIP	C-Maf-inducing protein; Plays a role in T-cell signaling pathway. Isoform 2 may play a role in
	122	ZNF513	Zinc finger protein 513; Transcriptional regulator that plays a role in retinal development and
	123	WDR1	WD repeat-containing protein 1; Induces disassembly of actin filaments in conjunction with A
	124	COX8A	Cytochrome c oxidase subunit 8A, mitochondrial; This protein is one of the nuclear-coded pol
	125	DNAJA3	DnaJ homolog subfamily A member 3, mitochondrial; Modulates apoptotic signal transductic
	126	IL36B	Interleukin-36 beta; Cytokine that binds to and signals through the IL1RL2/IL-36R receptor wh
	127	hsa-mir-127	microRNA WWP2
	128	RNF7	RING-box protein 2; Probable component of the SCF (SKP1-CUL1-F-box protein) E3 ubiquitin l
	129	hsa-mir-4445	microRNA RNF43, PPP3CA
	130	NEFL	Neurofilament light polypeptide; Neurofilaments usually contain three intermediate filamen
	131	SLC6A13	Sodium- and chloride-dependent GABA transporter 2; Sodium-dependent GABA and taurine t
	132	NTRK3	NT-3 growth factor receptor; Receptor tyrosine kinase involved in nervous system and probab
	133	RBMY1D	RNA-binding protein which may be involved in spermatogenesis. Required for sperm develop
	134	P2RY4	P2Y purinoceptor 4; Receptor for UTP and UDP coupled to G-proteins that activate a phospha
	135	CCDC19	WASH complex subunit 3; Acts at least in part as component of the WASH core complex whc
	136	CNTN3	Contactin-3; Contactins mediate cell surface interactions during nervous system developmen
	137	LMO7	LIM domain only protein 7; LIM domain containing
	138	PCED1B	PC-esterase domain containing 1B; Belongs to the PC-esterase family
	139	hsa-mir-4701	microRNA YWHAZ, XPO7
	140	SHISA6	Protein shisa-6 homolog; Shisa family member 6
	141	DMKN	Dermokine; May act as a soluble regulator of keratinocyte differentiation
	142	UBE3B	Ubiquitin-protein ligase E3B; E3 ubiquitin-protein ligase which accepts ubiquitin from an E2 u
	143	CCDC53	C-C motif chemokine 19; May play a role not only in inflammatory and immunological respor
	144	SNCG	Gamma-synuclein; Plays a role in neurofilament network integrity. May be involved in modul
	145	EGLN2	Egl nine homolog 2; Cellular oxygen sensor that catalyzes, under normoxic conditions, the po
	146	TMEM64	Transmembrane protein 64; Positively regulates TNFSF11-induced osteoclast differentiation.
	147	MOS	Belongs to the protein kinase superfamily. Ser/Thr protein kinase family
	148	hsa-mir-7641-2	microRNA
	149	APH1B	Gamma-secretase subunit APH-1B; Probable subunit of the gamma-secretase complex, an e
	150	CANX	Calnexin; Calcium-binding protein that interacts with newly synthesized glycoproteins in the

	151	LDLRAD1	Low density lipoprotein receptor class A domain containing 1
	152	CTSA	Lysosomal protective protein; Protective protein appears to be essential for both the activity
	153	TSHB	Thyrotropin subunit beta; Indispensable for the control of thyroid structure and metabolism;
	154	ZNF638	Zinc finger protein 638; Early regulator of adipogenesis that works as a transcription cofactor
	155	MGST2	Microsomal glutathione S-transferase 2; Can catalyze the production of LTC4 from LTA4 and
	156	DST	Dystonin; Cytoskeletal linker protein. Acts as an integrator of intermediate filaments, actin a
	157	MICU3	Calcium uptake protein 3, mitochondrial; May play a role in mitochondrial calcium uptake; EF
	158	MRPL4	Mitochondrial ribosomal protein L4; Belongs to the universal ribosomal protein uL4 family
	159	SLC16A14	Monocarboxylate transporter 14; Proton-linked monocarboxylate transporter. May catalyze th
	160	CERK	Ceramide synthase 3; Has (dihydro)ceramide synthesis activity with relatively broad substrat
	161	RSL24D1	Probable ribosome biogenesis protein RLP24; Involved in the biogenesis of the 60S ribosomal
	162	CETN3	Cilia and flagella associated protein 45; Belongs to the CFAP45 family
	163	RAB5B	Ras-related protein Rab-5B; Protein transport. Probably involved in vesicular traffic (By simil
	164	KRTAP17-1	Keratin-associated protein 17-1; In the hair cortex, hair keratin intermediate filaments are er
	165	TOM1L2	TOM1-like protein 2; Probable role in protein transport. May regulate growth factor-induced
	166	RIMBP3	RIMS-binding protein 3A; Component of the manchette, a microtubule-based structure which
	167	TCEA2	Transcription elongation factor A protein 2; Necessary for efficient RNA polymerase II transcr
	168	HLA-DMB	HLA class II histocompatibility antigen, DM beta chain; Plays a critical role in catalyzing the re
	169	PARD6G	Partitioning defective 6 homolog gamma; Adapter protein involved in asymmetrical cell divis
	170	TNKS	Tankyrase-1; Poly-ADP-ribosyltransferase involved in various processes such as Wnt signaling
	171	NARFL	Cytosolic Fe-S cluster assembly factor NARFL; Component of the cytosolic iron-sulfur protein
	172	WDPCP	WD repeat-containing and planar cell polarity effector protein fritz homolog; Probable effect
	173	BROX	BRO1 domain-containing protein BROX; BRO1 domain and CAAX motif containing; Belongs t
	174	OR51B6	Olfactory receptor 51B6; Odorant receptor; Olfactory receptors, family 51
	175	PP2D1	Protein phosphatase 2C like domain containing 1
	176	CT45A7	Cancer/testis antigen family 45 member A5; Belongs to the CT45 family
	177	KCNK2	Potassium channel subfamily K member 2; Ion channel that contributes to passive transmem
	178	LALBA	Alpha-lactalbumin; Regulatory subunit of lactose synthase, changes the substrate specificity
	179	IGF2BP2	Insulin-like growth factor 2 mRNA-binding protein 2; RNA-binding factor that recruits target
	180	ATP2C2	Calcium-transporting ATPase type 2C member 2; This magnesium-dependent enzyme catalyz
	181	EXOSC1	Exosome complex component CSL4; Non-catalytic component of the RNA exosome complex
	182	NLN	Neurolysin, mitochondrial; Hydrolyzes oligopeptides such as neurotensin, bradykinin and dync
	183	FAM46A	Putative nucleotidyltransferase FAM46A; Probable nucleotidyltransferase that may act as a r
	184	ATG4C	Cysteine protease ATG4C; Cysteine protease required for the cytoplasm to vacuole transport
	185	RIPK1	Receptor-interacting serine/threonine-protein kinase 1; Serine-threonine kinase which transc
	186	WTAP	Pre-mRNA-splicing regulator WTAP; Regulatory subunit of the WMM N6-methyltransferase c
	187	SERPINB10	Serpin B10; Protease inhibitor that may play a role in the regulation of protease activities du
	188	ABCC1	Multidrug resistance-associated protein 1; Mediates export of organic anions and drugs from
	189	SPATA31D3	Spermatogenesis-associated protein 31D1; May play a role in spermatogenesis
	190	XYLB	Xylulose kinase; Phosphorylates D-xylulose to produce D-xylulose 5- phosphate, a molecule th
	191	TFCP2	Alpha-globin transcription factor CP2; Binds a variety of cellular and viral promoters including
	192	LY6D	Lymphocyte antigen 6D; May act as a specification marker at earliest stage specification of ly
	193	TREM1	Triggering receptor expressed on myeloid cells 1; Stimulates neutrophil and monocyte-media
	194	GSDMD	Gasdermin-D; Gasdermin-D, N-terminal: Promotes pyroptosis in response to microbial infect
	195	AMOT	Angiomotin; Plays a central role in tight junction maintenance via the complex formed with A
	196	IVL	Involucrin; Part of the insoluble cornified cell envelope (CE) of stratified squamous epithelia;
	197	NCAPG	Condensin complex subunit 3; Regulatory subunit of the condensin complex, a complex requi
	198	TREH	Trehalase; Intestinal trehalase is probably involved in the hydrolysis of ingested trehalose; Be
	199	ZNF197	Zinc finger protein 197; May be involved in transcriptional regulation; SCAN domain containir
	200	ZNF227	Zinc finger protein 227; May be involved in transcriptional regulation; Belongs to the krueppe

	201	SAC3D1	SAC3 domain-containing protein 1; Involved in centrosome duplication and mitotic progressio
	202	MCM10	Protein MCM10 homolog; Acts as a replication initiation factor that brings together the MCM
	203	LINGO1	Leucine-rich repeat and immunoglobulin-like domain-containing nogo receptor-interacting pr
	204	CERS3	Centrin-3; EF-hand domain containing
	205	SCLY	Selenocysteine lyase; Catalyzes the decomposition of L-selenocysteine to L- alanine and elem
	206	hsa-mir-3186	microRNA
	207	ANGPTL4	Angiopoietin-related protein 4; Protein with hypoxia-induced expression in endothelial cells.
	208	hsa-mir-6759	microRNA H2AFZ,
	209	CSPG5	Chondroitin sulfate proteoglycan 5; May function as a growth and differentiation factor involv
	210	ADORA2B	Adenosine receptor A2b; Receptor for adenosine. The activity of this receptor is mediated by
	211	TMEM14B	Primate-specific protein involved in cortical expansion and folding in the developing neocorte
	212	SLC35A3	UDP-N-acetylglucosamine transporter; Uridine diphosphate-N-acetylglucosamine (UDP-GlcN
	213	RBP3	Retinol-binding protein 3; IRBP shuttles 11-cis and all trans retinoids between the retinol iso
	214	NARS2	Probable asparagine--tRNA ligase, mitochondrial; asparaginyl-tRNA synthetase 2, mitochonc
	215	ZNF133	Zinc finger protein 133; May be involved in transcriptional regulation as a repressor; Zinc fing
	216	TMEM215	Transmembrane protein 215
	217	C2CD4A	C2 calcium-dependent domain-containing protein 4A; May be involved in inflammatory proce
	218	MON1B	MON1 homolog B, secretory trafficking associated; Belongs to the MON1/SAND family
	219	TCF25	Transcription factor 25; May play a role in cell death control. Acts as a transcriptional repress
	220	FAM209A	Protein FAM209A; Family with sequence similarity 209 member A
	221	FAM129A	Protein Niban; Regulates phosphorylation of a number of proteins involved in translation reg
	222	SHOX_X	This gene belongs to the paired homeobox family and is located in the pseudoautosomal reg
	223	NPPB	Natriuretic peptides B; Cardiac hormone which may function as a paracrine antifibrotic facto
	224	GON4L	GON-4-like protein; Has transcriptional repressor activity, probably as part of a complex with
	225	BTBD17	BTB/POZ domain-containing protein 17; BTB domain containing 17
	226	TRIM72	Tripartite motif-containing protein 72; Muscle-specific protein that plays a central role in cell
	227	TRMT112	Multifunctional methyltransferase subunit TRM112-like protein; Acts as an activator of both
	228	MYO5B	Unconventional myosin-Vb; May be involved in vesicular trafficking via its association with th
	229	C8orf48	Uncharacterized protein C8orf48; Chromosome 8 open reading frame 48
	230	NEK6	Serine/threonine-protein kinase Nek6; Protein kinase which plays an important role in mitoti
	231	ZNF747	KRAB domain-containing protein ZNF747; Zinc finger protein 747; Zinc fingers C2H2-type
	232	TFAP2D	Transcription factor AP-2-delta; Sequence-specific DNA-binding protein that interacts with in
	233	hsa-mir-3166	microRNA TTC31, YWHAZ, OGT
	234	LENG8	Leukocyte receptor cluster member 8
	235	GRIA3	Glutamate receptor 3; Receptor for glutamate that functions as ligand-gated ion channel in t
	236	GPR124	Adhesion G protein-coupled receptor A2; Endothelial receptor which functions as a WNT7-sp
	237	FAM104A	Protein FAM104A; Family with sequence similarity 104 member A
	238	APOF	Apolipoprotein F; Minor apolipoprotein that associates with LDL. Inhibits cholesteryl ester tra
	239	TMED3	Transmembrane emp24 domain-containing protein 3; Potential role in vesicular protein traff
	240	MARCKS	Myristoylated alanine-rich C-kinase substrate; MARCKS is the most prominent cellular substr
	241	KIAA0907	Protein BLOM7; RNA-binding protein involved in pre-mRNA splicing. Interacts with the PRP19
	242	FNIP2	Folliculin-interacting protein 2; Acts as a co-chaperone of HSP90AA1. Inhibits the ATPase acti
	243	MS4A4A	Membrane-spanning 4-domains subfamily A member 4A; May be involved in signal transduc
	244	SLC25A4	ADP/ATP translocase 1; Involved in mitochondrial ADP/ATP transport. Catalyzes the exchange
	245	C17orf64	Uncharacterized protein C17orf64; Chromosome 17 open reading frame 64
	246	TRABD2A	Metalloprotease TIKI1; Metalloprotease that acts as a negative regulator of the Wnt signalin
	247	ATP6V0C	V-type proton ATPase 16 kDa proteolipid subunit; Proton-conducting pore forming subunit of
	248	HECTD3	E3 ubiquitin-protein ligase HECTD3; E3 ubiquitin ligases accepts ubiquitin from an E2 ubiquiti
	249	FUCA1	Tissue alpha-L-fucosidase; Alpha-L-fucosidase is responsible for hydrolyzing the alpha-1,6-lin
	250	ZNF853	Zinc finger protein 853; Zinc fingers C2H2-type

	251	TDGF1	Teratocarcinoma-derived growth factor 1; GPI-anchored cell membrane protein involved in N
	252	hsa-mir-3606	microRNA YWHAZ,
	253	MAGOH	Protein mago nashi homolog; Core component of the splicing-dependent multiprotein exon ju
	254	XAGE1D	X antigen family member 1A
	255	IGFBP4	Insulin-like growth factor-binding protein 4; IGF-binding proteins prolong the half-life of the
	256	FOXJ1	Forkhead box protein J1; Transcription factor specifically required for the formation of motile
	257	SLC25A35	Solute carrier family 25 member 35; Belongs to the mitochondrial carrier (TC 2.A.29) family
	258	ARHGEF35	Rho guanine nucleotide exchange factor 35
	259	CSNK2B	Casein kinase II subunit beta; Participates in Wnt signaling (By similarity). Plays a complex ro
	260	HHIPL1	HHIP-like protein 1; Scavenger receptor cysteine rich domain containing
	261	PARD6A	Partitioning defective 6 homolog alpha; Adapter protein involved in asymmetrical cell divisio
	262	SNURF	SNRPN upstream reading frame
	263	COX6A1	Cytochrome c oxidase subunit 6A1, mitochondrial; This protein is one of the nuclear-coded po
	264	GPR17	Uracil nucleotide/cysteinyl leukotriene receptor; Dual specificity receptor for uracil nucleotide
	265	LURAP1L	Leucine rich adaptor protein 1 like
	266	FSTL3	Follistatin-related protein 3; Isoform 1 or the secreted form is a binding and antagonizing pro
	267	TCF7	Transcription factor 7; Transcriptional activator involved in T-cell lymphocyte differentiation.
	268	RSL1D1	Ribosomal L1 domain-containing protein 1; Regulates cellular senescence through inhibition
	269	LDOC1	Protein LDOC1; May have an important role in the development and/or progression of some
	270	DEFB110	Beta-defensin 110; Has antibacterial activity; Defensins, beta
	271	LNX1	E3 ubiquitin-protein ligase LNX; E3 ubiquitin-protein ligase that mediates ubiquitination and
	272	ZNF485	Zinc finger protein 485; May be involved in transcriptional regulation; Zinc fingers C2H2-type
	273	RBM15	Putative RNA-binding protein 15; May function as an mRNA export factor, stimulating export
	274	IZUMO4	Izumo sperm-egg fusion protein 4; IZUMO family member 4; Belongs to the Izumo family
	275	CCDC12	Coiled-coil domain containing 12; Spliceosomal Bact complex
	276	MSRB3	Methionine-R-sulfoxide reductase B3; Catalyzes the reduction of free and protein-bound met
	277	ALOX5	Arachidonate 5-lipoxygenase; Catalyzes the first step in leukotriene biosynthesis, and thereby
	278	SMPD4	Sphingomyelin phosphodiesterase 4; Catalyzes the hydrolysis of membrane sphingomyelin to
	279	FAS	Tumor necrosis factor receptor superfamily member 6; Receptor for TNFSF6/FASLG. The ada
	280	PIN1	Peptidyl-prolyl cis-trans isomerase NIMA-interacting 1; Peptidyl-prolyl cis/trans isomerase (P
	281	PIWIL3	Piwi-like protein 3; May play a role during spermatogenesis by repressing transposable elem
	282	FAM163A	Protein FAM163A; Family with sequence similarity 163 member A; Belongs to the FAM163 fa
	283	OGFOD2	2-oxoglutarate and iron dependent oxygenase domain containing 2
	284	AKR1B15	Aldo-keto reductase family 1 member B15; Isoform 1: Mainly acts as a reductive enzyme tha
	285	MT3	Metallothionein-3; Binds heavy metals. Contains three zinc and three copper atoms per polyp
	286	CTF1	Cardiotrophin-1; Induces cardiac myocyte hypertrophy in vitro. Binds to and activates the ILST
	287	USPL1	SUMO-specific isopeptidase USPL1; SUMO-specific isopeptidase involved in protein desumoy
	288	ACOT2	Acyl-coenzyme A thioesterase 2, mitochondrial; Acyl-CoA thioesterases are a group of enzym
	289	IER3	Radiation-inducible immediate-early gene IEX-1; May play a role in the ERK signaling pathwa
	290	LILRA2	Leukocyte immunoglobulin-like receptor subfamily A member 2; Part of the innate immune r
	291	PAX1	Paired box protein Pax-1; This protein is a transcriptional activator. It may play a role in the fo
	292	B4GALT2	Beta-1,4-galactosyltransferase 2; Responsible for the synthesis of complex-type N-linked olig
	293	SPHK1	Sphingosine kinase 1; Catalyzes the phosphorylation of sphingosine to form sphingosine 1-ph
	294	SIPA1L1	Signal-induced proliferation-associated 1-like protein 1; Stimulates the GTPase activity of RA
	295	USP17L19	Ubiquitin carboxyl-terminal hydrolase 17-like protein 19; Deubiquitinating enzyme that remo
	296	MCAM	Cell surface glycoprotein MUC18; Plays a role in cell adhesion, and in cohesion of the endothe
	297	CALY	Neuron-specific vesicular protein calcyon; Interacts with clathrin light chain A and stimulates
	298	CLRN2	Clarin-2; Clarin 2; Clarins
	299	POFUT2	GDP-fucose protein O-fucosyltransferase 2; Catalyzes the reaction that attaches fucose throu

	300	RNF43	E3 ubiquitin-protein ligase RNF43; E3 ubiquitin-protein ligase that acts as a negative regulat
	301	OR6Y1	Olfactory receptor 6Y1; Odorant receptor; Olfactory receptors, family 6
	302	GLI3	Transcriptional activator GLI3; Has a dual function as a transcriptional activator and a repress
	303	ENOPH1	Enolase-phosphatase E1; Bifunctional enzyme that catalyzes the enolization of 2,3-diketo-5-r
	304	COL13A1	Collagen alpha-1(XIII) chain; Involved in cell-matrix and cell-cell adhesion interactions that ar
	305	ZMAT3	Zinc finger matrin-type protein 3; Acts as a bona fide target gene of p53/TP53. May play a ro
	306	ANGPT1	Angiopoietin-1; Binds and activates TEK/TIE2 receptor by inducing its dimerization and tyrosi
	307	NLRX1	NLR family member X1; Participates in antiviral signaling. Acts as a negative regulator of MA
	308	LHX1	LIM/homeobox protein Lhx1; Potential transcription factor. May play a role in early mesoderr
	309	NANOGNB	NANOG neighbor homeobox
	310	ZNF775	Zinc finger protein 775; May be involved in transcriptional regulation; Zinc fingers C2H2-type
	311	MSL2	E3 ubiquitin-protein ligase MSL2; Component of histone acetyltransferase complex responsib
	312	MOXD1	Monooxygenase DBH like 1; Belongs to the copper type II ascorbate-dependent monooxygena
	313	UGT2A2	UDP glucuronosyltransferase 2 family, polypeptide A2; UDP-glucuronosyltransferases catalyz
	314	LATS2	Serine/threonine-protein kinase LATS2; Negative regulator of YAP1 in the Hippo signaling pat
	315	BCMO1	Beta,beta-carotene 15,15'-dioxygenase; Symmetrically cleaves beta-carotene into two molec
	316	CT47B1	Cancer/testis antigen family 47, member B1
	317	C7orf57	Uncharacterized protein C7orf57; Chromosome 7 open reading frame 57
	318	CD80	Cyclin-dependent kinase inhibitor 2A; Acts as a negative regulator of the proliferation of norr
	319	SPRY1	Protein sprouty homolog 1; May function as an antagonist of fibroblast growth factor (FGF) p
	320	RNASE3	Eosinophil cationic protein; Cytotoxin and helminthotoxin with low-efficiency ribonuclease ac
	321	PEA15	Astrocytic phosphoprotein PEA-15; Blocks Ras-mediated inhibition of integrin activation and
	322	PREB	Prolactin regulatory element-binding protein; Guanine nucleotide exchange factor that specif
	323	THBS4	Thrombospondin-4; Adhesive glycoprotein that mediates cell-to-cell and cell-to-matrix intera
	324	HCN2	Potassium/sodium hyperpolarization-activated cyclic nucleotide-gated channel 2; Hyperpolar
	325	HPCAL1	Hippocalcin-like protein 1; May be involved in the calcium-dependent regulation of rhodopsin
	326	LATS1	Serine/threonine-protein kinase LATS1; Negative regulator of YAP1 in the Hippo signaling pat
	327	ITFG3	Protein FAM234A; Integrin alpha FG-GAP repeat containing 3; Belongs to the FAM234 family
	328	COL6A5	Collagen alpha-5(VI) chain; Collagen VI acts as a cell-binding protein; Collagens
	329	SLC24A4	Sodium/potassium/calcium exchanger 4; Transports 1 Ca(2+) and 1 K(+) in exchange for 4 Na
	330	IER5	Immediate early response gene 5 protein; Plays a role as a transcription factor. Mediates pos
	331	TRIM39	E3 ubiquitin-protein ligase TRIM39; E3 ubiquitin-protein ligase. May facilitate apoptosis by in
	332	KIAA0922	Transmembrane protein 131-like; Isoform 1: Membrane-associated form that antagonizes ca
	333	AP3S1	AP-3 complex subunit sigma-1; Part of the AP-3 complex, an adaptor-related complex which
	334	C19orf54	UPF0692 protein C19orf54; Chromosome 19 open reading frame 54; Belongs to the UPF0692
	335	TST	Thiosulfate sulfurtransferase; Formation of iron-sulfur complexes, cyanide detoxification or r
	336	FCRL5	Fc receptor-like protein 5; May be involved in B-cell development and differentiation in perip
	337	LRRC8B	Volume-regulated anion channel subunit LRRC8B; Non-essential component of the volume-re
	338	MAFG	Transcription factor MafG; Since they lack a putative transactivation domain, the small Mafs
	339	UNC5D	Netrin receptor UNC5D; Receptor for the netrin NTN4 that promotes neuronal cell survival (B
	340	ATXN1	Ataxin-1; Chromatin-binding factor that repress Notch signaling in the absence of Notch intra
	341	LIPJ	Lipase family member J; Belongs to the AB hydrolase superfamily. Lipase family
	342	USP5	Ubiquitin carboxyl-terminal hydrolase 5; Cleaves linear and branched multiubiquitin polymers
	343	QSOX1	Sulfhydryl oxidase 1; Catalyzes the oxidation of sulfhydryl groups in peptide and protein thiols
	344	ZNF488	Zinc finger protein 488; May be involved in transcriptional regulation; Belongs to the krueppe
	345	FPR2	N-formyl peptide receptor 2; Low affinity receptor for N-formyl-methionyl peptides, which ar
	346	HDAC10	Histone deacetylase 10; Responsible for the deacetylation of lysine residues on the N-termin
	347	ACD	Adrenocortical dysplasia protein homolog; Component of the shelterin complex (telosome) th
	348	POMC	Pro-opiomelanocortin; Met-enkephalin: Endogenous opiate; Belongs to the POMC family
	349	FAM193B	Protein FAM193B; Family with sequence similarity 193 member B; Belongs to the FAM193 fa
	350	ZNF860	Zinc finger protein 860; May be involved in transcriptional regulation; Zinc fingers C2H2-type

	351	ZC3H15	Zinc finger CCCH domain-containing protein 15; Protects DRG1 from proteolytic degradation;
	352	ZNF233	Zinc finger protein 233; May be involved in transcriptional regulation; Zinc fingers C2H2-type
	353	CDKN2A	Ceramide kinase; Catalyzes specifically the phosphorylation of ceramide to form ceramide 1-
	354	KCNJ16	Inward rectifier potassium channel 16; Inward rectifier potassium channels are characterized
	355	APITD1	Centromere protein S; DNA-binding component of the Fanconi anemia (FA) core complex. Re
	356	KIF3C	Kinesin-like protein KIF3C; Microtubule-based anterograde translocator for membranous orga
	357	LPCAT1	Lysophosphatidylcholine acyltransferase 1; Possesses both acyltransferase and acetyltransfer
	358	TUBB1	Tubulin beta-1 chain; Tubulin is the major constituent of microtubules. It binds two moles of
	359	ERC1	ELKS/Rab6-interacting/CAST family member 1; Regulatory subunit of the IKK complex. Proba
	360	ANXA11	Annexin A11; Binds specifically to calcyclin in a calcium-dependent manner (By similarity). R
	361	PDLIM5	PDZ and LIM domain protein 5; May play an important role in the heart development by scaff
	362	SLC6A17	Sodium-dependent neutral amino acid transporter SLC6A17; Functions as a sodium-dependen
	363	PMS2	Mismatch repair endonuclease PMS2; Component of the post-replicative DNA mismatch repa
	364	UGT3A2	UDP-glucuronosyltransferase 3A2; UDP-glucuronosyltransferases catalyze phase II biotransfo
	365	HES5	Transcription factor HES-5; Transcriptional repressor of genes that require a bHLH protein for
	366	OR4C46	Olfactory receptor 4C46; Odorant receptor; Olfactory receptors, family 4
	367	LUZP1	Leucine zipper protein 1
	368	FSCB	Fibrous sheath CABYR-binding protein; May be involved in the later stages of fibrous sheath
	369	GTF2F2	General transcription factor IIF subunit 2; TFIIF is a general transcription initiation factor that
	370	MRC2	C-type mannose receptor 2; May play a role as endocytotic lectin receptor displaying calcium
	371	FLYWCH2	FLYWCH family member 2
	372	IBSP	Bone sialoprotein 2; Binds tightly to hydroxyapatite. Appears to form an integral part of the r
	373	INTS3	Integrator complex subunit 3; Component of the Integrator (INT) complex. The Integrator cor
	374	L1CAM	Neural cell adhesion molecule L1; Neural cell adhesion molecule involved in the dynamics of
	375	EPHA3	Ephrin type-A receptor 3; Receptor tyrosine kinase which binds promiscuously membrane-bo
	376	AXL	Tyrosine-protein kinase receptor UFO; Receptor tyrosine kinase that transduces signals from
	377	hsa-mir-3191	microRNA NRAS
	378	HOXB4	Homeobox protein Hox-B4; Sequence-specific transcription factor which is part of a developn
	379	TCTEX1D4	Tctex1 domain containing 4
	380	CLPX	ATP-dependent Clp protease ATP-binding subunit clpX-like, mitochondrial; ATP-dependent sp
	381	GLS	Glutaminase kidney isoform, mitochondrial; Catalyzes the first reaction in the primary pathw
	382	OR52E8	Olfactory receptor 52E8; Odorant receptor; Olfactory receptors, family 52
	383	TNFSF14	Tumor necrosis factor ligand superfamily member 14; Cytokine that binds to TNFRSF3/LTBR.
	384	GPR39	G-protein coupled receptor 39; Zn(2+) acts as an agonist. This receptor mediates its action b
	385	BAIAP2	Brain-specific angiogenesis inhibitor 1-associated protein 2; Adapter protein that links memb
	386	RAC2	Ras-related C3 botulinum toxin substrate 2; Plasma membrane-associated small GTPase wh
	387	DDX53	Probable ATP-dependent RNA helicase DDX53; DEAD-box helicase 53
	388	GRIK2	Glutamate receptor ionotropic, kainate 2; Ionotropic glutamate receptor. L-glutamate acts as
	389	MANF	Mesencephalic astrocyte-derived neurotrophic factor; Selectively promotes the survival of do
	390	GAS2L3	GAS2-like protein 3; Cytoskeletal linker protein. May promote and stabilize the formation of
	391	SNX13	Sorting nexin-13; May be involved in several stages of intracellular trafficking. May play a rol
	392	LOC100130357	
	393	KCNK1	Potassium channel subfamily K member 1; Ion channel that contributes to passive transmem
	394	PRDM13	PR domain zinc finger protein 13; May be involved in transcriptional regulation; PR/SET doma
	395	SH3TC1	SH3 domain and tetratricopeptide repeats 1
	396	ALKBH4	Alpha-ketoglutarate-dependent dioxygenase alkB homolog 4; Dioxygenase that mediates der
	397	DGKK	Diacylglycerol kinase kappa; Phosphorylates diacylglycerol (DAG) to generate phosphatidic ac
	398	TTC19	Tetratricopeptide repeat protein 19, mitochondrial; Required for the preservation of the struc
	399	HELT	Hairy and enhancer of split-related protein HELT; Transcriptional repressor which binds prefe
	400	OR52I2	Olfactory receptor 52I2; Odorant receptor; Olfactory receptors, family 52

	401	FBLN7	Fibulin-7; An adhesion molecule that interacts with extracellular matrix molecules in develop
	402	GAREM	GRB2-associated and regulator of MAPK protein 1; Isoform 1: Acts as an adapter protein that
	403	SNRNP200	U5 small nuclear ribonucleoprotein 200 kDa helicase; RNA helicase that plays an essential ro
	404	FASTKD1	FAST kinase domain-containing protein 1, mitochondrial; FASTK mitochondrial RNA binding f
Enriched (biology)	405	ZRANB3	DNA annealing helicase and endonuclease ZRANB3; DNA annealing helicase and endonuclea
	406	RASSF9	Ras association domain-containing protein 9; May play a role in regulating vesicuar trafficking
	407	DDB2	DNA damage-binding protein 2; Required for DNA repair. Binds to DDB1 to form the UV- dan
	408	DNMT3B	DNA (cytosine-5)-methyltransferase 3B; Required for genome-wide de novo methylation and
	409	CHD1L	Chromodomain-helicase-DNA-binding protein 1-like; DNA helicase which plays a role in chron
	410	SUV420H2	Histone-lysine N-methyltransferase KMT5C; Histone methyltransferase that specifically trime
	411	MPHOSPH8	M-phase phosphoprotein 8; Heterochromatin component that specifically recognizes and bind
	412	HDAC3	Histone deacetylase 3; Responsible for the deacetylation of lysine residues on the N-termina
	413	XPC	DNA repair protein complementing XP-C cells; Involved in global genome nucleotide excision
	414	APOBEC3C	DNA dC->dU-editing enzyme APOBEC-3C; DNA deaminase (cytidine deaminase) which acts a
	415	MBD4	Methyl-CpG-binding domain protein 4; Mismatch-specific DNA N-glycosylase involved in DNA
	416	ZC3H12B	Probable ribonuclease ZC3H12B; May function as RNase and regulate the levels of target RN
	417	ZSCAN10	Zinc finger and SCAN domain-containing protein 10; Embryonic stem (ES) cell-specific transc
	418	ZNF354B	Zinc finger protein 354B; May be involved in transcriptional regulation; Zinc fingers C2H2-type
	419	ZFP14	Zinc finger protein 14 homolog; May be involved in transcriptional regulation; Zinc fingers C2
	420	PRDM5	PR domain zinc finger protein 5; Sequence-specific DNA-binding transcription factor. Repress
	421	TDRD9	ATP-dependent RNA helicase TDRD9; ATP-binding RNA helicase required during spermatoge
	422	DOT1L	Histone-lysine N-methyltransferase, H3 lysine-79 specific; Histone methyltransferase. Methy
	423	EHMT1	Histone-lysine N-methyltransferase EHMT1; Histone methyltransferase that specifically mon
	424	ATF7IP	Activating transcription factor 7-interacting protein 1; Recruiter that couples transcriptional f
	425	PRDM8	PR domain zinc finger protein 8; Probable histone methyltransferase, preferentially acting or
	426	KIAA1328	Protein hinderin; Competes with SMC1 for binding to SMC3. May affect the availability of SM
	427	ZBED9	SCAN domain-containing protein 3; Zinc finger BED-type containing 9; SCAN domain containi
	428	BEND3	BEN domain-containing protein 3; Transcriptional repressor which associates with the NoRC
	429	SLX1B	Structure-specific endonuclease subunit SLX1; Catalytic subunit of the SLX1-SLX4 structure-sp
	430	DTX3L	E3 ubiquitin-protein ligase DTX3L; E3 ubiquitin-protein ligase which, in association with ADP-
	431	TP53	Cellular tumor antigen p53; Acts as a tumor suppressor in many tumor types; induces growth
	432	TP53BP1	TP53-binding protein 1; Double-strand break (DSB) repair protein involved in response to DNA
	433	FAM111A	Protein FAM111A; Chromatin-associated protein required for PCNA loading on replication site
	434	TRIM49C	Tripartite motif containing 49C; Ring finger proteins
	435	TSN	Translin; DNA-binding protein that specifically recognizes consensus sequences at the breakp
	436	HELB	DNA helicase B; 5'-3' DNA helicase involved in DNA damage response by acting as an inhibito
	437	RAD1	Cell cycle checkpoint protein RAD1; Component of the 9-1-1 cell-cycle checkpoint response co
	438	APEX2	DNA-(apurinic or apyrimidinic site) lyase 2; Function as a weak apurinic/apyrimidinic (AP) en
	439	FANCE	Fanconi anemia group E protein; As part of the Fanconi anemia (FA) complex functions in DN
	440	POLL	DNA polymerase lambda; DNA polymerase that functions in several pathways of DNA repair.
	441	NCBP2	Nuclear cap-binding protein subunit 2; Component of the cap-binding complex (CBC), which b
	442	SRSF12	Serine/arginine-rich splicing factor 12; Splicing factor that seems to antagonize SR proteins i
	443	ZBP1	Z-DNA-binding protein 1; Participates in the detection by the host's innate immune system o
	444	MIR3179-3	microRNA RARG
	445	FAM46C	Putative nucleotidyltransferase FAM46C; Probable nucleotidyltransferase that may act as a n
	446	RNF146	E3 ubiquitin-protein ligase RNF146; E3 ubiquitin-protein ligase that specifically binds poly-AD
	447	RNF14	E3 ubiquitin-protein ligase RNF14; Might act as an E3 ubiquitin-protein ligase which accepts
	448	USP20	Ubiquitin carboxyl-terminal hydrolase 20; Deubiquitinating enzyme involved in beta-2 adrene
	449	DCAF11	DDB1- and CUL4-associated factor 11; May function as a substrate receptor for CUL4-DDB1 E
	450	PTPN14	Tyrosine-protein phosphatase non-receptor type 14; Protein tyrosine phosphatase which may

	451	LIMD1	LIM domain-containing protein 1; Adapter or scaffold protein which participates in the assem
	452	TP73	Tumor protein p73; Participates in the apoptotic response to DNA damage. Isoforms containi
	453	AJUBA	LIM domain-containing protein ajuba; Adapter or scaffold protein which participates in the as
	454	SETD7	Histone-lysine N-methyltransferase SETD7; Histone methyltransferase that specifically monc
	455	YWHAQ	14-3-3 protein theta; Adapter protein implicated in the regulation of a large spectrum of both
	456	ARHGAP17	Rho GTPase-activating protein 17; Rho GTPase-activating protein involved in the maintenanc
	457	TAOK3	Serine/threonine-protein kinase TAO3; Serine/threonine-protein kinase that acts as a regulat
	458	ABL1	Tyrosine-protein kinase ABL1; Non-receptor tyrosine-protein kinase that plays a role in many
	459	MOB1A	MOB kinase activator 1A; Activator of LATS1/2 in the Hippo signaling pathway which plays a
	460	SAV1	Protein salvador homolog 1; Regulator of STK3/MST2 and STK4/MST1 in the Hippo signaling
	461	RB1	Retinoblastoma-associated protein; Key regulator of entry into cell division that acts as a tun
	462	RBBP4	Histone-binding protein RBBP4; Core histone-binding subunit that may target chromatin asse
	463	RBBP7	Histone-binding protein RBBP7; Core histone-binding subunit that may target chromatin rem
	464	TET1	Methylcytosine dioxygenase TET1; Dioxygenase that catalyzes the conversion of the modified
	465	H2AFV	Histone H2A.V; Variant histone H2A which replaces conventional H2A in a subset of nucleoso
	466	DPPA2	Developmental pluripotency-associated protein 2; Binds to target gene promoters, including
	467	PRMT7	Protein arginine N-methyltransferase 7; Arginine methyltransferase that can both catalyze th
	468	SUPT20H	SPT20 homolog, SAGA complex component
	469	DMBX1	Diencephalon/mesencephalon homeobox protein 1; Functions as a transcriptional repressor.
	470	EZH1	Histone-lysine N-methyltransferase EZH1; Polycomb group (PcG) protein. Catalytic subunit of
	471	EPC1	Enhancer of polycomb homolog 1; Component of the NuA4 histone acetyltransferase (HAT) c
	472	AEBP2	Zinc finger protein AEBP2; DNA-binding transcriptional repressor. May interact with and stim
	473	CCND2	T-lymphocyte activation antigen CD80; Involved in the costimulatory signal essential for T- ly
	474	ZNF688	Zinc finger protein 688; May be involved in transcriptional regulation; Zinc fingers C2H2-type
	475	SAP130	Histone deacetylase complex subunit SAP130; Acts as a transcriptional repressor. May functi
	476	LBR	Lamin-B receptor; Anchors the lamina and the heterochromatin to the inner nuclear membra
	477	HBP1	HMG box-containing protein 1; Transcriptional repressor that binds to the promoter region of
	478	KRBA2	KRAB-A domain containing 2
	479	ZNF768	Zinc finger protein 768; May be involved in transcriptional regulation; Zinc fingers C2H2-type
	480	ZFP92	Zinc finger protein 92 homolog; May be involved in transcriptional regulation; Zinc fingers C2
	481	ZNF2	Zinc finger protein 2; May be involved in transcriptional regulation; Belongs to the krueppel C
	482	ZNF630	Zinc finger protein 630; May be involved in transcriptional regulation; Zinc fingers C2H2-type
	483	EID1	EP300-interacting inhibitor of differentiation 1; Interacts with RB1 and EP300 and acts as a r
	484	XRCC4	DNA repair protein XRCC4; Involved in DNA non-homologous end joining (NHEJ) required for
	485	TDP1	Tyrosyl-DNA phosphodiesterase 1; DNA repair enzyme that can remove a variety of covalent
	486	APOBEC3H	DNA dC->dU-editing enzyme APOBEC-3H; DNA deaminase (cytidine deaminase) which acts a
	487	PAIP2	Polyadenylate-binding protein-interacting protein 2; Acts as a repressor in the regulation of t
	488	RRP7A	Ribosomal RNA processing 7 homolog A; UTPc subcomplex
	489	IL17RD	Interleukin-17 receptor D; Feedback inhibitor of fibroblast growth factor mediated Ras-MAPK

References

Abu Dawud, R., Schreiber, K., Schomburg, D., and Adjaye, J. (2012). Human embryonic stem cells and embryonal carcinoma cells have overlapping and distinct metabolic signatures. PLoS One *7*, e39896.

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

Athanikar, J.N., Badge, R.M., and Moran, J.V. (2004). A YY1-binding site is required for accurate human LINE-1 transcription initiation. Nucleic Acids Res *32*, 3846-3855.

Barlev, N.A., Emelyanov, A.V., Castagnino, P., Zegerman, P., Bannister, A.J., Sepulveda, M.A., Robert, F., Tora, L., Kouzarides, T., Birshtein, B.K.*, et al.* (2003). A novel human Ada2 homologue functions with Gcn5 or Brg1 to coactivate transcription. Mol Cell Biol *23*, 6944-6957.

Beck, C.R., Collier, P., Macfarlane, C., Malig, M., Kidd, J.M., Eichler, E.E., Badge, R.M., and Moran, J.V. (2010). LINE-1 retrotransposition activity in human genomes. Cell *141*, 1159-1170.

Beck, C.R., Garcia-Perez, J.L., Badge, R.M., and Moran, J.V. (2011). LINE-1 elements in structural variation and disease. Annu Rev Genomics Hum Genet *12*, 187-215.

Blomen, V.A., Majek, P., Jae, L.T., Bigenzahn, J.W., Nieuwenhuis, J., Staring, J., Sacco, R., van Diemen, F.R., Olk, N., Stukalov, A.*, et al.* (2015). Gene essentiality and synthetic lethality in haploid human cells. Science *350*, 1092-1096.

Bollag, G., Hirth, P., Tsai, J., Zhang, J., Ibrahim, P.N., Cho, H., Spevak, W., Zhang, C., Zhang, Y., Habets, G.*, et al.* (2010). Clinical efficacy of a RAF inhibitor needs broad target blockade in BRAF-mutant melanoma. Nature *467*, 596-599.

Brouha, B., Meischl, C., Ostertag, E., de Boer, M., Zhang, Y., Neijens, H., Roos, D., and Kazazian, H.H., Jr. (2002). Evidence consistent with human L1 retrotransposition in maternal meiosis I. Am J Hum Genet *71*, 327-336.

Brouha, B., Schustak, J., Badge, R.M., Lutz-Prigge, S., Farley, A.H., Moran, J.V., and Kazazian, H.H., Jr. (2003). Hot L1s account for the bulk of retrotransposition in the human population. Proc Natl Acad Sci U S A *100*, 5280-5285.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A.*, et al.* (2013). Multiplex genome engineering using CRISPR/Cas systems. Science *339*, 819-823.

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

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

Cross, B.C., Lawo, S., Archer, C.R., Hunt, J.R., Yarker, J.L., Riccombeni, A., Little, A.S., McCarthy, N.J., and Moore, J.D. (2016). Increasing the performance of pooled CRISPR-Cas9 drop-out screening. Sci Rep *6*, 31782.

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

Doench, J.G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E.W., Donovan, K.F., Smith, I., Tothova, Z., Wilen, C., Orchard, R.*, et al.* (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nat Biotechnol *34*, 184-191.

Dombroski, B.A., Feng, Q., Mathias, S.L., Sassaman, D.M., Scott, A.F., Kazazian, H.H., Jr., and Boeke, J.D. (1994). An in vivo assay for the reverse transcriptase of human retrotransposon L1 in Saccharomyces cerevisiae. Mol Cell Biol *14*, 4485-4492.

Dombroski, B.A., Mathias, S.L., Nanthakumar, E., Scott, A.F., and Kazazian, H.H., Jr. (1991). Isolation of an active human transposable element. Science *254*, 1805-1808.

Doucet, A.J., Wilusz, J.E., Miyoshi, T., Liu, Y., and Moran, J.V. (2015). A 3' Poly(A) Tract Is Required for LINE-1 Retrotransposition. Mol Cell *60*, 728-741.

Faulkner, G.J., and Billon, V. (2018). L1 retrotransposition in the soma: a field jumping ahead. Mob DNA *9*, 22.

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

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

Feschotte, C., and Pritham, E.J. (2007). DNA transposons and the evolution of eukaryotic genomes. Annu Rev Genet *41*, 331-368.

Flasch, D.A., Macia, A., Sanchez, L., Ljungman, M., Heras, S.R., Garcia-Perez, J.L., Wilson, T.E., and Moran, J.V. (2019). Genome-wide de novo L1 Retrotransposition Connects Endonuclease Activity with Replication. Cell *177*, 837-851 e828.

Garcia-Perez, J.L., Doucet, A.J., Bucheton, A., Moran, J.V., and Gilbert, N. (2007a). Distinct mechanisms for trans-mediated mobilization of cellular RNAs by the LINE-1 reverse transcriptase. Genome Res *17*, 602-611.

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

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

Gilbert, N., Lutz, S., Morrish, T.A., and Moran, J.V. (2005). Multiple fates of L1 retrotransposition intermediates in cultured human cells. Mol Cell Biol *25*, 7780-7795.

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

Goodier, J.L. (2016). Restricting retrotransposons: a review. Mob DNA *7*, 16.

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

Grimaldi, G., Skowronski, J., and Singer, M.F. (1984). Defining the beginning and end of KpnI family segments. EMBO J *3*, 1753-1759.

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

Hart, T., Brown, K.R., Sircoulomb, F., Rottapel, R., and Moffat, J. (2014). Measuring error rates in genomic perturbation screens: gold standards for human functional genomics. Mol Syst Biol *10*, 733.

Hart, T., Chandrashekhar, M., Aregger, M., Steinhart, Z., Brown, K.R., MacLeod, G., Mis, M., Zimmermann, M., Fradet-Turcotte, A., Sun, S.*, et al.* (2015). High-Resolution CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities. Cell *163*, 1515-1526.

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

Hsu, P.D., Lander, E.S., and Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. Cell *157*, 1262-1278.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science *337*, 816-821.

Jurka, J. (1997). Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons. Proc Natl Acad Sci U S A *94*, 1872-1877.

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

Kazazian, H.H., Jr. (2004). Mobile elements: drivers of genome evolution. Science *303*, 1626-1632.

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

Kazazian, H.H., Jr., Wong, C., Youssoufian, H., Scott, A.F., Phillips, D.G., and Antonarakis, S.E. (1988). Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. Nature *332*, 164-166.

Kopera, H.C., Larson, P.A., Moldovan, J.B., Richardson, S.R., Liu, Y., and Moran, J.V. (2016). LINE-1 Cultured Cell Retrotransposition Assay. Methods Mol Biol *1400*, 139-156.

Kopera, H.C., Moldovan, J.B., Morrish, T.A., Garcia-Perez, J.L., and Moran, J.V. (2011). Similarities between long interspersed element-1 (LINE-1) reverse transcriptase and telomerase. Proc Natl Acad Sci U S A *108*, 20345-20350.

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

Kulpa, D.A., and Moran, J.V. (2005). Ribonucleoprotein particle formation is necessary but not sufficient for LINE-1 retrotransposition. Hum Mol Genet *14*, 3237-3248.

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

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W.*, et al.* (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860-921.

Leibold, D.M., Swergold, G.D., Singer, M.F., Thayer, R.E., Dombroski, B.A., and Fanning, T.G. (1990). Translation of LINE-1 DNA elements in vitro and in human cells. Proc Natl Acad Sci U S A *87*, 6990-6994.

Levin, H.L., and Moran, J.V. (2011). Dynamic interactions between transposable elements and their hosts. Nat Rev Genet *12*, 615-627.

Lewinski, M.K., and Bushman, F.D. (2005). Retroviral DNA integration--mechanism and consequences. Adv Genet *55*, 147-181.

Li, W., Xu, H., Xiao, T., Cong, L., Love, M.I., Zhang, F., Irizarry, R.A., Liu, J.S., Brown, M., and Liu, X.S. (2014). MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. Genome Biol *15*, 554.

Liu, X., Vorontchikhina, M., Wang, Y.L., Faiola, F., and Martinez, E. (2008). STAGA recruits Mediator to the MYC oncoprotein to stimulate transcription and cell proliferation. Mol Cell Biol *28*, 108-121.

Luan, D.D., Korman, M.H., Jakubczak, J.L., and Eickbush, T.H. (1993). Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. Cell *72*, 595-605.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013). RNA-guided human genome engineering via Cas9. Science *339*, 823-826.

Martin, S.L., Cruceanu, M., Branciforte, D., Wai-Lun Li, P., Kwok, S.C., Hodges, R.S., and Williams, M.C. (2005). LINE-1 retrotransposition requires the nucleic acid chaperone activity of the ORF1 protein. J Mol Biol *348*, 549-561.

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

Moldovan, J.B., Wang, Y., Shuman, S., Mills, R.E., and Moran, J.V. (2019). RNA ligation precedes the retrotransposition of U6/LINE-1 chimeric RNA. Proc Natl Acad Sci U S A *116*, 20612-20622.

Monot, C., Kuciak, M., Viollet, S., Mir, A.A., Gabus, C., Darlix, J.L., and Cristofari, G. (2013). The specificity and flexibility of l1 reverse transcription priming at imperfect T-tracts. PLoS Genet *9*, e1003499.

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

Morrish, T.A., Gilbert, N., Myers, J.S., Vincent, B.J., Stamato, T.D., Taccioli, G.E., Batzer, M.A., and Moran, J.V. (2002). DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. Nat Genet *31*, 159-165.

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

Nazarian, R., Shi, H., Wang, Q., Kong, X., Koya, R.C., Lee, H., Chen, Z., Lee, M.K., Attar, N., Sazegar, H.*, et al.* (2010). Melanomas acquire resistance to B-RAF(V600E) inhibition by RTK or N-RAS upregulation. Nature *468*, 973-977.

Olovnikov, I.A., Ad'ianova, Z.V., Galimov, E.P., Andreev, D.E., Terenin, I.M., Ivanov, D.S., Prasolov, V.S., and Dmitriev, S.E. (2007). [A key role of the internal region of the 5'-untranslated region in the human L1 retrotransposon transcription activity]. Mol Biol (Mosk) *41*, 508-514.

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

Ostertag, E.M., Prak, E.T., DeBerardinis, R.J., Moran, J.V., and Kazazian, H.H., Jr. (2000). Determination of L1 retrotransposition kinetics in cultured cells. Nucleic Acids Res *28*, 1418-1423.

Piskareva, O., Denmukhametova, S., and Schmatchenko, V. (2003). Functional reverse transcriptase encoded by the human LINE-1 from baculovirus-infected insect cells. Protein Expr Purif *28*, 125-130.

Piskareva, O., and Schmatchenko, V. (2006). DNA polymerization by the reverse transcriptase of the human L1 retrotransposon on its own template in vitro. FEBS Lett *580*, 661-668.

Pyzocha, N.K., Ran, F.A., Hsu, P.D., and Zhang, F. (2014). RNA-guided genome editing of mammalian cells. Methods Mol Biol *1114*, 269-277.

Ran, F.A., Hsu, P.D., Lin, C.Y., Gootenberg, J.S., Konermann, S., Trevino, A.E., Scott, D.A., Inoue, A., Matoba, S., Zhang, Y.*, et al.* (2013a). Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. Cell *154*, 1380-1389.

Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013b). Genome engineering using the CRISPR-Cas9 system. Nat Protoc *8*, 2281-2308.

Richardson, S.R., Doucet, A.J., Kopera, H.C., Moldovan, J.B., Garcia-Perez, J.L., and Moran, J.V. (2015). The Influence of LINE-1 and SINE Retrotransposons on Mammalian Genomes. Microbiol Spectr *3*, MDNA3-0061-2014.

Richardson, S.R., Gerdes, P., Gerhardt, D.J., Sanchez-Luque, F.J., Bodea, G.O., Munoz-Lopez, M., Jesuadian, J.S., Kempen, M.H.C., Carreira, P.E., Jeddeloh, J.A.*, et*

al. (2017). Heritable L1 retrotransposition in the mouse primordial germline and early embryo. Genome Res *27*, 1395-1405.

Romano, E., Pradervand, S., Paillusson, A., Weber, J., Harshman, K., Muehlethaler, K., Speiser, D., Peters, S., Rimoldi, D., and Michielin, O. (2013). Identification of multiple mechanisms of resistance to vemurafenib in a patient with BRAFV600E-mutated cutaneous melanoma successfully rechallenged after progression. Clin Cancer Res *19*, 5749-5757.

Sanjana, N.E., Shalem, O., and Zhang, F. (2014). Improved vectors and genome-wide libraries for CRISPR screening. Nat Methods *11*, 783-784.

Sarraf, S., Tejada, R., Abawi, M., Oberst, M., Dennis, T., Simon, K.C., and Blancato, J. (2005). The human ovarian teratocarcinoma cell line PA-1 demonstrates a single translocation: analysis with fluorescence in situ hybridization, spectral karyotyping, and bacterial artificial chromosome microarray. Cancer Genet Cytogenet *161*, 63-69.

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

Schorn, A.J., Gutbrod, M.J., LeBlanc, C., and Martienssen, R. (2017). LTR-Retrotransposon Control by tRNA-Derived Small RNAs. Cell *170*, 61-71 e11.

Scott, A.F., Schmeckpeper, B.J., Abdelrazik, M., Comey, C.T., O'Hara, B., Rossiter, J.P., Cooley, T., Heath, P., Smith, K.D., and Margolet, L. (1987). Origin of the human L1 elements: proposed progenitor genes deduced from a consensus DNA sequence. Genomics *1*, 113-125.

Scott, E.C., and Devine, S.E. (2017). The Role of Somatic L1 Retrotransposition in Human Cancers. Viruses *9*.

Shalem, O., Sanjana, N.E., Hartenian, E., Shi, X., Scott, D.A., Mikkelson, T., Heckl, D., Ebert, B.L., Root, D.E., Doench, J.G.*, et al.* (2014). Genome-scale CRISPR-Cas9 knockout screening in human cells. Science *343*, 84-87.

Solyom, S., Ewing, A.D., Rahrmann, E.P., Doucet, T., Nelson, H.H., Burns, M.B., Harris, R.S., Sigmon, D.F., Casella, A., Erlanger, B.*, et al.* (2012). Extensive somatic L1 retrotransposition in colorectal tumors. Genome Res *22*, 2328-2338.

Sperger, J.M., Chen, X., Draper, J.S., Antosiewicz, J.E., Chon, C.H., Jones, S.B., Brooks, J.D., Andrews, P.W., Brown, P.O., and Thomson, J.A. (2003). Gene expression patterns in human embryonic stem cells and human pluripotent germ cell tumors. Proc Natl Acad Sci U S A *100*, 13350-13355.

Swergold, G.D. (1990). Identification, characterization, and cell specificity of a human LINE-1 promoter. Mol Cell Biol *10*, 6718-6729.

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

Szklarczyk, D., Gable, A.L., Nastou, K.C., Lyon, D., Kirsch, R., Pyysalo, S., Doncheva, N.T., Legeay, M., Fang, T., Bork, P.*, et al.* (2021). The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. Nucleic Acids Res *49*, D605-D612.

Tainsky, M.A., Cooper, C.S., Giovanella, B.C., and Vande Woude, G.F. (1984). An activated rasN gene: detected in late but not early passage human PA1 teratocarcinoma cells. Science *225*, 643-645.

Weber, M.J. (2006). Mammalian small nucleolar RNAs are mobile genetic elements. PLoS Genet *2*, e205.

Wei, W., Gilbert, N., Ooi, S.L., Lawler, J.F., Ostertag, E.M., Kazazian, H.H., Boeke, J.D., and Moran, J.V. (2001). Human L1 retrotransposition: cis preference versus trans complementation. Mol Cell Biol *21*, 1429-1439.

Wei, W., Morrish, T.A., Alisch, R.S., and Moran, J.V. (2000). A transient assay reveals that cultured human cells can accommodate multiple LINE-1 retrotransposition events. Anal Biochem *284*, 435-438.

Zeuthen, J., Norgaard, J.O., Avner, P., Fellous, M., Wartiovaara, J., Vaheri, A., Rosen, A., and Giovanella, B.C. (1980). Characterization of a human ovarian teratocarcinoma-derived cell line. Int J Cancer *25*, 19-32.

Zhang, A., Dong, B., Doucet, A.J., Moldovan, J.B., Moran, J.V., and Silverman, R.H. (2014). RNase L restricts the mobility of engineered retrotransposons in cultured human cells. Nucleic Acids Res *42*, 3803-3820.

Chapter 3

NF2/merlin is Necessary for Efficient Silencing of L1 Retrotransposition in Human PA-1 EC Cells

This chapter is the product of a continued collaboration with Dr. Peter Larson and Dr. Jacob Kitzman. Illumina sequencing was performed in collaboration with the Kitzman laboratory. I performed all experiments and analyses discussed in this chapter.

Abstract

An average human genome contains approximately 100 Long Interspersed Element-1 (LINE-1 or L1) sequences capable of retrotransposition. Human embryonic carcinoma-derived cell lines (hECs) initiate and maintain epigenetic silencing of reporter genes delivered by L1 retrotransposition, a process termed L1-REPEL (L1-delivered REPorter gEne siLencing). In Chapter 2, we performed a genome-wide screen that identified *NF2* as our top candidate gene necessary for efficient L1-REPEL in PA-1 hECs. Here, we demonstrate that population-based and clonal CRISPR/Cas9-mediated knockout of *NF2* attenuates L1-REPEL, suggesting that the NF2/merlin protein is necessary for efficient L1-REPEL in PA-1 cells. Expression of the predominant NF2/merlin isoform 1 cDNA efficiently restored L1-REPEL in *NF2* knockout cells.

However, *NF2* knockout is not sufficient to reactivate a previously silenced L1-delivered reporter gene in a clonal PA-1 cell line containing a single, silenced L1 insertion (pk5 cells), suggesting that NF2/merlin may be necessary to initiate L1-REPEL in PA-1 cells. Finally, we found that culturing *NF2* knockout cells in differentiation media further suppressed L1-REPEL, suggesting that *NF2* knockout and culturing cells in differentiation medium may act independently or in combination to attenuate L1-REPEL. Our data indicate that NF2/merlin is necessary to establish efficient L1 REPEL in PA-1 hECs, suggesting that a tumor suppressor gene implicated in human disease may also play a role in silencing L1 retrotransposition events during early human development.

Introduction

Long Interspersed Element-1 (LINE-1 or L1) is an endogenous non-LTR retrotransposon that has proliferated throughout mammalian evolution (Lander et al., 2001; Smit, 1996). Retrotransposition-competent human L1s (RC-L1s) are ~6 kb in length and encode two proteins (ORF1p and ORF2p) necessary for autonomous retrotransposition (Ergun et al., 2004; Holmes et al., 1992; Kulpa and Moran, 2005, 2006; Leibold et al., 1990; Martin et al., 2005; Moran et al., 1996; Wei et al., 2001). The average human genome contains ~100 retrotransposition-competent L1s that can retrotranspose by target-primed reverse transcription (TPRT) (Beck et al., 2010; Brouha et al., 2003; Cost et al., 2002; Feng et al., 1996; Kulpa and Moran, 2006; Luan et al., 1993). Notably, TPRT is unique to non-LTR retrotransposons and differs from integration mechanisms used by LTR retrotransposons, DNA transposons, and retroviruses (Beck et al., 2011; Feschotte and Pritham, 2007; Kazazian and Moran, 2017; Lewinski and Bushman, 2005; Schorn et al., 2017).

L1 can retrotranspose in the germline, during early development, and in select somatic cells (Coufal et al., 2009; Faulkner and Billon, 2018; Faulkner and Garcia-Perez, 2017; Garcia-Perez et al., 2007; Garcia-Perez et al., 2010; Kano et al., 2009; Kazazian, 2004; Kubo et al., 2006; Muotri et al., 2005; Richardson et al., 2017). The resultant L1-mediated retrotransposition events can alter gene expression, generate structural variation either upon or after integration, and, on occasion, can create pathogenic mutations (Beck et al., 2011; Kazazian and Moran, 2017; Kazazian et al., 1988; Richardson et al., 2015; Scott and Devine, 2017; Solyom et al., 2012).

Previous studies established that human embryonic carcinoma-derived cell lines (hECs) differ from many somatic cancer cell lines in their ability to epigenetically silence reporter genes delivered by L1 retrotransposition (Garcia-Perez et al., 2010). PA-1 cells are a human ovarian teratocarcinoma-derived cell line containing a single reciprocal translocation between chromosomes 15 and 20 (Sarraf et al., 2005; Zeuthen et al., 1980). PA-1 cells, like many hEC cell lines, exhibit early developmental gene expression profiles (Abu Dawud et al., 2012; Sperger et al., 2003) and efficiently express a cohort of endogenous L1s (Garcia-Perez et al., 2007; Garcia-Perez et al., 2010; Hohjoh and Singer, 1996; Skowronski and Singer, 1985). Despite high levels of endogenous L1 expression, engineered L1s are efficiently and stably silenced upon retrotransposition in hEC cells by a process that we have termed L1-REPEL (L1-delivered REPorter gEne siLencing) (Garcia-Perez et al., 2010). Importantly, the following data suggest that L1-REPEL appears to involve both initiation and maintenance phases: (1) L1-REPEL results in the stable mitotic silencing of reporter gene expression; (2) treating cells with the histone deacetylase inhibitor TSA reverses

L1-REPEL; (3) L1-REPEL status correlates with histone modifications at the L1 integration site; and (4) the removal of TSA results in re-establishment of L1-REPEL, suggesting an epigenetic memory to the system.

PA-1 cells preferentially differentiate into an ectodermal-like lineage and, in a way, can be considered as immortalized neural progenitor cells (Garcia-Perez et al., 2010). Culturing PA-1 cells in differentiation media (DM) during the L1 retrotransposition assay led to a significant decrease in L1-REPEL efficiency (*i.e.*, ~30-fold increase in EGFP expressing cells) compared to PA-1 cells cultured in FBS-containing media (Garcia-Perez et al., 2010). The subsequent treatment of cells grown in DM with TSA led to an additional ~3-fold increase in EGFP expression, as opposed to the ~20-30-fold increase observed in PA-1 cells grown in FBS-containing media (Garcia-Perez et al., 2010). By comparison, culturing pk5 cells in DM media was not sufficient to reactivate a previously silenced L1-REPEL event (Garcia-Perez et al., 2010). Together, these results led to the hypothesis that host factors required for the initiation of L1-REPEL are efficiently expressed in PA-1 hECs, then undergo downregulation during cellular differentiation.

In Chapter 2, we implemented a genome-wide CRISPR/Cas9-based knockout screen to identify cellular factors necessary for L1-REPEL in PA-1 cells and identified *NF2* as our top candidate gene. Neurofibromin 2 (*NF2*) is a known tumor suppressor gene that encodes the NF2/merlin (moesin-ezrin-radixin-like) protein (Rouleau et al., 1993; Trofatter et al., 1993). Loss-of-function mutations in human *NF2* lead to neurofibromatosis type 2 (NF2), which is a benign tumor forming disease of the nervous system (Petrilli and Fernandez-Valle, 2016). Studies using animal models demonstrated that rat *Nf2* is widely expressed during embryogenesis, but is restricted to nervous

tissue and the testis in adult animals (Gutmann et al., 1995). Subsequent studies revealed that mouse embryos containing homozygous *Nf2* mutations lacked an organized extraembryonic ectoderm, resulting in embryonic lethality at approximately day E7 of gestation (McClatchey et al., 1997).

NF2 comprises 17 exons; NF2/merlin isoform 1 encodes the predominant 595 amino acid isoform of NF2/merlin, which lacks exon 16 due to exon skipping (Bianchi et al., 1994; Golovnina et al., 2005; Gutmann et al., 1995). The amino-terminus of NF2/merlin encodes a conserved FERM (4.1, ezrin, radixin, moesin) domain, which is followed by a helical domain and a carboxyl-terminal domain (Figure 3.1A).

NF2/merlin can form homodimers as well as heterodimers with other ERM proteins (Gronholm et al., 1999; Pearson et al., 2000; Shimizu et al., 2002; Stokowski and Cox, 2000). Merlin homodimers exhibit a head-to-tail intramolecular arrangement that is critical for its tumor suppressor activity (Gronholm et al., 1999; Nguyen et al., 2001; Xu and Gutmann, 1998). NF2/merlin has various activities, which upon its loss, are thought to contribute to tumor suppressor activity. NF2/merlin is thought to function as: (1) a membrane-cytoskeleton scaffold protein that regulates receptor mediated signaling; (2) an important factor implicated in embryonic developmental gene expression pathways *(e.g.*, the Hippo, WNT/β-catenin, TGF-β, receptor tyrosine kinase, Notch, and Hedgehog pathways (Perrimon et al., 2012; Stamenkovic and Yu, 2010); and (3) a factor necessary for contact-dependent inhibition of cell proliferation (Lallemand et al., 2003; Morrison et al., 2001; Okada et al., 2005; Rouleau et al., 1993; Shaw et al., 1998; Trofatter et al., 1993; Xiao et al., 2005).

Post-translational modifications also are implicated in regulating NF2/merlin conformation and function. For example, phosphorylation of NF2/merlin at Ser10, Thr230, and/or Ser315 by protein kinase B (PKB or Akt) promotes its proteasomal degradation (Laulajainen et al., 2011; Tang et al., 2007). Phosphorylation of Ser10 by protein kinase A (PKA) can influence actin cytoskeleton dynamics (Laulajainen et al., 2008). Phosphorylation at Ser518 by the p21-activated kinase (PAK) is reported to hinder NF2/merlin intramolecular complex formation and promotes an "open" conformation of the protein, which allows NF2/merlin to function as a cytoplasmic scaffold protein (Alfthan et al., 2004; Rong et al., 2004; Shaw et al., 2001; Surace et al., 2004; Xiao et al., 2002). Dephosphorylation of Ser518 by myosin phosphatase (MYPT1) is thought to promote the formation of a "closed" conformation of NF2/merlin that allows it to localize to the nucleus in confluent and growth factor-deprived cells, which is necessary for NF2/merlin tumor suppressor activity (Jin et al., 2006). Intriguingly, the dephosphorylated Ser518 form of NF2/merlin can bind DDB1 and CUL4 associated factor 1 (DCAF1), also known as vpr binding protein (VPRBP), and is thought to inhibit $CRL4^{DCAF1}$ E3 ubiquitin ligase activity (Li et al., 2010).

Here, we investigate the role of NF2/merlin in L1-REPEL. We demonstrate that *NF2* knockout attenuates L1-REPEL in PA-1 cells and that the subsequent expression of NF2/merlin isoform 1 can efficiently restore L1-REPEL in *NF2* knockout PA-1 cells. Furthermore, we demonstrate that knockout of *NF2* is not sufficient to reactivate L1-REPEL in a clonal PA-1 derivative cell line (pk5), which contains a stably silenced L1-delivered *EGFP* gene. Finally, we report that L1-REPEL is less efficient in PA-1 cells undergoing active differentiation in culture. Together, this work suggests that NF2/merlin

is necessary to establish efficient L1 REPEL in PA-1 hECs, revealing a potential role for NF2/merlin in silencing *de novo* engineered L1 retrotransposition events in PA-1 cells.

Results

"Population knockout" of *NF2* decreases L1-REPEL in PA-1 cells

To determine whether *NF2* was necessary for efficient silencing of L1-delivered reporter genes, we transfected PA-1 cells with PX459-derived vectors (Ran et al., 2013a; Ran et al., 2013b) containing sgRNAs targeting *NF2*. The *NF2* encoded protein, merlin, contains an amino-terminal FERM (4.1, ezrin, radixin and moesin) domain, which is followed by a helical domain and a carboxyl-terminal domain (Figure 3.1 A).

To generate *NF2* knockouts, we used three different sgRNAs that target exons encoding the FERM domain (Figure 3.1 A). *NF2*_sgRNA_1 and NF2_sgRNA_2 target the 3' end of exon 4, whereas *NF2*_sgRNA_3 targets exon 8 (Figure 3.1 A). As a control, we also transfected PA-1 cells with an empty PX459 vector. Transfected PA-1 cells were selected with puromycin to isolate a population of cells containing potential *NF2* mutations. This population of knockout cells then was subjected to the L1-*neo* retrotransposition assay using a wild-type (WT) L1 expression vector (Figure 3.1B; pJM101/L1.3; see Chapter 2 and Methods for details).

We transiently transfected pJM101/L1.3 into separate populations of PA-1 cells that either express the sgRNAs mentioned above or an empty PX459 vector (Figure 3.1 B) (Wei et al., 2001). As controls, we also transfected each population of putative *NF2* knockout cells with pCDNA3, an expression vector that constitutively expresses the neomycin phosphotransferase resistance gene, or pJM105/L1.3, a retrotransposition-

defective derivative of pJM101/L1.3 that contains and inactivating missense mutation in the ORF2p reverse-transcriptase domain (D702A).

Consistent with the data presented at the end of Chapter 2, puromycin-resistant PA-1 cell populations expressing the PX459 empty vector, and then transfected pJM101/L1.3, generally had fewer than 1 G418-resistant focus per well of a six well tissue culture plate (Figure 3.1 C). By comparison, individual puromycin-resistant PA-1 cell populations expressing one of the three different sgRNAs targeting *NF2*, and then transfected pJM101/L1.3, generally contained greater than 20 G418-resistant foci per well. As controls, we demonstrated that transfection of the above cell populations with pCDNA3 yielded similar numbers of G418-resistant foci, whereas cell populations transfected with pJM105/L1.3 did not result in G418-resistant foci. Thus, sgRNAs that target the FERM domain of *NF2* allow a subset of PA-1 cells to escape L1-REPEL (Figure 3.1 C).

As additional controls, we also performed the above assays in other immortalized non-hEC derived cell lines that do not appear to silence engineered L1 retrotransposition events by L1-REPEL (Figure 3.2 A). Briefly, we transfected either the *NF2*_sgRNA_1, NF2_sgRNA_2, or PX459 vectors into HeLa-JVM (Figure 3.2 B) and HAP1 (Figure 3.2 C) cells (near-haploid, leukemia-derived human cells) and then conducted the pJM101/L1.3 retrotransposition assay as described above. We observed a ~3-fold decrease in G418-resistant foci in cells transfected with the *NF2* sgRNAs in both HeLa-JVM and HAP1 cells transfected with pJM101/L1.3 when compared to the PX459 control population of cells. Thus, as opposed to PA-1 cells, population knockout of *NF2* in HeLa-JVM or HAP1 cells lead to slightly reduced L1 retrotransposition levels

when compared to PX459 cells, suggesting that *NF2*/merlin may differentially influence L1-REPEL and/or L1 retrotransposition efficiency in human embryonic cells compared to other immortalized somatic cell types

Generation of clonal PA-1 *NF2* knockout cell lines

To investigate the role of *NF2*/merlin in L1-REPEL we aimed to generate clonal PA-1 cell lines lacking functional *NF2/*merlin expression. To generate clonal *NF2* knockout cells, we first generated a population of puromycin-resistant PA-1 *NF2* knockout cells, and then performed dilution cloning (Figure 3.3 A) to isolate candidate single cells that may contain a knockout of the *NF2* gene. We then tested whether the resultant single cell-derived clonal cell lines express NF2/merlin. Briefly, as described in the above section, PA-1 cells were transfected with the PX459-derived vectors expressing either the *NF2*_sgRNA_1 or the NF2_sgRNA_2 and were subjected to puromycin selection to produce a population of *NF2* sgRNA-expressing cells. The resultant cell lines then were collected, counted, and plated at dilutions of 0.5-, 1-, 2- and 5-cells per well in 96-well cloning plates. Light microscopy was used to identify individual wells containing a single cell and the resultant cells were monitored throughout clonal outgrowth. Upon reaching confluency, the clonal cell lines were trypsinized and serially transferred to increasingly larger tissue culture plates (*i.e.*, 24-well, 6-well, and then 10cm). The cells from the 10cm tissue culture plates were trypsinized and aliquots were either cryo-frozen to create clonal cell stocks or subjected to further analyses to characterize *NF2*/merlin expression. We generated 15 clones using *NF2*_sgRNA_1 (N19) and 12 clones using NF2_sgRNA_2 (N60).

To identify clonal cell lines that lacked functional *NF2*/merlin expression, we screened clones for stable expression of the *NF2*-encoded protein, merlin. Briefly, whole cell lysates (WCLs) from the candidate clonal cell lines were collected and subjected to western blot analysis to assay for NF2/merlin protein expression. We used three different antibodies to detect the expression of the endogenous human NF2/merlin protein: (1) a NF2 (B-12) mouse monoclonal antibody raised against amino acids 336-595 (Santa Cruz, sc-55575); (2) a NF2 rabbit polyclonal antibody raised against amino acids 465-590 (Sigma-Aldrich, HPA003097); and (3) a NF2 rabbit polyclonal antibody raised against amino acids 65-95 (Invitrogen, PA535316). We repeatedly detected a robust 70 kD band in wild-type PA-1 cells using the C-terminal Santa Cruz antibody (Figure 3.3 B). Similar results were obtained using the Sigma-Aldrich C-terminal and Invitrogen N-terminal antibodies (not shown). Clones N19_7, N19_11, N19_13, N60_4, N60_9, N60_10, N60_11 and N60_12 lacked the 70 kD NF2/merlin band, suggesting knockout of *NF2*. We detected NF2/merlin in clones N60_3 and N60_8 at levels similar to wild-type PA-1 cells (Figure 3.3 B). Thus, some of the single-cell derived clonal PA-1 cell lines appear to lack full-length stable NF2/merlin expression.

To further characterize the clonal *NF2* knockout cell lines, we examined the predicted CRISPR/Cas-9 sgRNA-targeting sites to determine whether a mutation(s) had occurred at the *NF2* locus. Briefly, genomic DNA was collected from the candidate clonal PA-1 cell lines and PCR was conducted using two primer sets (A and B) that flank the predicted sgRNA-target site within *NF2* (Figure 3.3 C). Because both sgRNAs (N19 and N60) target the 3' end of *NF2* exon 4, we were able to amplify either of the sgRNA-targets sites using the A and B primer sets.

PCR amplification using primer set B yielded the predicted ~312bp product in clones N60_4, N60_9, N60_12, N19_11 and N19_7, as well as products of greater than 312bp for clones N60_4 and N60_9, suggesting the presence of an insertion mutation at the sgRNA-targeting site in clones N60_4 and N60_9 (Figure 3.3 C). Cloning and sequencing the PCR products revealed that clone N60_4 contained a 254 bp insertion mapping to *NF2* exon 3, whereas clone N60_9 contained a 60 bp insertion mapping to *Cas9* sequence within the PX459 vector (Figure 3.3 D). Each clone lacking stable *NF2*/merlin protein expression contained genomic edits in the vicinity of the sgRNA-target site in both alleles of *NF2* (Figure 3.3 D); some edits resulted in frameshift mutations, whereas others resulted in an in-frame insertion or deletion (Figure 3.3 D). For example, examination of the amino acid conservation at the sgRNA-target site using the UCSC genome browser revealed that the AVQ amino acids that are deleted in clone N19_7 are conserved throughout Zebrafish, which spans ~450 million years of evolution (Figure 3.3 E). Further analysis revealed that every recovered insertion or deletion mutation disrupted the conserved AVQ site. Thus, these results suggest that the AVQ residues encoded at the 3' end of exon 4 are necessary for the structural integrity of the FERM domain and NF2/merlin protein stability. In sum, we identified 12 *NF2* knockout clones lacking functional NF2/merlin expression.

The NF2 knockout clones N19_7 and N60_4 do not exhibit increased proliferation

Previous reports have demonstrated that loss of the *NF2* tumor suppressor gene can result in increased cell proliferation (Lallemand et al., 2003; Morrison et al., 2001; Okada et al., 2005; Rouleau et al., 1993; Trofatter et al., 1993; Xiao et al., 2005). Thus, we performed a cell growth assay to determine if our *NF2* knockout clones exhibited a

proliferative advantage when compared to PA-1 cells. We plated wild-type PA-1 cells along with select *NF2* knockout clones at $5x10^4$ cells per well in a 6-well plate. Cells then were microscopically imaged (Figure 3.4 A) and counted (Figure 3.4 B) daily. As expected, wild-type PA-1 cells and clone N60_3 cells, which expresses the NF2/merlin protein, grew at similar rates during the course of our assay (Figure 3.4 B). By comparison, *NF2* knockout clones N19_7 and N60_4 proliferated at slower rates when compared to wild-type PA-1 cells (Figure 3.4 B), suggesting that *NF2* knockout does confer a proliferative advantage for the N19_7 and N60_4 clonal cell lines.

NF2 is necessary for efficient silencing of L1-delivered reporter genes

We next performed the L1 retrotransposition assay in clonal *NF2* knockout cells to assess L1-REPEL efficiency. We transiently transfected pJM101/L1.3 into clonal *NF2* knockout cells and performed the L1-*neo* retrotransposition assay (Figure 3.5 A). We also transfected clonal *NF2* knockout cells with pCDNA3, to control for transfection efficiency, and the negative control pJM105/L1.3 (RT-defective) L1 expression vector. We observed robust levels of G418-resistant foci in the clonal *NF2* knockout cells that exhibited a loss of NF2/merlin protein expression (Figures 3.3 and 3.5), indicating that the loss of NF2/merlin expression attenuates L1-REPEL. By comparison, wild-type PA-1 cells and clone N60_3, which express the NF2/merlin protein (Figure 3.3), did not show any evidence of G418-resistant foci. As expected, controls revealed the absence of G418-resistant foci in cells transfected with pJM105/L1.3 (Figure 3.5 A). Notably, all cell lines exhibited G418-resistant foci, albeit at various efficiencies (which likely is due to differences in transfection efficiencies), in cells transfected with pCDNA3.

To further examine the efficiency of L1-REPEL in clonal *NF2* knockout cells we next used the L1-*GFP* retrotransposition assay. Unlike the L1-*neo* assay, the L1-*GFP* assay allows us to reactivate silenced L1-delivered *EGFP* reporter genes to measure the L1-REPEL efficiency. Previous studies have demonstrated that treating cells with trichostatin A (TSA) a class-one pan histone deacetylase inhibitor efficiently reactivated silenced L1-delivered *EGFP* reporter genes (Garcia-Perez et al., 2010). Briefly, we transiently transfected p99EGFP/LRE3 (which contains a puromycin resistance gene on the vector backbone) into clonal *NF2* knockout cell lines, selected for puromycin resistant (2 μg/mL) cells 24 hours post-transfection, incubated the cells for five more days, and then treated the cell in either the presence of TSA (500 nM) or vehicle (DMSO) for 18 hours. The TSA-treated and untreated cells then were collected and subjected to flow cytometry to identify EGFP expressing (+) cells (Figure 3.5 B). As a negative control, we also independently transfected the clonal *NF2* knockout cell lines with p99EGFP/LRE3_111, a retrotransposition-deficient L1 mutant that contains two missense mutations within the carboxyl-terminal RNA binding domain of ORF1p (RR261-262AA). Flow cytometry revealed that untreated N19_7 and N60_4 cells exhibited a 6-fold and 7-fold increase in EGFP-positive cells, respectively, when compared to control wild-type PA-1 cells (Figure 3.5 C), suggesting that the clonal knockout of *NF2* attenuated L1-REPEL.

To determine if *NF2* knockout led to a complete abolition of L1-REPEL, we next compared the numbers of EGFP-positive cells in TSA-treated *vs*. untreated cell lines. Consistent with previous reports, TSA treatment of wild-type PA-1 cells induced a ~24-fold increase in the number of EGFP-positive cells (Figure 3.5 D). In contrast, TSA

treatment of the N19_7 or N60_4 clonal *NF2* knockout cell lines only resulted in a modest increase (~2.5-fold and ~3-fold, respectively) in the number of EGFP-positive cells (Figure 3.5 D). Thus, these results demonstrate that loss of NF2/merlin diminishes, but does not abolish, L1-REPEL in PA-1 cells.

The generation of NF2/merlin expression vectors to rescue L1-REPEL

We next sought to determine if the reintroduction of NF2/merlin into select clonal *NF2* knockout cell lines was sufficient to rescue efficient L1-REPEL. Because our results demonstrate that NF2/merlin is necessary for efficient L1-REPEL (Figure 3.5), we hypothesized that re-expressing wild-type *NF2* from a cDNA may be sufficient to re-establish efficient L1-REPEL in clonal *NF2* knockout cells (Figure 3.6 A).

The human *NF2* gene encodes 17 exons that are subject to alternative splicing and previous studies revealed that two predominant *NF2* isoforms, and several minor isoforms, are expressed in human cells (Golovnina et al., 2005; Gutmann et al., 1995; Meng et al., 2000; Zoch et al., 2015). Isoform 1 and isoform 2 represent the most abundant, full-length, *NF2* isoforms and can be differentiated by the presence or absence of exon 16; skipping of exon 16 leads to isoform 1 formation, whereas the retention of exon 16 leads to isoform 2 formation.

We first analyzed published PA-1 RNA-seq datasets (Flasch et al., 2019) to determine which *NF2* isoform is predominantly expressed in PA-1 hEC cells. Although we identified ~400 sequence reads at exons 14, 15, and 17, we only observed a background sequence level of reads at exon 16, indicating that *NF2* isoform 1 is predominantly expressed in PA-1 cells (Figure 3.6 B). Intriguingly, the predicted

AlphaFold structure of the NF2/merlin isoform 1 exhibits head-to-tail intramolecular association (Figure 3.6 C) (AlphaFold: NF2/merlin).

To gain preliminary insight into regions of NF2/merlin that may be required for efficient L1-REPEL, we used site-directed to generate a panel of mutant *NF2* isoform 1 cDNAs based upon the published literature about NF2/merlin biology. We then expressed these cDNAs in *NF2* knockout cells using a mammalian expression vector (OriGene, SC124024) (Figure 3.7 A; also see text below) (Alfthan et al., 2004; Gutmann et al., 1995; Kim et al., 1997; Laulajainen et al., 2008; Laulajainen et al., 2011; Rong et al., 2004; Shaw et al., 2001; Surace et al., 2004; Tang et al., 2007; Xiao et al., 2002).

We first introduced the NF2 patient derived L64P mutation into the NF2 isoform 1 cDNA. This mutation is thought to alter the conformation of the NF2/merlin amino-terminus, thereby eliminates self-association between the amino- and carboxyl termini, and is thought to promote an “open” protein confirmation NF2/merlin that inhibits its tumor suppressor function (Gutmann et al., 1998).

We also introduced putative phosphorylation-deficient (S518A) or phospho-mimetic (S518D) mutations at serine 518 into the NF2 isoform 1 cDNA. Previous studies revealed that serine 518 dephosphorylation is thought to promote a closed conformation necessary for NF2/merlin tumor suppressor activity (Alfthan et al., 2004; Rong et al., 2004; Shaw et al., 2001; Surace et al., 2004; Xiao et al., 2002), whereas serine 518 phosphorylation inhibits intramolecular complex formation, promoting an “open” confirmation of the protein, allowing it to serve as a scaffold protein at the plasma membrane for receptor-mediated signal transduction.

PKA also phosphorylates NF2/merlin at serine 10 (Laulajainen et al., 2008; Laulajainen et al., 2011). Thus, we introduced S10A and S10D substitutions into the isoform 1 *NF2* cDNA. Previous studies suggest that NF2/merlin phosphorylation by PKA may be important for the cyclic AMP-PKA signaling axis (Kim et al., 1997). Furthermore, AKT phosphorylates at threonine 230 and serine 315, which promotes formation of the NF2/merlin "open" conformation, decreases NF2/merlin binding to phosphoinositides, and decreases NF2/merlin ubiquitination (Tang et al., 2007). Thus, we introduced the following substitutions into NF2 isoform 1 cDNA: T230A, T230D, S315A and S315D.

Using a similar approach, we also generated truncating mutations throughout NF2/merlin to identify protein domains important in re-establishing L1-REPEL. We introduced the following nonsense mutations into NF2 isoform 1 cDNA: A145X, R310X, N371X, Y481XX, K510X and L580X (Figure 3.7 A). The A145X mutation is located at the conserved AVQ sgRNA-target site and could mimic a potential nonsense mutation generated by the CRISPR/Cas9 knockout system. The AVQdel mutant lacks 9 bp at the sgRNA-target site present in clone N19_7 (see Figure 3.3 E). The 1bpins mutant contains a single adenine nucleotide insertion at the sgRNA-target site present in clone N19_7 (see Figure 3.3 E), resulting in a nonsense mutation 5 residues downstream of the insertion. The R310X mutant produces a nonsense mutation at the FERM-helical domain border. The Y481X mutant produces two nonsense mutations in exon 13 of the helical domain. The K510X mutant produces a nonsense mutation at the border of the helical domain and C-terminal domain. The L580X mutant lacks 15 carboxyl-terminal amino acid residues specific to *NF2* isoform 1.

Transient expression of NF2/merlin cDNAs in *NF2* knockout cells

We transiently transfected wild-type and mutant *NF2* cDNAs into N19_7 cells to assess stable expression of NF2/merlin protein (Figure 3.7 B). Briefly, whole cell lysates (WCLs) were collected and subjected to western blot analysis for NF2/merlin protein expression using the Invitrogen polyclonal N-terminal antibody raised against residues 65-95. We observed a robust 70 kD band in WCLs derived from cells transected with the wild-type *NF2* isoform 1 cDNA (Figure 3.7 B), but not in WCLs derived from cells transfected with the mutant AVQdel, 1bpins, or A145X constructs (Figure 3.7 B). These results indicate that the highly conserved residues at the 3' end of exon 4 likely are necessary for NF2/merlin protein stability. We also observed weak bands in WCLs derived from cells transfected with FERM domain mutants L64P and R310X, suggesting that these mutations within the highly organized FERM domain disrupt protein stability. By comparison, we observed robust NF2/merlin expression in WCLs derived from cells transfected with the Y481XX, K510X and L580X mutant cDNAs, indicating stable expression of the truncated NF2/merlin proteins. We also detected stable expression of the Ser518 mutant proteins, S518A and S518D.

Expression of full-length *NF2* cDNA rescues L1-REPEL in *NF2* knockout cells

To determine whether NF2/merlin expression could re-establish L1-REPEL, we transiently co-transfected N19_7 *NF2* knockout cells with candidate *NF2* isoform 1 cDNAs and p99EGFP/LRE3 and performed the L1-*GFP* retrotransposition assay (Figure 3.8 A: top). The co-transfection of wild-type *NF2*, L580X, or S518A resulted in similarly low levels of EGFP-positive cells, suggesting their expression efficiently restored L1 REPEL. By comparison, expression of the other *NF2* mutants cDNAs

showed varying levels of EFGP-positive cells (ranging from no suppression [R310X] to milder degrees of suppression [L64P, AVQdel, 1bpins, A145X, Y481X, K510X, and S518D]). The variation in this assay suggests that the expression of some isoform 1 NF2/merlin cDNAs did not or only mildly allowed the re-establishment of L1-REPEL (Figure 3.8 A: bottom).

To gain an independent evaluation of whether NF2/merlin isoform 1 cDNA expression could re-establish L1-REPEL, we transiently transfected pJM101/L1.3 or pJJ101/L1.3, a retrotransposition-competent L1 equipped with a *mblastI* retrotransposition indicator cassette, into clonal *NF2* knockout cells (Figure 3.8). For the *L1-neo* assays, the *NF2* knockout cell lines were co-transfected with a vector expressing *EGFP* (pCEP-GFP) as a transfection control and either wild-type *NF2* or mutant R310X *NF2* isoform 1 cDNAs. We observed robust levels of G418-resistant foci in clonal *NF2* knockout cells co-transfected with pCEP4-GFP and mutant *NF2* R310X. By comparison, expression of the wild-type *NF2* isoform 1 cDNA led to ~3-fold fewer G418-resistant foci when compared to the results from the pCEP-GFP and *NF2* R310X co-transfection experiments (Figure 3.8 B).

For the L1*blast* assays, knockout *NF2* cell lines were co-transfected with pJJ101/L1.3 and either wild-type *NF2*, mutant *NF2* A145X, or mutant *NF2* Y481XX isoform 1 cDNAs. We observed robust levels of blasticidin-resistant foci in clonal *NF2* knockout cells co-transfected with pJJ101/L1.3 and mutant *NF2* A145X or mutant *NF2* Y481XX isoform 1 cDNAs (Figure 3.8 B). By comparison, expression of the wild-type *NF2* isoform 1 cDNA led to ~5-fold fewer blasticidin-resistant foci when compared to the results from the *NF2* A145X and *NF2* Y481XX co-transfection experiments (Figure 3.8

B). Notably, blasticidin selection occurs much quicker than G418 selection (~5 days vs ~12 days). Together, the above results suggest that transient expression of wild-type *NF2* is sufficient to partially re-establish L1-REPEL in clonal *NF2* knockout cells.

In an effort to more efficiently re-establish L1-REPEL in clonal *NF2* knockout cells, we modified our experimental approach (Figure 3.8 C). We reasoned that L1-retrotransposition events occurring before the NF2/merlin protein is adequately expressed may escape L1-REPEL. Thus, to ensure that NF2/merlin protein is adequately expressed before the onset of L1-retrotransposition, we initially co-transfected clonal *NF2* knockout cells with a *NF2* isoform 1 cDNA expression vector and pCDNA6 (blast^R) and then subjected cells to blasticidin selection for two days (Figure 3.8 C: A). The resultant blasticidin-resistant cells then were co-transfected with pJM101/L1.3 and a *NF2* isoform 1 cDNA expression vector and subjected to the L1-*neo* retrotransposition assay (Figure 3.8 C: B). The serial transfection (A+B) approach resulted in a more severe reduction of G418-resistant foci (*i.e.*, 0 to 1 focus per well) when compared to our prior experiments (Figures 3.8B and 3.8 C: A), indicating an efficient re-establishment of L1-REPEL. Consistent with our previous results, the serial transfection approach using the *NF2* Y481XX or *NF2* R310X isoform 1 did not change the levels of G418-reistant foci. Control co-transfection experiments further indicated that the expression of *NF2* isoform 1 cDNA and pCDNA3 (neo^R) exhibited similar levels of G418-resistant foci, suggesting that *NF2* over-expression does not result in cell toxicity or decreased cellular proliferation in clonal *NF2* knockout cells.

Expression of *NF2* cDNA does not influence L1-retrotransposition in somatic cells

We next tested whether the expression of *NF2*/merlin influenced levels of L1-delivered reporter gene expression in other representative immortalized somatic cell types (Figure 3.9). Briefly, we co-transfected HeLa-JVM and HAP1 cells with pJM101/L1.3 and either pCEP-GFP, wild-type *NF2*, or mutant *NF2* R310X isoform 1 cDNAs. We observed similar levels of G418-resistant foci in each transfection condition, suggesting that overexpression of NF2/merlin isoform 1 cDNA not influence L1-delivered reporter gene expression in these cell lines.

NF2/merlin is not required to maintain L1-REPEL

We next tested whether NF2/merlin expression affects the initiation and/or maintenance steps of L1-REPEL. We utilized clonal pk5 and pc39 cells, which harbor mitotically stable silenced L1-GFP retrotransposition events, to assess whether knocking out *NF2* led to an increase in the numbers of EGFP-positive cells. We reasoned that if NF2/merlin is necessary to maintain L1-REPEL, knockout of *NF2* in both the clonal pk5 and pc39 cell lines would result in the reactivation of EGFP expression.

To determine whether *NF2* was necessary to maintain L1-REPEL, we transfected pk5 and pc39 cells with PX459-derived vectors containing sgRNAs targeting *NF2* (N_19 and N_60) (Figure 3.10 A). As controls, we also transfected pk5 and pc39 cells with an empty PX459 vector or a vector containing a sgRNA targeting *XPO7.* Briefly, the transfected pk5 and pc39 cells were selected with puromycin to generate a population of knockout cells. Six days post-transfection, the cells then were treated with TSA (500 nM) or left untreated. Eighteen hours later, TSA-treated, and untreated cells were

collected and subjected to flow cytometry to quantify the numbers of EGFP-positive cells (Figure 3.10 A). Flow cytometry analyses revealed similar levels of EGFP-positive cells in untreated "population knockout" pk5 cells independent of the sgRNA used in these experiments (Figure 3.10 B: blue bars). Similarly, TSA treatment efficiently reactivated EGFP expression in all the "population knockout" pk5 cells (Figure 3.10 B: red bars). Similar trends also were observed in pc39 cells (Figure 3.10 C). Together, these results suggest that knockout of *NF2* is not sufficient to reactivate a previously silenced L1-*GFP* retrotransposition event in pk5 or pc39 cells.

Differentiation reduces L1-REPEL in clonal *NF2* knockout cells

Previous studies demonstrated that L1-REPEL is more efficient in PA-1 cells than in actively differentiating PA-1 cells (Garcia-Perez et al., 2010). To determine if differentiation influenced L1-REPEL in clonal *NF2* knockout cells, we performed the L1-GFP-based retrotransposition assay in the presence of differentiation media. Briefly, we transfected p99EGFP/LRE3 into wild-type PA-1 and clonal *NF2* knockout cells. Twenty-four hours post-transfection, cells were selected in puromycin (2 µg/mL) and the L1-GFP-based retrotransposition assay was conducted the presence of either normal FBS-containing media (FBS) or differentiation media (DM) (Figure 3.11 A). On day 6 post-transfection, the cells were treated with TSA (500 nM) or left untreated. Eighteen hours later, the TSA-treated and untreated cells were collected and subjected to flow cytometry to quantify the number of EGFP-positive cells (Figure 3.11 A).

Consistent with previous results (Garcia-Perez et al., 2010), we observed a ~10-fold increase in the number of EGFP-positive cells when PA-1 cells where cultured in DM *vs.* FBS media (Figure 3.11 B). Furthermore, PA-1 cells cultured in DM only exhibited a

~3-fold increase in EGFP-positive cells upon TSA treatment when compared to the ~24-fold increase observed in TSA-treated PA-1 cells cultured in FBS (Figure 3.11 B).

Clonal N60_4 *NF2* knockout cells exhibited a ~6-fold increase in the number of EGFP-positive cells when PA-1 cells where cultured in FBS *vs.* DM media (Figure 3.13 C). Furthermore, N60_4 cells cultured in FBS only exhibited a ~3-fold increase in the number of EGFP-positive cells upon TSA treatment when compared to untreated N60_4 cells (Figure 3.11 C). By comparison, N60_4 *NF2* knockout cells cultured in DM only exhibited a mild increase in the number of EGFP-positive cells upon TSA treatment (~32% untreated *vs.* ~42% TSA) (Figure 3.11 C). These data suggest that L1-REPEL is less efficient in actively differentiating *NF2* knockout cells.

Previous fluorescent microscopy studies demonstrated that culturing PA-1 cells in DM media decreased the expression of the pluripotency marker OCT3/4 (Garcia-Perez et al., 2010). To evaluate OCT3/4 expression in our differentiation experiments, whole cell lysates were collected and subjected to western blot analysis using a OCT3/4 monoclonal N-terminal antibody raised against residues 1-134 (Santa Cruz, sc-5279) (Figure 3.11 D). Ribosomal protein S6 was used as a loading control (Cell Signaling, 2217). Consistent with previous results (Garcia-Perez et al., 2010), PA-1 cells cultured in DM media expressed less OCT3/4 compared to wild-type PA-1 cells cultured in FBS media. Intriguingly, N60_4 *NF2* knockout cells cultured in FBS media expressed less OCT3/4 compared to wild-type PA-1 cells cultured in FBS media (Figure 3.11 D). However, and somewhat paradoxically, N60_4 cells cultured in DM media expressed more OCT3/4 compared to N60_4 cells cultured in FBS media (Figure 3.11 D). Together, these results demonstrate that culturing wild-type PA-1 cells, but not N60_4

NF2 knockout cells, in DM media reduces the expression of the pluripotency marker OCT3/4.

Discussion

"Population knockout" of NF2 decreases L1-REPEL in PA-1 cells

We previously identified NF2/merlin as a candidate host factor necessary for efficient L1-REPEL in PA-1 hEC cells (Chapter 2). To validate that *NF2* was necessary for efficient L1-REPEL, we generated a population of *NF2* knockout cells by transiently expressing plasmid vectors containing *NF2*-targeting sgRNAs and the Cas9 protein. We demonstrated that knocking out *NF2* in a population of PA-1 cells resulted in increased L1-delivered reporter gene expression, suggesting that *NF2* is necessary for efficient L1-REPEL (Figure 3.1). Importantly, control L1-*neo* retrotransposition assays revealed that a population of *NF2* knockout cells transfected with pJM105/L1.3 did not lead to the formation of G418-resistant foci, demonstrating that G418-resistant foci only arise upon *bona fide* L1-mediated retrotransposition. By comparison, knocking out *NF2* in a population of HeLa JVM or HAP1 cells actually led to a mild decrease in the levels L1-delivered reporter gene expression (Figure 3.2). Together, these results suggest that *NF2*/merlin expression is necessary for efficient L1-REPEL in PA-1 hECs and potentially could differentially impact L1 biology in human embryonic-derived cells *vs.* other somatic-cell types.

Generation of clonal *NF2* knockout cell lines

We used CRISPR/Cas9-based approaches with two different sgRNAs (N19 and N60), in conjunction with dilution cloning, to generate several *NF2* knockout clonal cell

lines lacking functional NF2/merlin expression (Figure 3.3). Western blot analysis revealed the absence of NF2/merlin protein expression in several clonal cell lines (N19_7, N19_11, N19_13, N60_4, N60_9, N60_10, N60_11 and N60_12), whereas two clonal cell lines (N60_3 and N60_8) exhibited stable NF2/merlin expression (Figure 3.3 B). PCR followed by Sanger sequencing of the resultant products revealed that cell lines lacking NF2/merlin protein expression contained two mutant *NF2* alleles (Figure 3.3 D). Interestingly, clones N60_4 and N60_9 contained large insertions of 254 bp and 60 bp, respectively, at one allele of *NF2*. The 254 bp insertion within N60_4 mapped to *NF2* intron 3, suggesting that local sequences were used as a template to repair the Cas9-induced double strand break (Morrish et al., 2002). The 60bp insertion within N60_9 mapped to the Cas9 coding sequence encoded by the PX459 vector. Notably, some cell lines contained in-frame deletions (N19_7, N19_11, N60_4 and N60_12), which led to the deletion of amino acids in the *NF2* protein that have been conserved over the last ~450 million years of evolution (Figure 3.3 E), suggesting that these amino acids are necessary for the structural integrity and/or stability of NF2/merlin.

NF2 knockout clones do not exhibit increased proliferation

Cell growth and survival is regulated by a variety of environmental cues, including intercellular and matrix adhesions as well as growth factor signaling (Brizzi et al., 2012; DeMali et al., 2014) Proliferating cells typically undergo growth arrest after contacting adjacent cells and form intercellular junctions (McClatchey and Yap, 2012). Cancer cells typically evade contact inhibition, leading to abnormal growth and tumor formation (Hanahan and Weinberg, 2011). The *NF2* tumor suppressor gene encodes the FERM-containing domain protein merlin, which has been implicated in intercellular adhesion

and attachment to the cell membrane (Cooper and Giancotti, 2014; Shaw et al., 1998). Previous reports have demonstrated that loss of the *NF2* tumor suppressor gene can result in increased cell proliferation (Lallemand et al., 2003; Morrison et al., 2001; Okada et al., 2005; Rouleau et al., 1993; Trofatter et al., 1993; Xiao et al., 2005). Thus, to determine if loss of *NF2* increased cell proliferation, we performed a cell growth assay in *NF2* knockout clones (Figure 3.4). Interestingly, *NF2* knockout clones actually exhibited a decrease in cell proliferation when compared to wild-type PA-1 cells, suggesting that *NF2* knockout does not lead to a proliferative advantage in PA-1 cells.

Although *NF2* knockout did not confer a proliferative advantage to PA-1 cells, we did notice microscopically visible differences in cell volume, which possibly could be attributed to disrupted organization of the actin cytoskeleton (Laulajainen et al., 2008). Intriguingly, cell volume and cytoskeletal remodeling are hallmarks of the osmotic stress response. Indeed, a recent study demonstrated that osmotic stress promoted NF2/merlin lipid-binding at the plasma membrane, which, in turn, activated Hippo signaling (Hong et al., 2020). However, NF2/merlin mutants unable to bind PI (4,5)P_2 did not activate Hippo signaling in response to osmotic stress. In addition, hypo-osmotic cellular conditions have been implicated in the loosening of chromatin loosening and increased RNA polymerase II activity (Lima et al., 2018). Thus, it is reasonable to speculate that prolonged osmotic stress in *NF2* knockout cells may influence L1-REPEL.

It is noteworthy that a small molecule drug screen identified cardiac glycosides as candidate drugs that reversed L1-REPEL in pc39 cells (unpublished, Aurelien Doucet). Cardiac glycosides inhibit the sodium-potassium pump, leading to intracellular sodium

accumulation (Smith, 1989). These changes in ion homeostasis typically affect the osmotic regulation of cell volume. Thus, these preliminary results may further support a role for osmotic stress in maintaining L1-REPEL.

NF2 is necessary for efficient L1-REPEL

We performed the L1 retrotransposition assay in clonal *NF2* knockout cell lines to assess the efficiency of L1-REPEL. *NF2* knockout clonal cell lines subjected to the L1-*neo* retrotransposition assay resulted in an increase in G418-resistant foci (Figure 3.5 A). Notably, multiple *NF2* knockout clones, which were derived using two different *NF2*-targeting sgRNAs (N19 and N60), exhibited similar trends in L1-delivered reporter expression. Thus, these data indicate that *NF2* is necessary for efficient L1-REPEL in hECs. Moreover, these results further validate the utility of the transient Cas9-sgRNA-mediated knockout system in generating a population of *NF2* knockout cells to study L1-REPEL. We imagine that similar approaches will allow the identification of other candidate genes affecting L1-REPEL.

To examine the efficiency of L1-REPEL we subjected clonal *NF2* knockout cells to the L1-*GFP* retrotransposition assay (Figure 3.5 B). *NF2* knockout clones exhibited a ~6-fold (N19_7) and ~7-fold (N60_4) increase, respectively, in the number of EGFP-positive cells when compared wild-type PA-1 cells when grown in normal FBS-containing culture medium (Figure 3.5 C). Treatment of cells with TSA resulted in a ~2.5 to 3-fold increase in the number of EGFP-positive cell in the *NF2* 19_7 and 60_4 knockout cells as compared to a ~24-fold increase in the number of EGFP-positive PA-1 cells when the cells were grown in normal FBS-containing culture medium (Figure 3.5 D: compare blue and red bars). It is noteworthy that the TSA-treatment experiments

indicate that only a subset of L1-*GFP* retrotransposition events appear to be subjected to L1-REPEL in PA-1 cells; indeed, we speculate that redundant silencing mechanisms may act to ensure efficient silencing of L1-delivered reporter genes. Moreover, if we assume the number of TSA-treated EGFP-positive cells represent the total number of cells containing L1-*GFP* retrotransposition events, we can conclude that *NF2* knockout does not lead to more retrotransposition in PA-1 cells. In sum, our results demonstrate that loss of NF2/merlin significantly diminishes L1-REPEL in PA-1 cells.

Expression of full-length *NF2* cDNA rescues L1-REPEL in *NF2* knockout cells

Previous studies identified two predominant and several minor *NF2* isoforms expressed in human cells (Golovnina et al., 2005; Gutmann et al., 1995; Meng et al., 2000; Zoch et al., 2015). RNA-seq analysis revealed isoform 1 as the predominant *NF2* isoform expressed in PA-1 cells (Figure 3.6 B). Although we did not identify appreciable difference in RNA-seq read-depth indicative of other *NF2* isoforms, it remains possible that the low-level expression of other *NF2* isoforms could impact L1-REPEL. However, based on our RNA-seq analysis, we decided to test *NF2* isoform 1 in the L1-REPEL *NF2* rescue experiment (3.6 A).

We generated wild-type *NF2* isoform 1 as well as a variety mutant *NF2* isoform 1 cDNA expression constructs (Figure 3.7 A). We demonstrated that mutations disrupting the FERM domain generally depleted NF2/merlin expression. Intriguingly, transient expression of the AVQ mutant cDNA did not produce stable NF2/merlin protein (Figure 3.7 B), supporting our hypothesis that highly conserved residues at the *NF2* sgRNA-target site are necessary for protein stability (see Figure 3.3 E).

We tested wild-type *NF2* along with several mutants in the L1-REPEL *NF2* rescue assay. Expression of wild-type *NF2* in clonal *NF2* knockout cells decreased the number of EGFP-positive cells ~6-fold (Figure 3.8: pCMV6-empty *vs*. WT NF2) in the L1-*GFP* assay and led to a ~3-fold and ~5-fold decrease in the number of drug resistant foci in the L1-*neo* and L1-*blast* assays, respectively. Blasticidin selection occurs more quickly than G418 selection in PA-1 cells (~5 days *vs*. ~12 days). Thus, the timing and duration of selection could explain the increased levels of drug-resistant foci in the L1-*blast* assay compared to the L1-*neo* assay. By comparison, the expression of the A145X, R310X, and Y481XX truncation mutants did not drastically alter the levels of drug resistant cells in the L1 retrotransposition assay (Figure 3.8). Because A145X, R310X, and Y481XX mutant cDNAs were expressed at similar levels as the wild-type NF2 cDNA in PA-1 cells, these data argue that these mutants do not act via a dominant-negative mechanism. In sum, the above data suggest that transient expression of wild-type *NF2* is sufficient to partially re-establish L1-REPEL in a subset of clonal *NF2* knockout cells.

We hypothesized that incomplete rescue of L1-REPEL observed in the above experiments was, at least in part, due to L1 retrotransposition events occurring before adequate levels of NF2/merlin protein were expressed in *NF2* knockout cell lines. Consistent with this interpretation, a serial transfection approach, which allowed the establishment of NF2/merlin protein before assaying for retrotransposition, allowed the efficient re-establishment (a.k.a., rescue) of L1-REPEL (Figure 3.8 D). Furthermore, control experiments revealed that *NF2* cDNA expression did not result in toxicity or decreased proliferation of clonal *NF2* knockout cells (Figure 3.8 D) and did not alter L1-

neo retrotransposition in HeLa-JVM and HAP1 cells did not influence levels of G418-resistant foci in the L1-*neo* assay (Figure 3.9). Thus, the expression of NF2/merlin cannot simply establish L1-REPEL in non-hEC cell lines. Together, these data suggest that NF2/merlin isoform 1 cDNA expression is sufficient to re-establish efficient L1-REPEL in *NF2* knockout PA-1 cells. Future studies using the serial transfection approach should allow a rigorous means to test additional *NF2* mutant cDNAs to identify aspects of NF2/merlin biology necessary for L1-REPEL.

Intramolecular association between the NF2/merlin amino-terminal FERM domain and its carboxyl-terminus tail is necessary for NF2/merlin to adopt a “closed” conformation that is critical for its nuclear localization and tumor suppressor activity. Intriguingly, *NF2* isoform 1 has an extended carboxyl-terminal tail when compared to *NF2* isoform 2. Moreover, isoform 2 lacks carboxyl-terminal amino acids necessary for intramolecular binding, leading to an open conformation (Gonzalez-Agosti et al., 1999; Sherman et al., 1997). Thus, it is tempting to speculate that the ability of *NF2* isoform 1 to adopt a “closed” conformation may be necessary for L1-REPEL. Interestingly, “closed” NF2/merlin can translocate to the nucleus where it binds to DCAF1 (VPRBP) and modulates $CRL4^{DCAF1}$ E3 ubiquitin ligase activity (Li et al., 2010). DCAF1 has been implicated in regulating histones (H2A and H3) and HDACs, which appear to be key components of the L1-REPEL silencing mechanism (Kim et al., 2012; Kim et al., 2013). Thus, the NF2/merlin-mediated inhibition of $CRL4^{DCAF1}$ might be necessary for efficient L1-REPEL in PA-1 cells.

NF2 is not required to maintain L1-REPEL

Our working model predicts that L1-REPEL requires both initiation and maintenance phases to efficiently and stably silence L1-delivered reporter genes in PA-1, and perhaps other, hEC cells. Thus, we investigated the role of NF2/merlin in the maintenance step of L1-REPEL by knocking out *NF2* in clonal pk5 and pc39 cell lines (Figure 3.10). The pk5 cell line harbors a single full-length silenced L1-*GFP* retrotransposition event, whereas the pc39 cell line harbors three silenced L1-*GFP* retrotransposition events (Flasch et al., 2019; Garcia-Perez et al., 2010). We reasoned that if NF2/merlin is necessary to maintain L1-REPEL, knockout of *NF2* in pk5 and pc39 cells would result in the reactivation of EGFP reporter gene expression.

To test the above hypothesis, we used the N19 and N60 sgRNAs to knockout NF2 expression in a population of pk5 or pc39 cells. The knockout of *NF2* using this approach did not lead to an increase the number of EGFP-positive pk5 or pc39 cells when compared to the background levels observed in PX459-empty vector transfected cells (Figures 3.10 B and C). Importantly, TSA treatment efficiently reactivated EGFP expression in the pk5 and pc39 “population knockout” cells independent of the sgRNAs used in our experiments (Figure 3.10 B and C: red bars). These data suggest that the knockout of *NF2* is not sufficient to reactivate previously silenced L1-*GFP* retrotransposition events in pk5 or pc39 cells and argue against a role of *NF2* in the maintenance of L1-REPEL. Instead, we propose that *NF2* may be necessary to directly or indirectly establish and/or initiate efficient L1-REPEL in PA-1 hEC cells.

Differentiation reduces L1-REPEL in clonal *NF2* knockout cells

Previous studies demonstrated that L1-REPEL is attenuated in PA-1 cells undergoing differentiation (Garcia-Perez et al., 2010). The differentiation media used in those studies consisted of knockout serum replacement (in the place of FBS) supplemented with the addition of all-*trans* retinoic acid (1μM); differentiation could not reactivate a previously silenced L1-*GFP* retrotransposition event in pk5 cells (Garcia-Perez et al., 2010).

We confirmed that PA-1 cells subjected to the *L1-GFP*-based assay exhibited a ~10-fold increase in the number of EGFP-positive cells when cultured in DM *vs.* FBS-containing media. If we assume that TSA treatment reactivates all *L1-GFP* retrotransposition events, we estimate that ~30% of actively differentiating cells escape L1-REPEL (Figure 3.11 B: DM blue bar vs DM red bar). Strikingly, these results are similar to those observed in *NF2* knockout cell lines cultured in FBS media (Figure 3.11 C), raising the possibility that *NF2* knockout potentially may, in part, phenocopy the effect of differentiation on L1-REPEL. However, we must state that culturing *NF2* knockout cells in DM led to an even further reduction of L1-REPEL and that under the conditions used in our assays ~75% of actively differentiating *NF2* knockout cells escaped L1-REPEL (Figure 3.11 C: DM blue bar vs DM red bar). These data suggest that L1-REPEL is less efficient in actively differentiating *NF2* knockout cells and that *NF2* knockout and differentiation may act in combination and/or synergistically to alleviate L1-REPEL.

OCT3/4 (*POU5F1*) is a critical “Yamanaka factor” that functions in the self-renewal of undifferentiated embryonic stem cells (Takahashi and Yamanaka, 2006). Previous

fluorescence microscopy studies demonstrated that PA-1 hEC cells express OCT3/4 (Garcia-Perez et al., 2010) and that culturing PA-1 cells in DM media resulted in decreased OCT3/4 expression (Garcia-Perez et al., 2010). Consistent with previous results (Garcia-Perez et al., 2010), we demonstrated that wild-type PA-1 cells cultured in DM media downregulated OCT3/4 expression when compared to PA-1 cells cultured in FBS-containing media (Figure 3.11 D). OCT3/4 expression was also decreased in *NF2* knockout cells cultured in FBS-containing media, but somewhat paradoxically, OCT3/4 expression was increased when cultured in DM media (Figure 3.11 D). Whether these results are peculiar to the individual *NF2* knockout cell line used in this study requires further investigation. However, to date, these results provide preliminary *prima facie* evidence that culturing PA-1 cells, but not *NF2* knockout cells, in DM media reduces the expression of the pluripotency marker OCT3/4. Notably, it also is possible that retinoic acid signaling is differentially regulated in *NF2* knockout cells when compared to PA-1 cells.

Conclusion

We demonstrated that NF2/merlin protein expression is necessary for efficient L1-REPEL in PA-1 cells. Additionally, we determined that the expression of the NF2/merlin isoform 1 cDNA efficiently re-established L1-REPEL, indicating that NF2/merlin expression is sufficient to rescue L1-REPEL in *NF2* knockout cells. We further demonstrated that *NF2* knockout is not sufficient to reactivate L1-REPEL in clonal pk5 and pc39 cells, suggesting that NF2/merlin is required, either directly or indirectly, L1-REPEL initiation. Finally, we found that culturing cells in differentiation media attenuated L1-REPEL, suggesting that *NF2* knockout and culturing cells in differentiation medium

may act independently or, perhaps, synergistically to attenuate L1-REPEL. In sum, our data indicate that the NF2/merlin tumor suppressor may contribute to L1-REPEL in PA-1 cells.

Materials and Methods

Cell lines and cell culture conditions

PA-1 cells were purchased from American Type Culture Collection and cultured as previously described (Garcia-Perez et al., 2010). All PA-1-derived cells were cultured in Minimum Essential Media (MEM) supplemented with 10% heat-inactivated fetal bovine serum, 2 mM Glutamax, 100 U/ml penicillin, 0.1 mg/ml streptomycin and 0.1 mM non-essential amino acids. Controlled heat inactivation of the fetal bovine serum was critical for conducting assays in PA-1 cells. HeLa JVM cells were cultured in Dulbecco's Modified Eagle Medium (DMEM) high glucose supplemented with 10% fetal bovine serum (FBS), 2 mM Glutamax, 100 U/ml penicillin and 0.1mg/ml streptomycin. HAP1 cells were cultured in Iscove's Modified Dulbecco's Medium (IMDM) supplemented with 10% fetal bovine serum (FBS), 2 mM Glutamax, 100 U/ml penicillin and 0.1mg/ml streptomycin. All cell lines were grown at 37ºC in a humidified 7% CO2 incubator.

Conditioned media for dilution cloning

PA-1 media was supplemented with 25% conditioned media. Conditioned media was harvested from wild-type PA-1 cell cultures at 70% confluence within 48 hours of the previous passage. Conditioned media was clarified by centrifugation at 1000 x g for 10 minutes in a bench top centrifuge and then passed through a 0.2 µm sterile filter. Conditioned media was stored at 4ºC for up to 1 week.

Expression vectors

All vectors were propagated in *Escherichia coli* strain DH5a (F-f80lacZDM15D[lacZYA-argF] U169 recA1 endA1 hsdR17 [rk-, mk+] phoA supE44 l- thi-1 gyrA96 relA1)

(Invitrogen). Competent *E. coli* were prepared and transformed using previously described methods (Inoue et al., 1990; Moran et al., 1996). Plasmids were prepared using the Qiagen Plasmid Midi Kit according to the manufacturer's instructions.

pJM101/L1.3: contains a full-length L1 (L1.3, GenBank: L19088) that includes the *mneoI* retrotransposition indicator cassette within its 3'UTR (Sassaman et al., 1997; Wei et al., 2001). A CMV promoter and SV40 polyadenylation signal in the pCEP4 plasmid backbone facilitate L1.3 expression. This vector was used to assay for L1-*neo* retrotransposition.

pJM105/L1.3: is identical to pJM101/L1.3 except for the presence of a missense mutation (D702A) in the L1.3 ORF2p reverse transcriptase (RT) domain, which renders L1.3 retrotransposition-defective (Wei et al., 2001). This vector was used as a negative control in L1-*neo* retrotransposition assays.

pCDNA3: expresses a neomycin resistance gene from the vector backbone This vector was used as a positive control for transfection and G418 drug selection in the L1-*neo* retrotransposition assays.

pJJ101/L1.3: is similar to pJM101/L1.3, but contains an *mblastI* retrotransposition indicator cassette within its 3'UTR (Goodier et al., 2007; Kopera et al., 2011; Morrish et al., 2002). A CMV promoter and SV40 polyadenylation signal in the pCEP4 plasmid backbone facilitate L1.3 expression. This vector was used to assay for L1-*blast* retrotransposition assays.

pJJ105/L1.3: is identical to pJJ101/L1.3 except for the presence of a missense mutation (D702A) in the L1.3 ORF2p reverse transcriptase (RT) domain, which renders L1.3

retrotransposition-defective (Goodier et al., 2007; Kopera et al., 2011; Morrish et al., 2002). This vector was used as a negative control for L1-*blast* retrotransposition assays.

pCDNA6: expresses a blasticidin resistance gene from the vector backbone This vector was used as a positive control for transfection and blasticidin drug selection in the L1-*blast* retrotransposition assays.

p99EGFP/LRE3: contains a full-length RC-L1 (LRE3) with an *mEGFPI* retrotransposition indicator cassette within its 3'UTR. LRE3 expression is driven from its native 5'UTR (Ostertag et al., 2000). The LRE3 expression construct was cloned into a version of pCEP4 that lacks the CMV promoter. A puromycin-resistance selectable marker replaced the hygromycin-resistance selectable marker in pCEP4 (Garcia-Perez et al., 2010). This vector was used to assay for L1-*GFP* retrotransposition.

p99EGFP/LRE3_111: is identical to p99EGFP/LRE3 except that it contains two missense mutations in LRE3 ORF1p (RR261-262AA), which renders LRE3 retrotransposition-defective(Zhang et al., 2014). This vector was used as a negative control in L1-*GFP* retrotransposition assays.

pCEP4/GFP: contains the coding sequence of the humanized Renilla reniformis green fluorescent protein (hrGFP) from phrGFP-C (Stratagene). GFP expression is driven by a cytomegalovirus (CMV) immediate early promoter and terminated at a simian virus 40 (SV40) late polyadenylation signal present in the pCEP4 plasmid backbone (Alisch et al., 2006). This vector was used to calculate transfection efficiencies.

Generation of candidate gene knockout vectors

The sgRNAs used in this study were identified using CRISPick software (Broad Institute) with the following parameters: Human GRCh38 reference genome; CRISPRko mechanism; SpyoCas9 enzyme. (https://portals.broadinstitute.org/gppx/crispick/public). *Bbs*I restriction sites were appended to sgRNA-containing oligonucleotides for cloning into *Bbs*I-digested PX459 plasmid (Ran et al., 2013a; Ran et al., 2013b). Sense (forward) and antisense (reverse) oligonucleotides were ordered from Integrated DNA Technologies (Table 3.1). Cloning of oligonucleotides into the PX459 vector was performed as previously described (Ran et al., 2013a; Ran et al., 2013b). Sense and antisense oligonucleotides were phosphorylated using T4 PNK (NEB) and then were annealed to create double stranded DNA. PX459 was digested by *Bbs*I (NEB), then the phosphorylated-and-annealed oligonucleotides were ligated into the *Bbs*I-digested PX459 vector. Digestion-ligation reactions were treated with PlasmidSafe exonuclease (Lucigen) and transformed into competent bacteria. Bacterial transformations were plated on LB-ampicillin plates and individual clones were picked and cultured overnight in LB-ampicillin broth at 37°C. Plasmid DNA was extracted, and Sanger sequenced to confirm the sgRNA insert.

Candidate gene "population knockout"

Cells were plated at $3x10^5$ (PA-1), $2x10^4$ (HeLa JVM) or $1x10^6$ (HAP1) cells/well of a six-well dish. Twenty-four hours later, each well was transfected with a FuGENE 6 mix containing 1 µg candidate knockout vector DNA and 3 µl FuGENE 6 (Promega, E2692) up to a final volume of 105 µL Opti-MEM (Life Technologies). One day later, the transfection-media was replaced with fresh media containing puromycin (2 µg/mL: PA-1

and HAP1; 5 μg/mL HeLa JVM) to select for vector expression. Once untransfected control cells died, puro-media was replaced with fresh PA-1 media. Cells were cultured and passaged once reaching confluency. 12 days post-transfection, “population knockout” cells were collected for cryopreservation and re-plated for the L1 retrotransposition assay.

L1 retrotransposition assay in “population knockout” cells

The L1 retrotransposition assay was conducted as previously described (Garcia-Perez et al., 2010; Kopera et al., 2016; Moran et al., 1996; Wei et al., 2001). “Population knockout” cells were plated at $3x10^5$ (PA-1), $2x10^4$ (HeLa JVM) or $1x10^6$ (HAP1) cells/per well of a six-well dish. Eighteen hours after plating, each well was transfected with a FuGENE 6 mix containing 1 μg plasmid DNA and 3 μl FuGENE 6 (Promega) up to a final volume of 105 μL Opti-MEM (Gibco, 31985062). One day later, the transfection-media was replaced with fresh media. Three days post-transfection, media was supplemented with 210 μg/mL of (G418) (Gibco, 10131035) to select for L1 retrotransposition events expressing the neomycin resistance reporter gene. After ~14 days of G418 selection, G418-resistant foci were washed with PBS and then fixed for 30 minutes at room temperature in a 1X PBS solution containing 2% paraformaldehyde (Sigma Aldrich) and 0.4% glutaraldehyde (Sigma Aldrich). Fixed foci were then stained with a 0.1% crystal violet solution for 2 hours with gentle rotation at room temperature for visualization.

Dilution cloning; generation of clonal NF2 knockout cell lines

“Population knockout” cells were counted using the Countess hemocytometer (Invitrogen, C10227) following the manufactures recommendations. Cells were diluted

to concentrations of 0.5-, 1-, 2- and 5-cells per 200µl of conditioned PA-1 media. 200µl of volume was dispensed into each well of two 96-well plates for each dilution. 5 days after plating, wells were visually inspected daily for established colonies. Wells containing single colonies were maintained in conditioned media during clonal outgrowth. Once cells reached confluency, cells were trypsinized and moved to larger plates (96-well, 24-well, 6-well, 10 cm dish). After reaching confluency in 10cm plates, cells were collected for cryopreservation.

Western blotting

The following protocol contains minor changes from the original protocol developed by Dr. John Moldovan and Dr. Peter Larson (Moldovan and Moran, 2015). Cells were washed with PBS, trypsinized (0.25% Trypsin-EDTA) and pelleted at 500×g for 2 minutes. Whole cell lysates were prepared by incubating cell pellets in ~500 µL (1 mL lysis buffer per 100 mg of cell pellet) of lysis buffer (20 mM Tris-HCl (pH 7.5), 150 mM NaCl, 10% glycerol, 1 mM EDTA, 0.1% NP-40 (Sigma), 1X complete EDTA-free protease inhibitor cocktail (Roche) on ice for 30 minutes. Lysates were centrifuged at 15,000×g for 10 minutes at 4°C and supernatants were transferred to a clean tube. Protein concentration was determined using the Bradford reagent assay (BioRad). For SDS-PAGE, samples were diluted in 4x Laemmli buffer (BioRad) containing 10% BME and incubated at 95°C for 5 minutes. 30µg of protein was loaded per well of a 4–15% precast mini-PROTEAN TGX gel (Bio-Rad) run at 200V for 35 minutes in 1X Tris/Glycine/SDS (25 mM Tris-HCL, 192 mM glycine, 0.1% SDS, pH 8.3) buffer (BioRad). Protein was transferred using the Trans-Blot Turbo Mini PVDF Transfer Packs (BioRad) with the Trans-Blot Turbo Transfer System (BioRad) at 2.5A and 25V

for 3 minutes. The membranes then were incubated at room temperature in Intercept blocking buffer (LI-COR) for 1 hour. Following blocking, fresh blocking buffer containing 0.1% Tween 20 (Sigma) was added with the following primary antibodies: NF2 (B-12) mouse monoclonal (Santa Cruz, sc-55575) at 1:1000; NF2 rabbit polyclonal (Sigma, HPA003097), at 1:1000; NF2 rabbit polyclonal (Invitrogen, PA535316), 1:1000; S6 Ribosomal Protein (5G10) rabbit monoclonal (Cell Signaling, 2217) at 1:2000; b-actin mouse monoclonal (ThermoFisher, MA1-744) at 1:2000. The membrane was incubated overnight at 4°C. The next day, the membrane was washed 5X with 1X PBS and fresh blocking solution containing 0.1% Tween 20 (Sigma) and 0.02% SDS was added with a combination of the following secondary antibodies: Anti-Mouse or anti-Rabbit IRDye 680LT (LICOR) at 1:15,000, and anti-Rabbit or anti-Mouse IRDye 800CW (LI-COR) at 1:15,000. The membrane was incubated for 1 hour at room temperature. Lastly, the membrane was washed 5X with 1X PBS and scanned using the Odyssey CLx scanner (LI-COR).

Genomic sequencing of clonal NF2 mutations

Genomic DNA was extracted from ~$1x10^6$ cells using the Wizard Genomic DNA Purification Kit (Promega, A1120) per the manufacturer recommendations. 200ng of genomic DNA was subjected to PCR amplification using Phusion High-Fidelity DNA polymerase (M0530L) following the manufacturers recommendations. Primer set A (Forward: 5'-ctggcagccctcattagaac-3'; Reverse: 5'-caaattaacgcccaggaaaa-3') amplification utilized a 61°C annealing temperature, whereas Primer set B (Forward: 5'-gagtgatcccatgacccaaat-3'; Reverse: 5'-ccacctgtctgcatcagtaaa-3') utilized a 63°C annealing temperature. PCR products were subjected to electrophoresis using a 1%

agarose gel. PCR products were TOPO cloned using the Zero Blunt TOPO PCR Cloning Kit (Invitrogen, 450245) and submitted for sanger sequencing.

Cell proliferation assay

Wild-type PA-1 and clonal *NF2* knockout cells were plated at $5x10^4$ cells/well of a 6-well plate in 2 mL of PA-1 media. Each day following plating, cells were imaged at 20x and counted using the Countess hemocytometer (Invitrogen, C10227) following the manufactures recommendations. All cell counts achieved >95% viability as determined by trypan blue staining. PA-1 media was replaced daily.

L1-neo and L1-blast retrotransposition assays in clonal NF2 knockout cells

The transient L1 retrotransposition assay was conducted as previously described (Garcia-Perez et al., 2010; Kopera et al., 2016; Moran et al., 1996; Wei et al., 2001). Cells were plated at $3x10^5$ cells/per well of a six-well dish. Eighteen hours after plating, each well was transfected with a FuGENE 6 mix containing 1 µg plasmid DNA and 3 µl FuGENE 6 (Promega) up to a final volume of 105 µL Opti-MEM (Gibco, 31985062). One day later, the transfection-media was replaced with fresh media. For L1-*neo* assays, media was supplemented with 210 µg/mL of geneticin (G418) (Gibco, 10131035) three days post-transfection to select for L1 retrotransposition events expressing the neomycin resistance reporter gene. For L1-*blast* assays, media was supplemented with 10 µg/mL of blasticidin (Gibco, A1113903) three days post-transfection to select for L1 retrotransposition events expressing the blasticidin resistance gene. After ~10 days (G418) or 5 days (blasticidin) of selection, drug-resistant foci were washed with PBS and then fixed for 30 minutes at room temperature in a 1X PBS solution containing 2% paraformaldehyde (Sigma Aldrich) and 0.4%

glutaraldehyde (Sigma Aldrich). Fixed foci were then stained with a 0.1% crystal violet solution for 2 hours with gentle rotation at room temperature for visualization.

NF2 rescue L1 retrotransposition assays

The transient L1 retrotransposition assay was conducted as previously described with the following modification: cells were co-transfected with the indicated NF2/merlin expression vector and a retrotransposition-competent L1-reporter construct (p99EGFP/LRE3, pJM101/L1.3 or pJJ101/L1.3). HeLa JVM cells were plated at $2x10^4$ and HAP1 at $1x10^6$ cells/well. For serial *NF2* rescue experiments, *NF2* knockout cells were co-transfected with the indicated NF2/merlin expression vector along with pCDNA6. Cells were selected with 10 µg/mL blasticidin until untransfected cells died, typically 48-72 hours post-transfection. Blasticidin-resistant cells were collected, counted, and plated at $3x10^5$ cells/well of a 6-well plate. Eighteen hours later, cells were co-transfected with the indicated NF2/merlin expression vector and pJM101/L1.3 or pcDNA3, then subjected to the L1-*neo* retrotransposition assay.

L1-GFP retrotransposition assays in clonal NF2 knockout cells

The L1-*GFP* retrotransposition assays was conducted as previously described (Garcia-Perez et al., 2010) Cells were plated at $3x10^5$ cells/per well of a six-well dish. Eighteen hours after plating, each well was transfected with a FuGENE 6 mix containing 1 µg plasmid DNA (p99EGFP/LRE3 or p99EGFP/LRE3 _111) and 3 µl FuGENE 6 (Promega) up to a final volume of 105 µL Opti-MEM (Gibco, 31985062). One day later, the transfection-media was replaced with puromycin (Gibco, A1113803) supplemented media (2 µg/mL). Once all untransfected cells died (~48 hours), puro-media was replaced with normal PA-1 media. 6 days post-transfection, cells were left untreated or

were treated for 18 hours with 500 nM Trichostatin A (TSA, Sigma). Flow cytometry determined the percentage of EGFP expressing cells in both the untreated and TSA-treated samples. The percentage of EGFP expressing cells obtained with a retrotransposition-defective L1 (p99EGFP/LRE3_111) was used as a negative control. Experiments were conducted.

Site-directed mutagenesis

The QuikChange II XL Site-Directed Mutagenesis Kit (Agilent, 200521) was used to generate mutations in a NF2/merlin expression vector (OriGene, SC124024). Mutant primers are listed in Table 3.2.

L1-REPEL maintenance assay

NF2 "population knockout" pk5 and pc39 cells were generated as previously described (Chapter 2: validation assays). Cells were plated at $3x10^5$ cells/well of a six-well plate in PA-1 media. Twenty-four hours later, cells were transfected with a FuGENE 6 mix containing 1 µg PX459 vector containing the indicated sgRNA (Table 3.1). 24 hours later, the transfection-media was replaced with puromycin (Gibco, A1113803) supplemented media (2 µg/mL). Once all untransfected cells died (~48 hours), puro-media was replaced with normal PA-1 media. 6 days post-transfection, cells were treated with or without 500 nM Trichostatin A (TSA, Sigma) for 18 hours. Flow cytometry determined the percentage of EGFP expressing cells in both the untreated and TSA-treated samples. The percentage of EGFP expressing cells obtained with a retrotransposition-defective L1 (p99EGFP/LRE3_111) served as our negative control. Experiments were conducted in triplicate.

Differentiation of PA-1 and NF2 knockout cells

PA-1 or *NF2* knockout cells were cultured in differentiation media containing 10% knockout serum replacement (KOSR, Invitrogen), substituting for 10% FBS, with the addition of 1 μM all-trans retinoic acid (RA, Sigma). Notably, in differentiation experiments, cells were plated and transfected in normal PA-1 media. 24 hours after transfection, media was replaced with differentiation media for the duration of the assay.

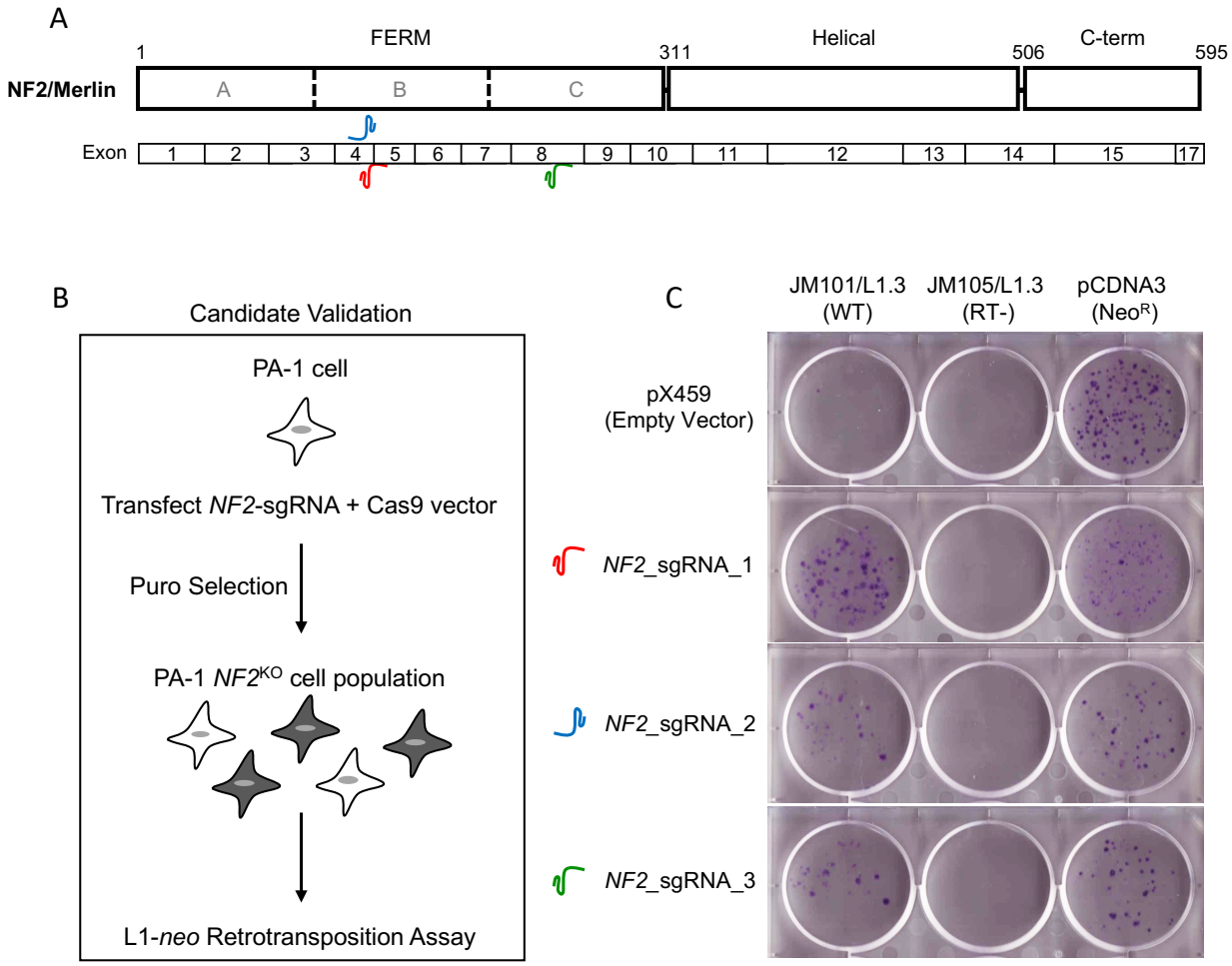


Figure 3.1: NF2 "population knockout" attenuates L1-REPEL in hEC PA-1 cells.

(A) Schematic of NF2/merlin isoform 1. The N-terminus of NF2/merlin encodes a conserved FERM (4.1, ezrin, radixin, moesin) domain followed by an α-helical domain and a C-terminal domain. sgRNA target-sites are indicated by ↯. (B) PA-1 cells were transfected with PX459-derived vectors expressing sgRNAs targeting *NF2* and Cas9. Transfected cells were selected with puromycin generate *NF2* "population knockout" cells that were subjected to the L1-*neo* retrotransposition assay. (C) Results of the L1-*neo* retrotransposition assay in *NF2* "population knockout" cells. pJM101/L1.3, a retrotransposition-competent L1; pJM105/L1.3, a retrotransposition-defective L1; pCDNA3, a vector that constitutively expresses the neomycin resistance gene. G418-resistant colonies that expressed the retrotransposed neomycin resistance gene were fixed and stained with crystal violet for visualization.

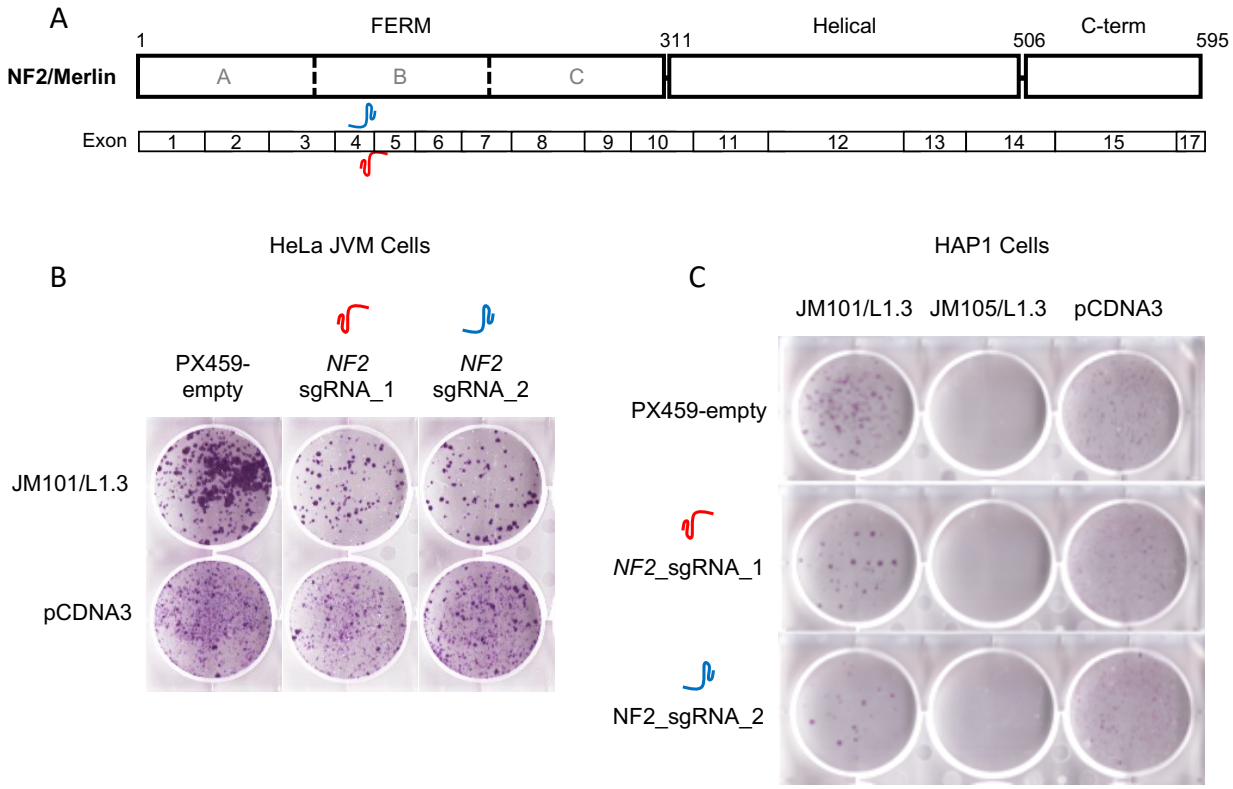
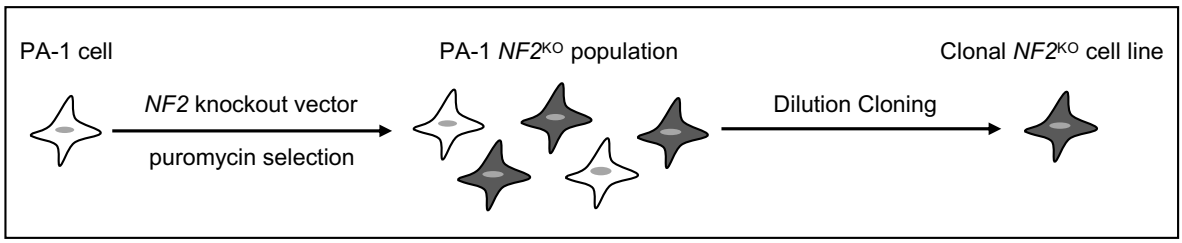


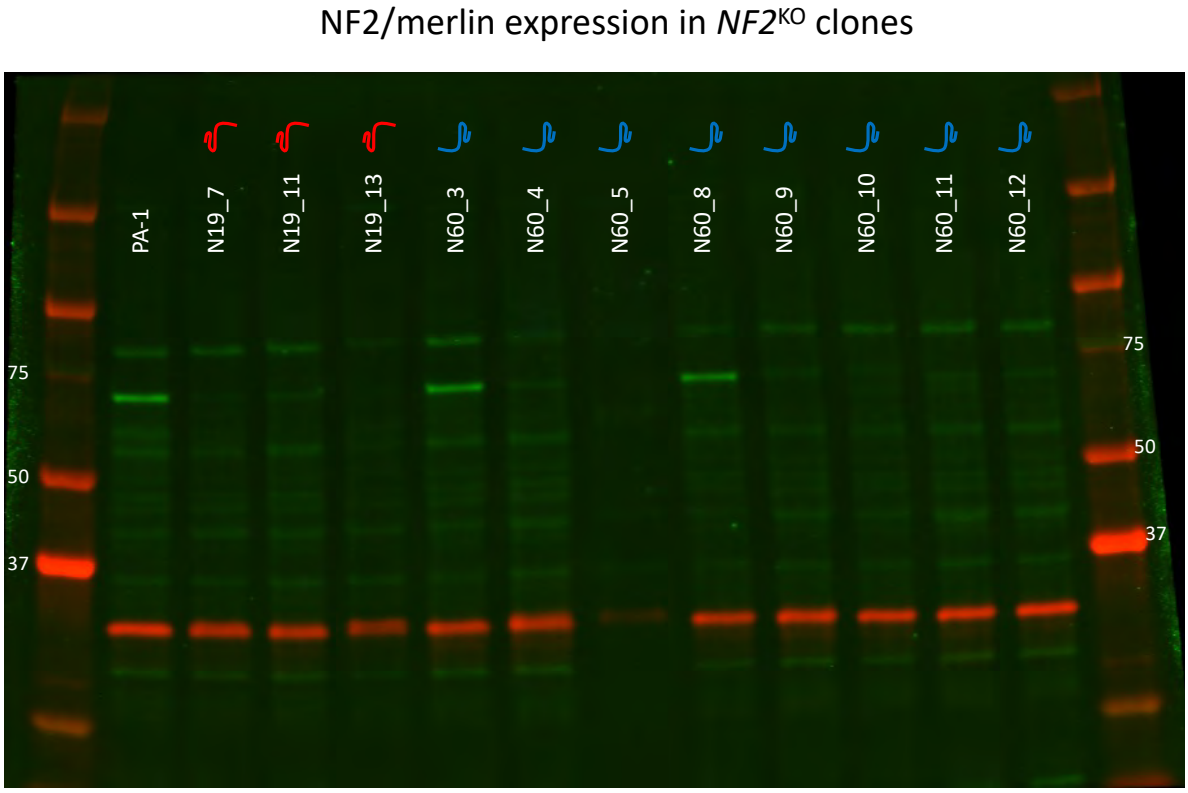
Figure 3.2: NF2 "population knockout" does not reduce L1-REPEL in somatic HeLa JVM and HAP1 cells.

(A) Schematic of NF2/merlin isoform 1 illustrating the sgRNA target-sites (). (B) Results of the L1-*neo* retrotransposition assay in *NF2* "population knockout" HeLa JVM cells. (C) Results of the L1-*neo* retrotransposition assay in *NF2* "population knockout" HAP1 cells. pJM101/L1.3, a retrotransposition-competent L1; pJM105/L1.3, a retrotransposition-defective L1; pCDNA3, a vector that constitutively expresses the neomycin resistance gene. G418-resistant colonies that expressed the retrotransposed neomycin resistance gene were fixed and stained with crystal violet for visualization.

A


PA-1 cell
PA-1 NF2KO population
Clonal NF2KO cell line
NF2 knockout vector
puromycin selection
Dilution Cloning

B


NF2/merlin expression in NF2KO clones
PA-1
N19_7
N19_11
N19_13
N60_3
N60_4
N60_5
N60_8
N60_9
N60_10
N60_11
N60_12
75
50
37
75
50
37
NF2/Merlin
70 kD
Ribosomal S6
32 kD

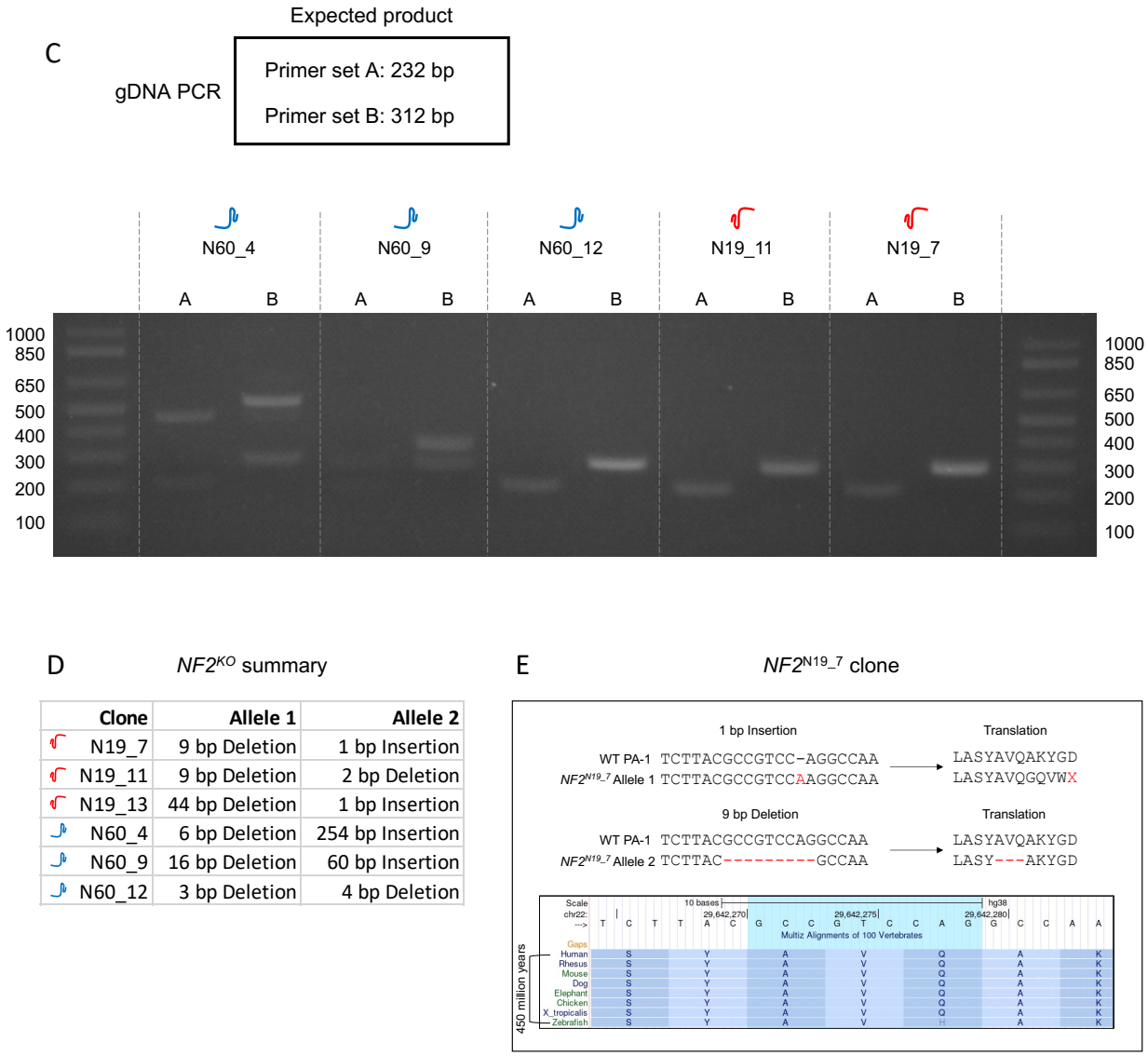
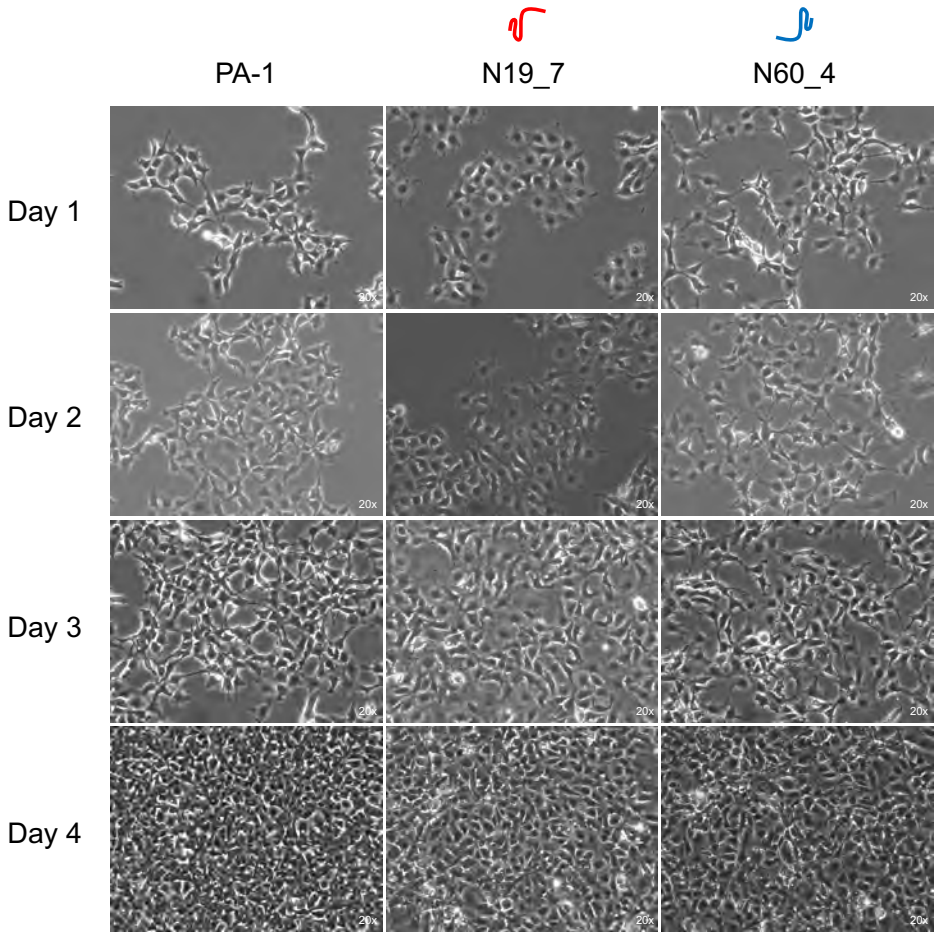


Figure 3.3: Generation of clonal NF2 knockout cell lines.

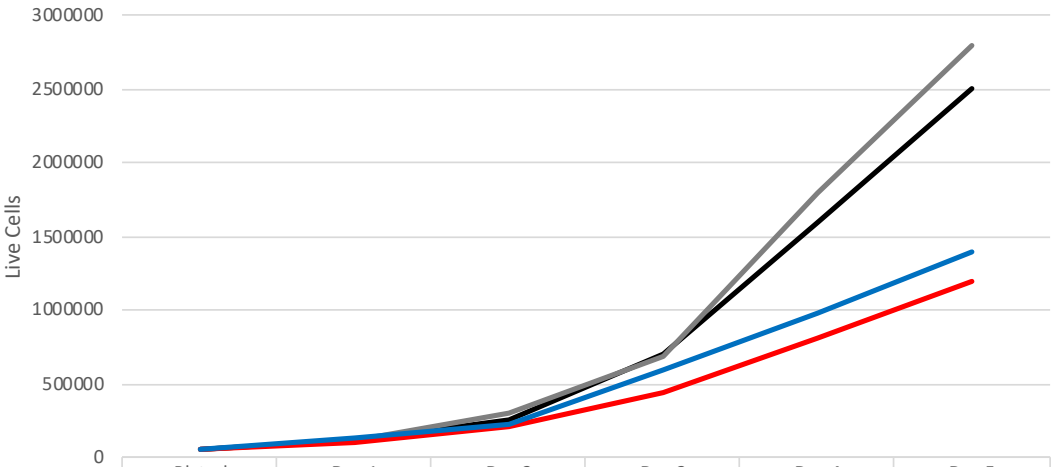
(A) Workflow for generating clonal *NF2* knockout cell lines. PA-1 cells were transfected with PX459-derived vectors expressing *NF2*-targeting sgRNAs and Cas9 to generate *NF2* "population knockout" cells. *NF2* population knockout cells were subjected to dilution cloning to establish clonal cell lines for characterization of NF2/merlin. (B) Western blot showing NF2/merlin protein expression (70 kD band) in established clonal cell lines. Ribosomal S6 protein expression (32 kD band) is shown as a loading control. *NF2*-targeting sgRNAs used to generate clonal cell lines are indicated (N19, ; N60,). (C) PCR assay for genomic edits at the *NF2* sgRNA-target locus. Two primer sets (A and B) flanking the sgRNA-target sites were used to amplify genomic edits. PCR products were run on a 1% agarose gel for visualization. (D) Summary of allelic mutations identified

by Sanger sequencing of TOPO cloned PCR products. (E) Genomic edits within the N19_7 clonal cell line. Allele 1 contains a single nucleotide insertion at the sgRNA-target site, resulting in a frameshift mutation. Allele 2 contains a 9-nucleotide deletion, resulting in an in-frame deletion. Sequence conservation analysis (UCSC genome browser) of the deleted nucleotides within allele 2 (bottom panel). The AVQ residues at the sgRNA-target site are conserved back to Zebrafish, or ~450 million years.

A



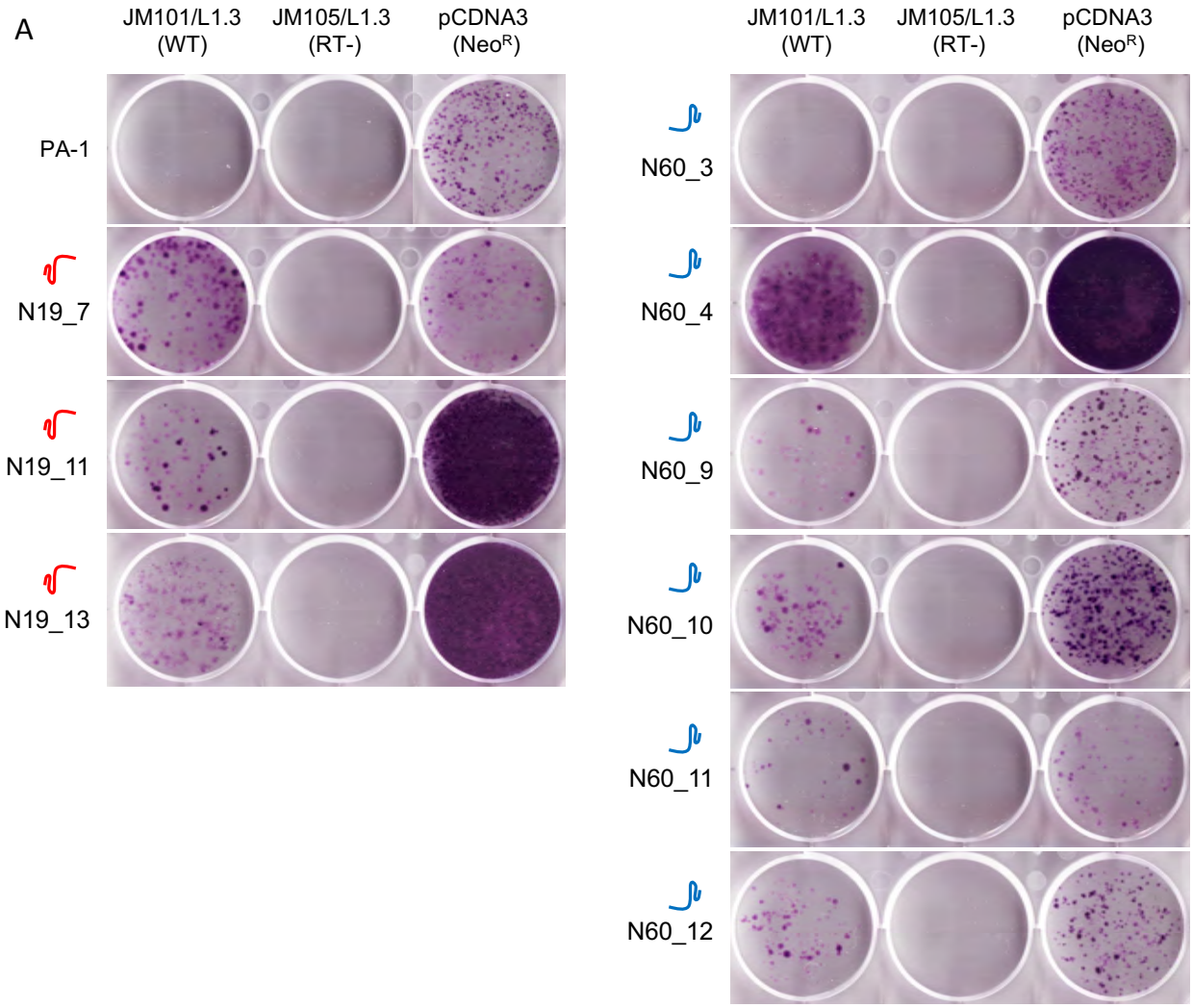
B



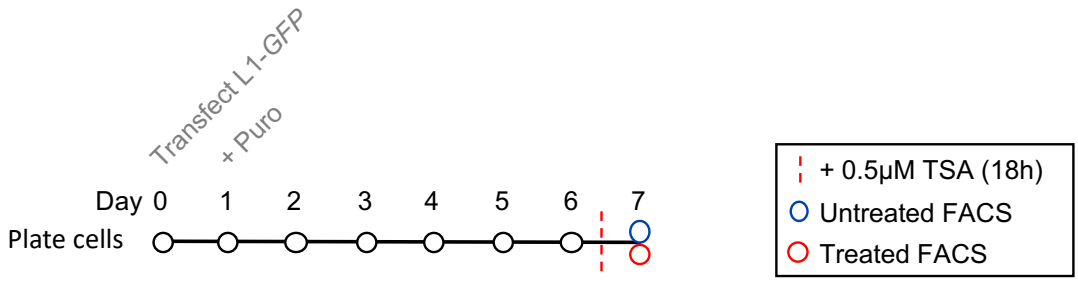
	Plated	Day 1	Day 2	Day 3	Day 4	Day 5
PA-1	50000	120000	250000	700000	1600000	2500000
N60_3	50000	120000	300000	680000	1800000	2800000
N19_7	50000	110000	210000	440000	810000	1200000
N60_4	50000	130000	230000	590000	980000	1400000

Figure 3.4: NF2 knockout cells do not exhibit increased proliferation.

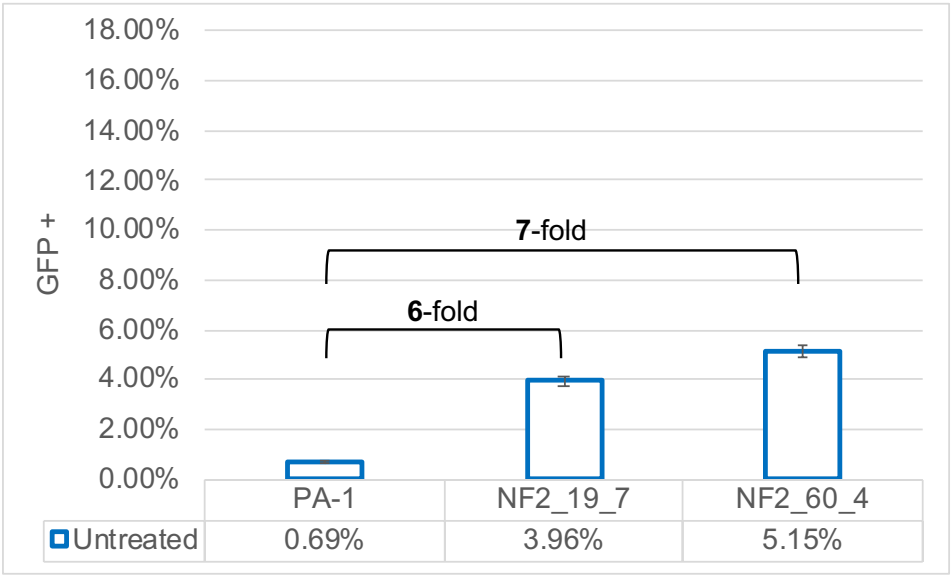
(A) NF2 knockout cell proliferation assay. Cells were plated at $5x10^4$ cells/well of a 6-well plate. Bright-field images were taken daily at 20x magnification. (B) Quantification of the *NF2* knockout cell proliferation assay. Cells were collected and counted daily using the Countess hemocytometer. Notably, N60_3 cells contain wild-type *NF2* alleles. All cell counts achieved >95% viability as determined by trypan blue staining.


A
JM101/L1.3 (WT)
JM105/L1.3 (RT-)
pCDNA3 (Neo^R)
PA-1
N19_7
N19_11
N19_13
JM101/L1.3 (WT)
JM105/L1.3 (RT-)
pCDNA3 (Neo^R)
N60_3
N60_4
N60_9
N60_10
N60_11
N60_12

B


Transfect L1-GFP
+ Puro
Day 0 1 2 3 4 5 6 7
Plate cells
+ 0.5μM TSA (18h)
Untreated FACS
Treated FACS

C


18.00%
16.00%
14.00%
12.00%
10.00%
8.00%
6.00%
4.00%
2.00%
0.00%
GFP +
7-fold
6-fold
PA-1
NF2_19_7
NF2_60_4
Untreated
0.69%
3.96%
5.15%

D

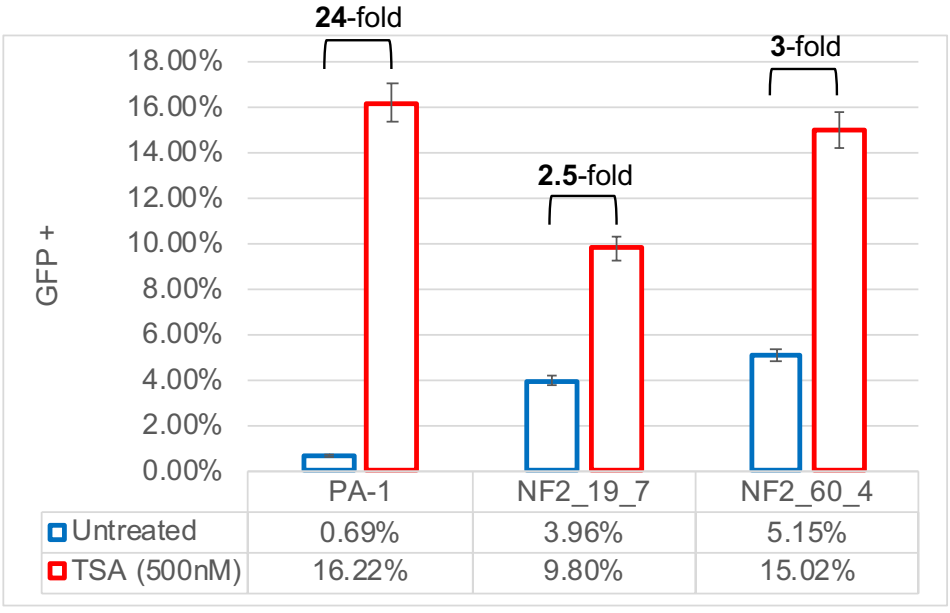

24-fold
3-fold
2.5-fold
18.00%
16.00%
14.00%
12.00%
10.00%
8.00%
6.00%
4.00%
2.00%
0.00%
GFP +
PA-1
NF2_19_7
NF2_60_4
Untreated
0.69%
3.96%
5.15%
TSA (500nM)
16.22%
9.80%
15.02%

Figure 3.5: NF2/merlin is necessary for efficient L1-REPEL.

(A) L1 retrotransposition assays in *NF2* knockout clones. Results of the L1-*neo* and L1-*blast* retrotransposition assays in clonal *NF2* knockout cell lines. Notably, N60_3 cells contain wild-type *NF2* alleles. pJM101/L1.3 and pJJ101/L1.3, retrotransposition-competent L1s; pJM105/L1.3 and pJJ105/L1.3, retrotransposition-defective L1s; pCDNA3, a vector that constitutively expresses the neomycin resistance gene; pCDNA6, a vector that constitutively expresses the blasticidin resistance gene. Drug-resistant colonies were fixed and stained with crystal violet for visualization. (B) Schematic of the L1-*GFP* assay. 6 days post-transfection, puromycin-resistant cells are treated with TSA or left untreated, then subjected to flow cytometry 18 hours later. Puromycin was used to select for the episomal plasmid. (C) Results of the L1-*GFP* retrotransposition assay in untreated cells transfected with p99EGFP/LRE3, a retrotransposition competent L1. The percentage of EGFP expressing cells (GFP +) and standard deviation (n=3) is indicated. (D) Results of the L1-*GFP* retrotransposition assay in untreated and TSA-treated cells transfected with p99EGFP/LRE3. The percentage of EGFP expressing cells (GFP +) and standard deviation (n=3) is indicated. Fold-change indicates the L1-REPEL efficiency.

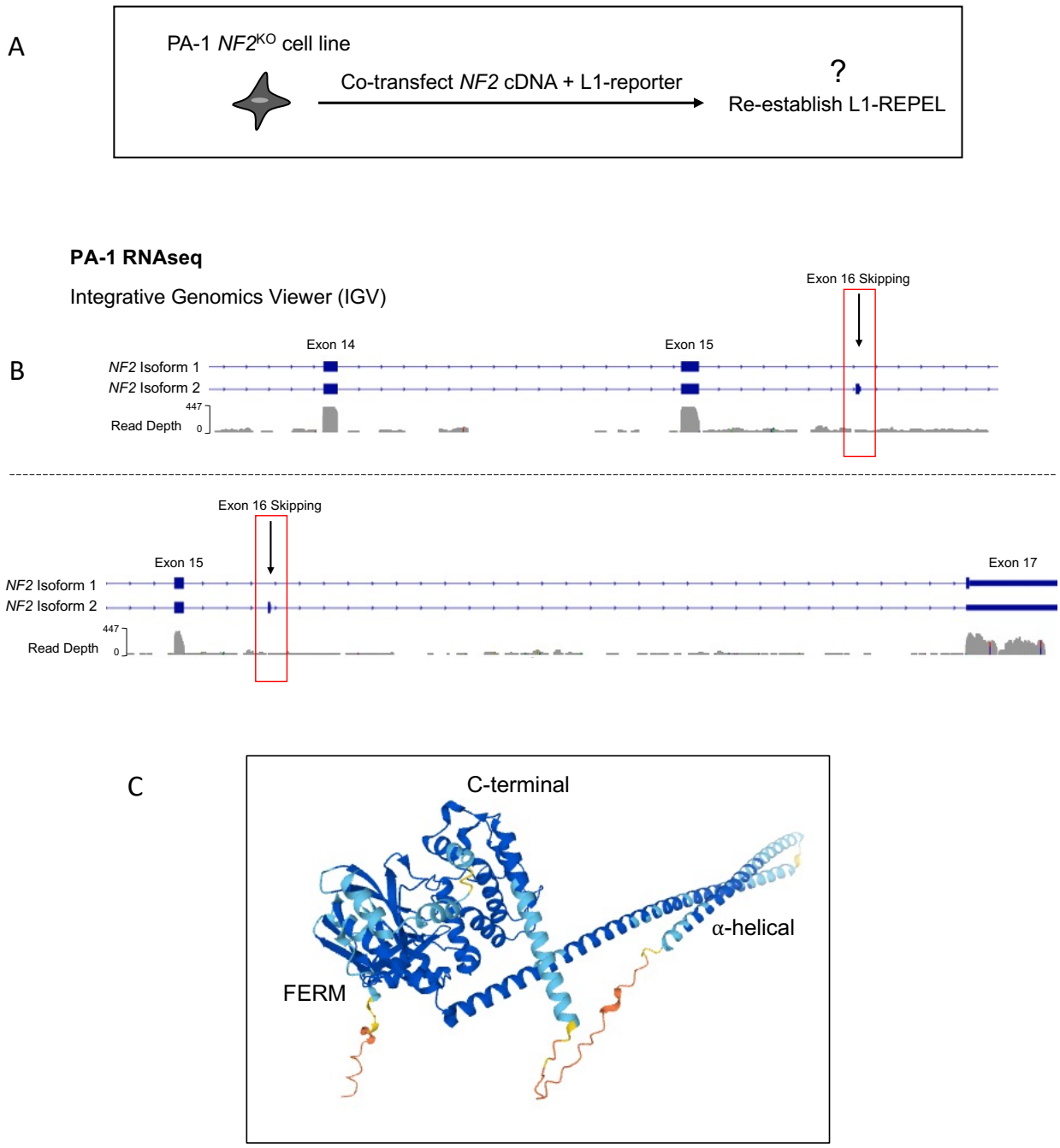


Figure 3.6: NF2/merlin isoform 1 is predominantly expressed in PA-1 cells.

(A) *NF2* rescue assay schematic. *NF2* knockout cells are co-transfected with *NF2* cDNA and L1-*neo* (pJM101/L1.3) or L1-*blast* (pJJ101/L1.3). (B) PA-1 RNA-seq datasets were

viewed using the Integrative Genomics Viewer (IGV). *NF2* isoforms are represented by NCBI RefSeq gene tracks (blue tracks). The bottom track illustrates read depth. Notably, *NF2* isoform 1 lacks exon 16, whereas isoform 2 retains exon 16. The lack of read depth at exon 16 suggests exon skipping in the majority of mature RNA (red box). (C) AlphaFold predicted structure of NF2/merlin isoform 1 exhibiting intramolecular association between the N-terminal FERM domain and the C-terminal domain.

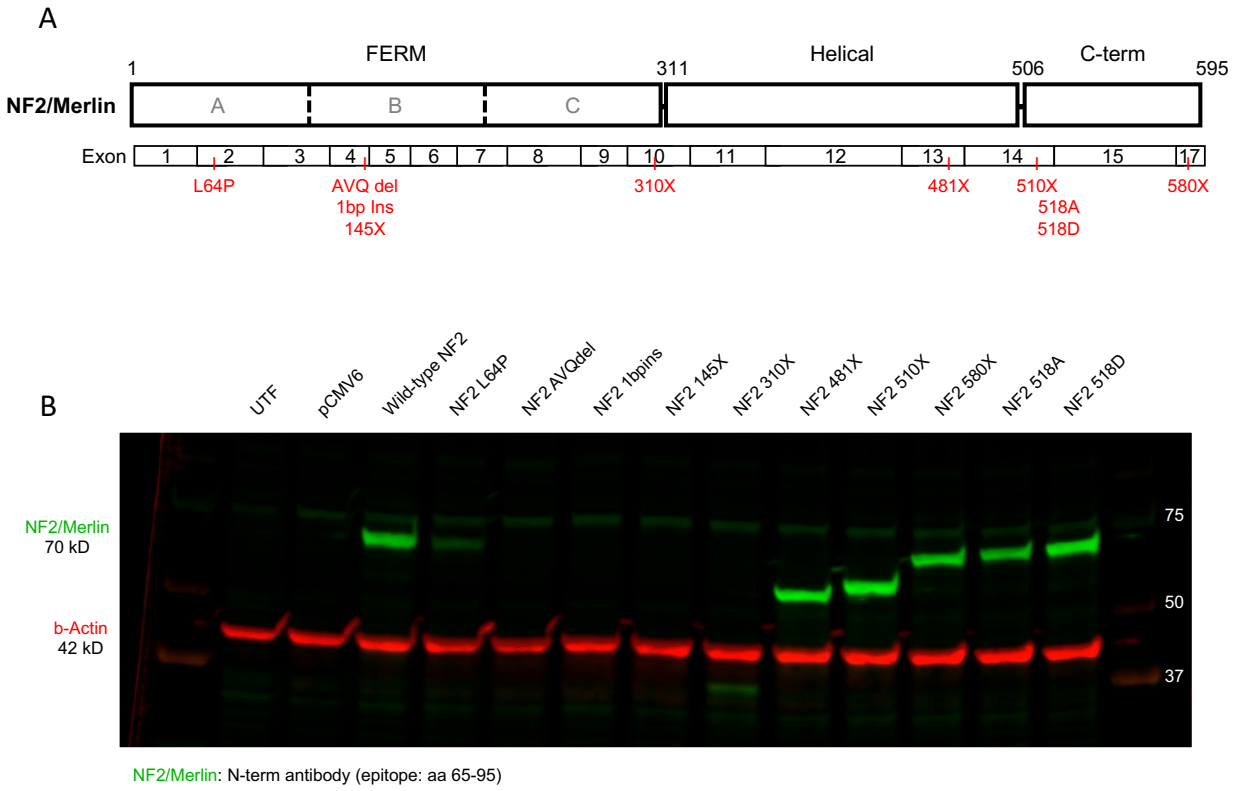
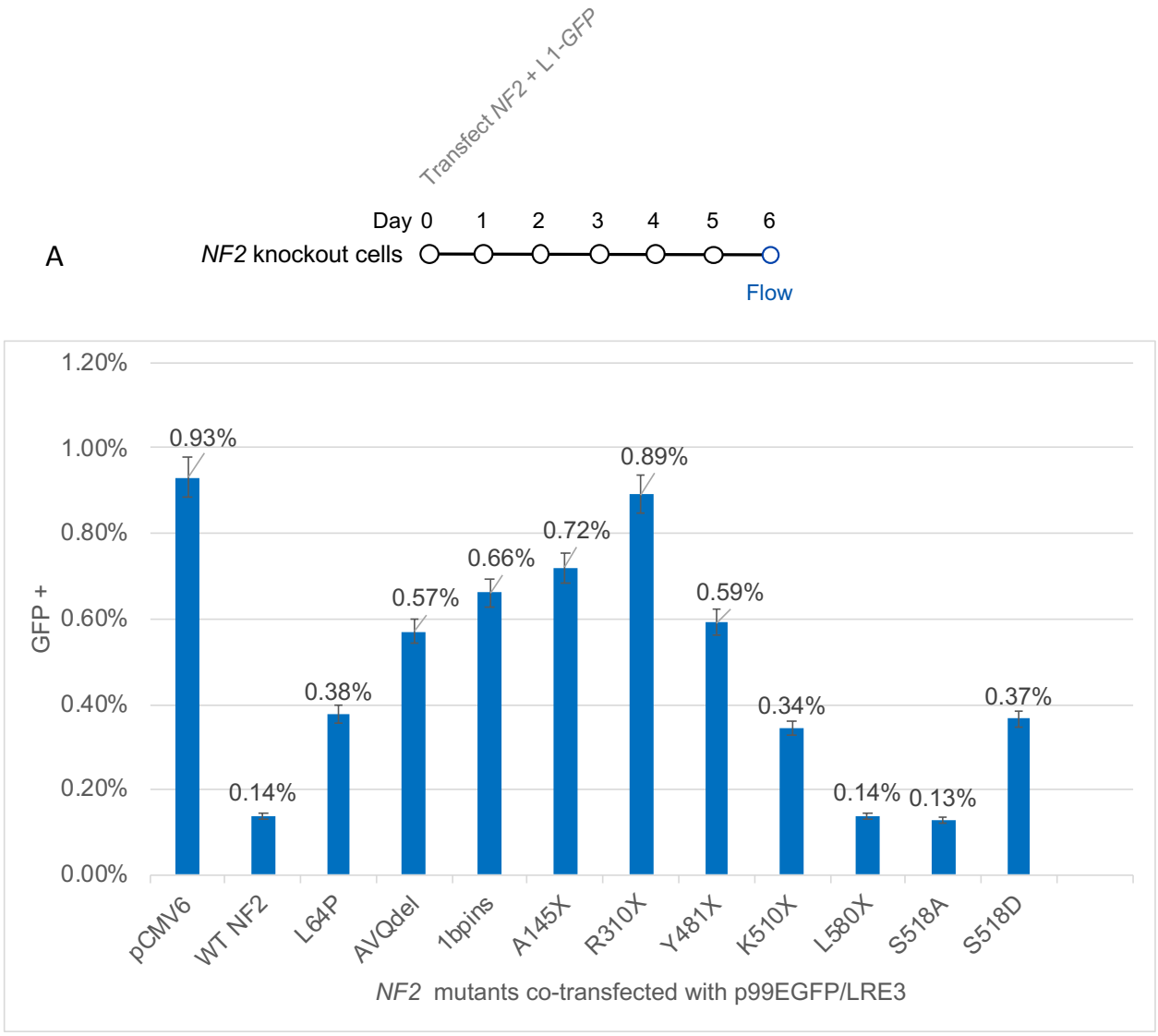
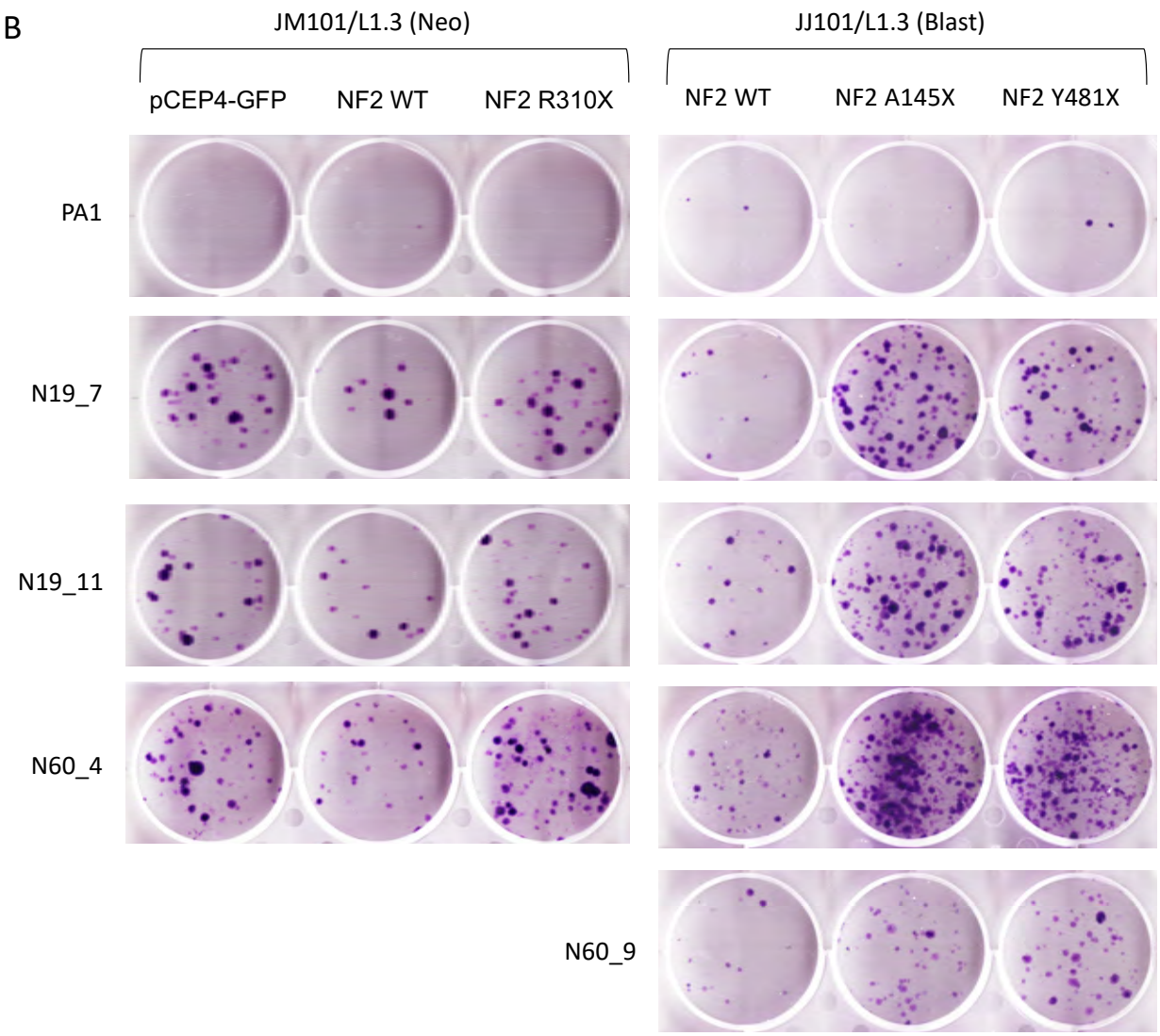


Figure 3.7: Generation of NF2/merlin mutants.

(A) Schematic illustrating site-directed mutations within NF2/merlin. (B) Western blot showing NF2/merlin protein expression (70 kD band) in N19_7 cells transfected with the indicated *NF2* cDNAs. Notably, an antibody raised against the N-terminus (residues 65-95) was used to detect NF2/merlin protein expression. b-actin protein expression (42 kD band) is shown as a loading control.


A
Transfect NF2 + L1-GFP
Day 0 1 2 3 4 5 6
NF2 knockout cells
Flow
1.20%
1.00%
0.80%
0.60%
0.40%
0.20%
0.00%
GFP +
0.93%
0.14%
0.38%
0.57%
0.66%
0.72%
0.89%
0.59%
0.34%
0.14%
0.13%
0.37%
pCMV6
WT NF2
L64P
AVQdel
1bpins
A145X
R310X
Y481X
K510X
L580X
S518A
S518D
NF2 mutants co-transfected with p99EGFP/LRE3


B
JM101/L1.3 (Neo)
JJ101/L1.3 (Blast)
pCEP4-GFP
NF2 WT
NF2 R310X
NF2 WT
NF2 A145X
NF2 Y481X
PA1
N19_7
N19_11
N60_4
N60_9

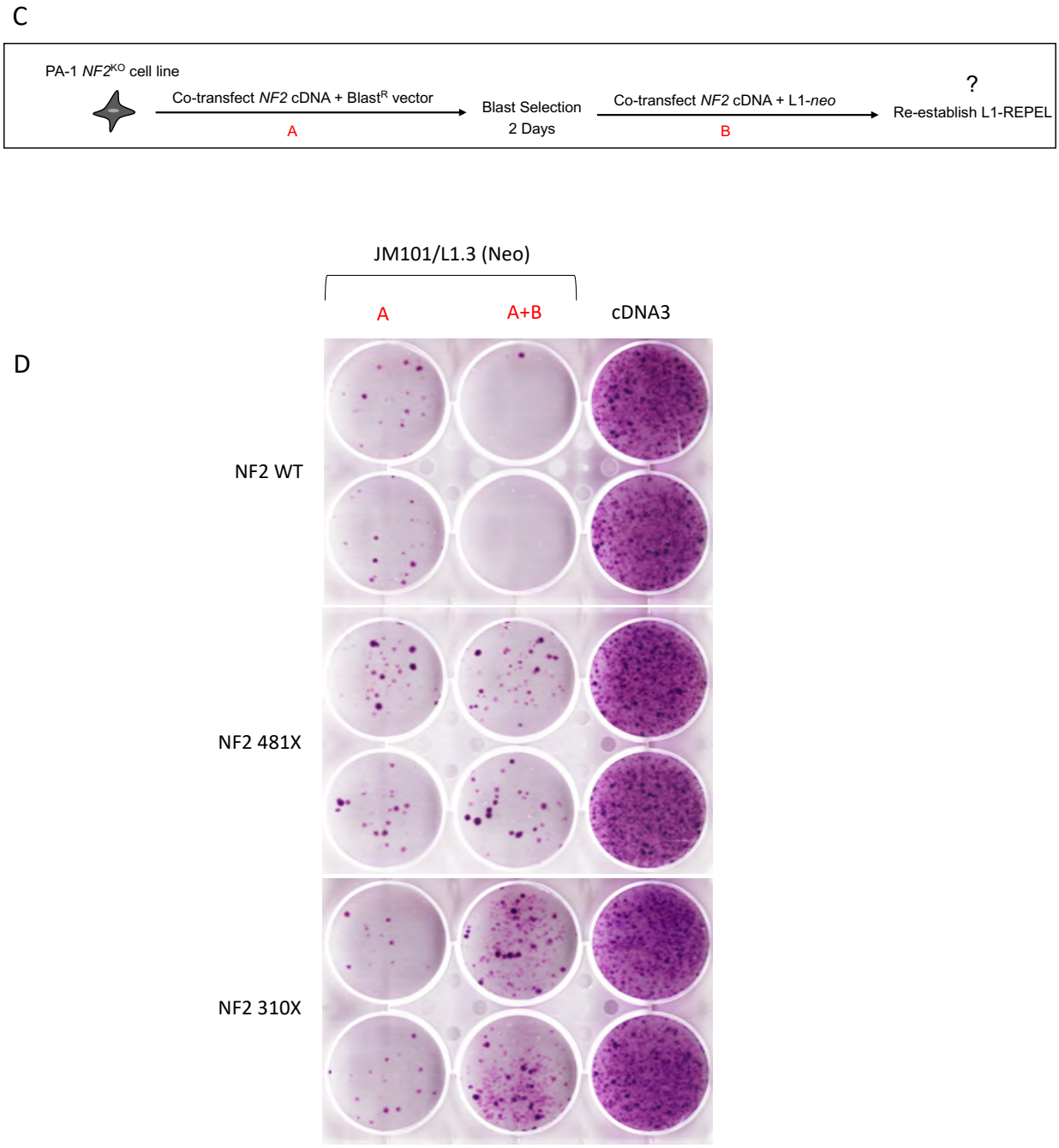


Figure 3.8: NF2/merlin expression rescues L1-REPEL in NF2 knockout cells.

(A) Top: Schematic of the *NF2* rescue L1-*GFP* assay. *NF2* knockout cells are co-transfected with *NF2* cDNA expression vectors and p99EGFP/LRE3. 6 days post-transfection, cells are subjected to flow cytometry. Bottom: results of the L1-*GFP* NF2 rescue assay in N19_7 *NF2* knockout cells. The percentage of EGFP expressing cells

(GFP +) and standard deviation (error bars, n=3) is indicated. (B) Results of the *NF2* rescue L1-*neo* and L1-*blast* retrotransposition assays in clonal *NF2* knockout cell lines. Notably, N60_3 cells contain wild-type *NF2* alleles. pJM101/L1.3 (L1-*neo*) and pJJ101/L1.3 (L1-*blast*), retrotransposition-competent L1s. Drug-resistant colonies were fixed and stained with crystal violet for visualization. (C) Schematic of the serial *NF2* rescue assay. *NF2* knockout cells are co-transfected with *NF2* cDNA expression vectors and pCDNA6 (A), selected, then co-transfected with *NF2* cDNA expression vectors and pJM101/L1.3 or pCDNA3 (B) and subjected to the L1-*neo* assay. (D) Results of the *NF2* rescue L1-*neo* retrotransposition assay in N19_7 *NF2* knockout cells. pJM101/L1.3 (L1-*neo*), retrotransposition-competent L1; pCDNA3, a vector that expresses neomycin resistance. G418-resistant colonies were fixed and stained with crystal violet for visualization.

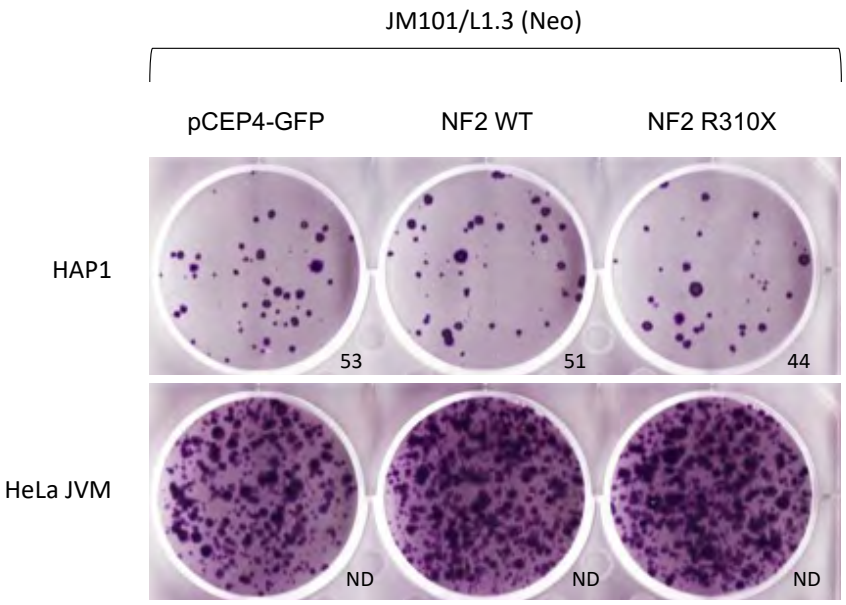
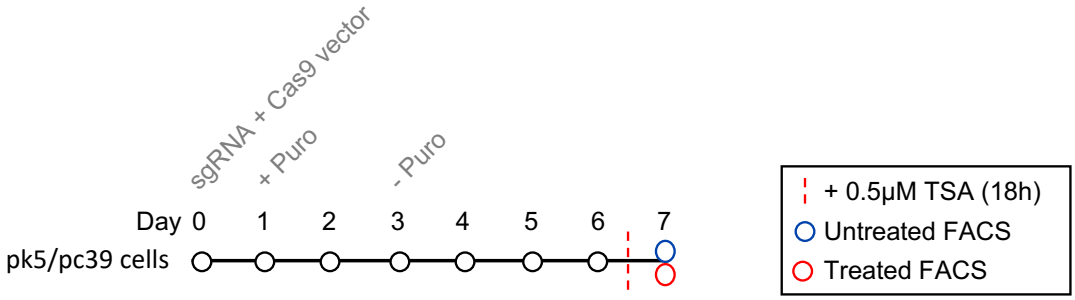


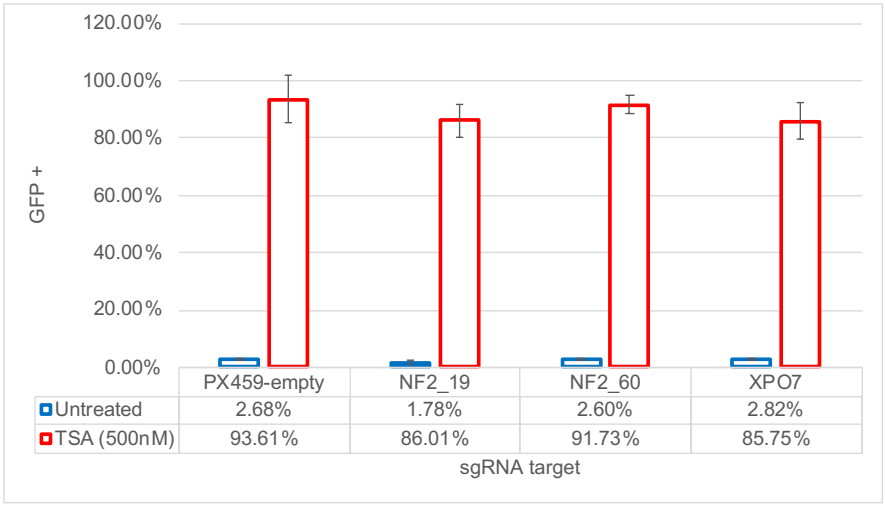
Figure 3.9: NF2/merlin expression does not hinder L1 retrotransposition in somatic cells.

Results of the *NF2* rescue L1-*neo* retrotransposition assays in somatic HeLa JVM and HAP1 cells. Cells were co-transfected with pJM101/L1.3 and *NF2* cDNA expression vectors or pCEP-GFP. pJM101/L1.3 (L1-*neo*), retrotransposition-competent L1. pCEP4-GFP serves as a control expression vector. G418-resistant colonies were fixed and stained with crystal violet for visualization.

A


sgRNA + Cas9 vector
+ Puro
- Puro
Day 0 1 2 3 4 5 6 7
pk5/pc39 cells
+ 0.5μM TSA (18h)
Untreated FACS
Treated FACS

B


pk5
120.00%
100.00%
80.00%
60.00%
40.00%
20.00%
0.00%
GFP +
PX459-empty NF2_19 NF2_60 XPO7
Untreated 2.68% 1.78% 2.60% 2.82%
TSA (500nM) 93.61% 86.01% 91.73% 85.75%
sgRNA target

C

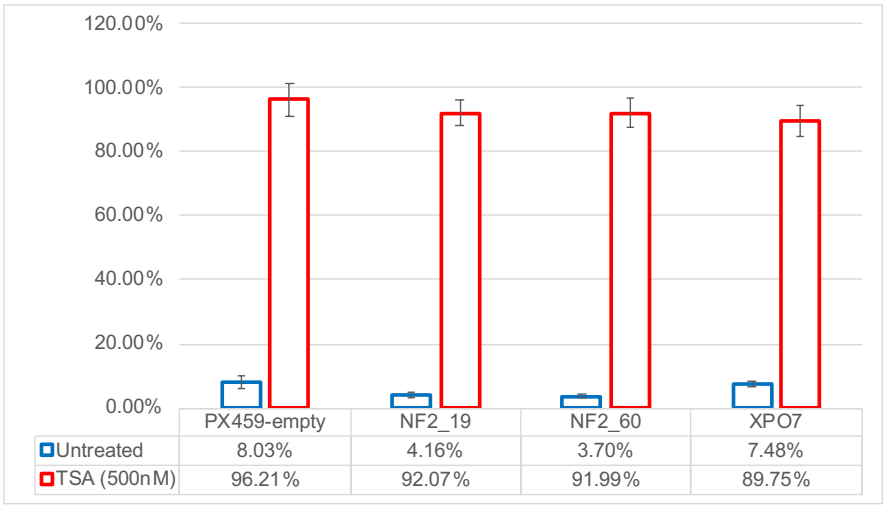
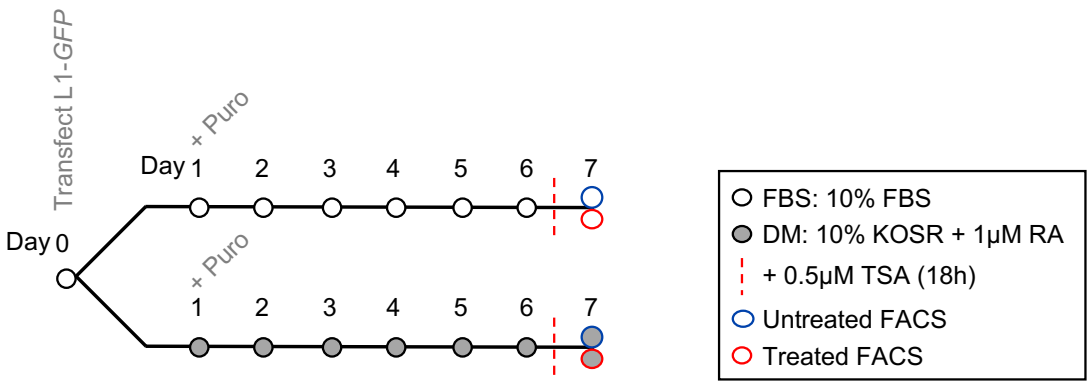

pc39
120.00%
100.00%
80.00%
60.00%
40.00%
20.00%
0.00%
PX459-empty NF2_19 NF2_60 XPO7
Untreated 8.03% 4.16% 3.70% 7.48%
TSA (500nM) 96.21% 92.07% 91.99% 89.75%

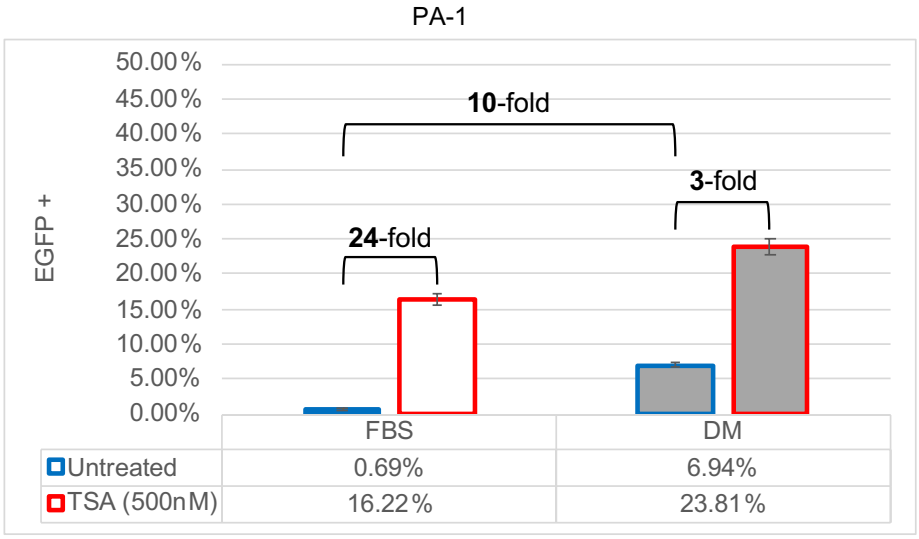
Figure 3.10: NF2 knockout is not sufficient to reverse L1-REPEL.

(A) Schematic of the L1-REPEL maintenance assay. pk5 or pc39 are transfected with *NF2* (N19, N60) or XPO7 knockout vectors. 6 days post-transfection, puromycin-resistant pk5 or pc39 cells are treated with TSA or left untreated, then subjected to flow cytometry 18 hours later. Results of the L1-REPEL maintenance assay in pk5 (B) or pc39 (C) cells. The percentage of EGFP expressing cells (GFP +) and standard deviation (error bars, n=3) is indicated.

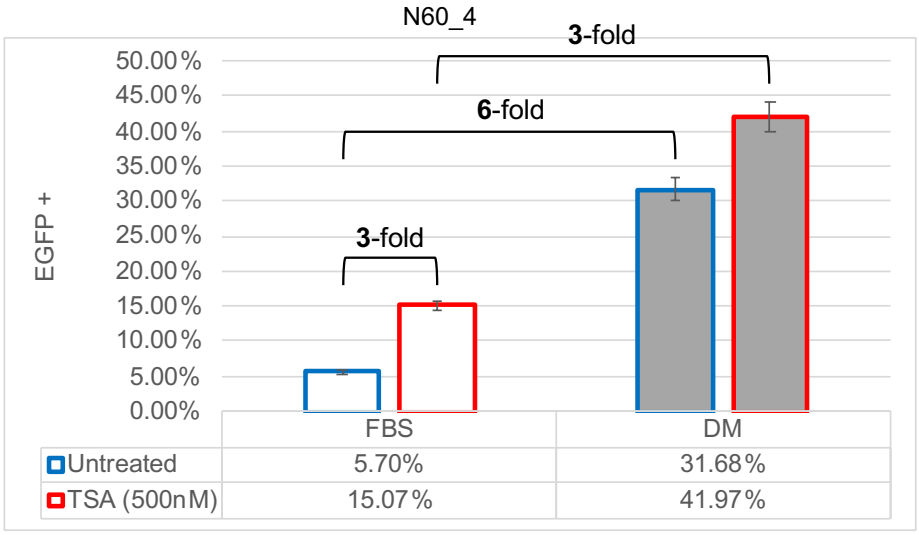
A


Transfect L1-GFP
+ Puro
Day 0
Day 1
2
3
4
5
6
7
+ Puro
1
2
3
4
5
6
7
FBS: 10% FBS
DM: 10% KOSR + 1μM RA
+ 0.5μM TSA (18h)
Untreated FACS
Treated FACS

B


PA-1
EGFP +
50.00%
45.00%
40.00%
35.00%
30.00%
25.00%
20.00%
15.00%
10.00%
5.00%
0.00%
10-fold
3-fold
24-fold
FBS
DM
Untreated 0.69% 6.94%
TSA (500nM) 16.22% 23.81%

C


N60_4
EGFP +
50.00%
45.00%
40.00%
35.00%
30.00%
25.00%
20.00%
15.00%
10.00%
5.00%
0.00%
3-fold
6-fold
3-fold
FBS
DM
Untreated 5.70% 31.68%
TSA (500nM) 15.07% 41.97%

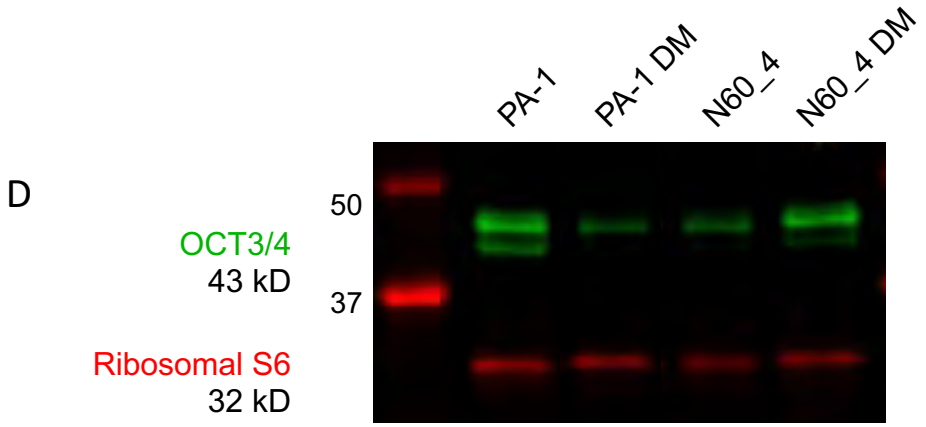


Figure 3.11: L1-REPEL is attenuated in differentiation media.

(A) Schematic of the L1-REPEL L1-*GFP* retrotransposition assays in differentiation media. 24 hours after transfection, cells are cultured in either 10% FBS or Differentiation media (DM) containing 10% KnockOut Serum Replacement (KOSR) plus 1μM all-trans retinoic acid for the duration of the L1-*GFP* assay. 6 days post-transfection, puromycin-resistant cells are treated with TSA or left untreated, then subjected to flow cytometry 18 hours later. Puromycin was used to select for the episomal plasmid. Results of the L1-REPEL L1-*GFP* retrotransposition assays in wild-type PA-1 (B) or N60_4 *NF2* knockout (C) cells. White filled bars indicate samples cultured in 10% FBS. Grey filled bars indicate samples cultured in DM. (D) Western blot showing protein expression of the pluripotency marker OCT3/4 (~43 kD band) in wild-type PA-1 or N60_4 *NF2* knockout cells. Ribosomal S6 protein expression (32 kD band) is shown as a loading control. DM indicates that cells were cultured in differentiation media.

Table 3.1: Candidate gene sgRNA oligonucleotides.

sgRNA	Forward oligo (5'-3')	Reverse oligo (5'-3')
*NF2*_sgRNA_1 (N19)	CACCGTGAGCCTACCTTGGCCTGGA	AAACTCCAGGCCAAGGTAGGCTCAC
*NF2*_sgRNA_2 (N60)	CACCGCCTGGCTTCTTACGCCGTCC	AAACGGACGGCGTAAGAAGCCAGGC
*NF2*_sgRNA_3 (N18)	CACCGATTCCACGGGAAGGAGATCT	AAACAGATCTCCTTCCCGTGGAATC
XPO7	CACCGAGACACAACCACTCGACTCC	AAACGGAGTCGAGTGGTTGTGTCTC

Table 3.2: Site-directed mutagenesis primers.

mutant	mutant primer sequence	wild-type sequence
L64P	aacctggttctttggaccgcagtacacaatcaagg	aacctggttctttggact**t**gcagtacacaatcaagg
1bpins	cttcttacgccgtcctaggccaagtatggtg	cttcttacgccgtccaggccaagtatggtg
AVQdel	gctcctggcttcttacgccaagtatggtgact	gctcctggcttcttacgcc**gtccaggcc**aagtatggtgact
A145X	gtgctcctggcttcttactgagtccaggccaagtatgg	gtgctcctggcttcttac**gcc**gtccaggccaagtatgg
R310X	accatgatctatttatgaggtgaaggaaagccgattctttg	accatgatctatttatgagg**a**gaaggaaagccgattctttg
Y481XX	gattgccaccaagcccacgtagtagcccatgaacccaatt	gattgccaccaagcccacgta**ccc**gcccatgaacccaatt
K510X	cagcctgtctttcgacttctgagatactgacatgaagcgg	cagcctgtctttcgacttc**aaa**gatactgacatgaagcgg
L580X	gcagcaagcacaataccattaaaaagtgaaccttgcagagcgccaag	gcagcaagcacaataccattaaaaag**ctc**accttgcagagcgccaag
S518A	actgacatgaagcggcttgccatggagatagaga	actgacatgaagcggctt**tcc**atggagatagaga
S518D	aagatactgacatgaagcggcttgacatggagatagagaaagaaaaag	aagatactgacatgaagcggctt**tcc**atggagatagagaaagaaaaag

References

Abu Dawud, R., Schreiber, K., Schomburg, D., and Adjaye, J. (2012). Human embryonic stem cells and embryonal carcinoma cells have overlapping and distinct metabolic signatures. PLoS One *7*, e39896.

Alfthan, K., Heiska, L., Gronholm, M., Renkema, G.H., and Carpen, O. (2004). Cyclic AMP-dependent protein kinase phosphorylates merlin at serine 518 independently of p21-activated kinase and promotes merlin-ezrin heterodimerization. J Biol Chem *279*, 18559-18566.

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

Beck, C.R., Collier, P., Macfarlane, C., Malig, M., Kidd, J.M., Eichler, E.E., Badge, R.M., and Moran, J.V. (2010). LINE-1 retrotransposition activity in human genomes. Cell *141*, 1159-1170.

Beck, C.R., Garcia-Perez, J.L., Badge, R.M., and Moran, J.V. (2011). LINE-1 elements in structural variation and disease. Annu Rev Genomics Hum Genet *12*, 187-215.

Bianchi, A.B., Hara, T., Ramesh, V., Gao, J., Klein-Szanto, A.J., Morin, F., Menon, A.G., Trofatter, J.A., Gusella, J.F., Seizinger, B.R.*, et al.* (1994). Mutations in transcript isoforms of the neurofibromatosis 2 gene in multiple human tumour types. Nat Genet *6*, 185-192.

Brizzi, M.F., Tarone, G., and Defilippi, P. (2012). Extracellular matrix, integrins, and growth factors as tailors of the stem cell niche. Curr Opin Cell Biol *24*, 645-651.

Brouha, B., Schustak, J., Badge, R.M., Lutz-Prigge, S., Farley, A.H., Moran, J.V., and Kazazian, H.H., Jr. (2003). Hot L1s account for the bulk of retrotransposition in the human population. Proc Natl Acad Sci U S A *100*, 5280-5285.

Cooper, J., and Giancotti, F.G. (2014). Molecular insights into NF2/Merlin tumor suppressor function. FEBS Lett *588*, 2743-2752.

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

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

DeMali, K.A., Sun, X., and Bui, G.A. (2014). Force transmission at cell-cell and cell-matrix adhesions. Biochemistry *53*, 7706-7717.

Ergun, S., Buschmann, C., Heukeshoven, J., Dammann, K., Schnieders, F., Lauke, H., Chalajour, F., Kilic, N., Stratling, W.H., and Schumann, G.G. (2004). Cell type-specific expression of LINE-1 open reading frames 1 and 2 in fetal and adult human tissues. J Biol Chem *279*, 27753-27763.

Faulkner, G.J., and Billon, V. (2018). L1 retrotransposition in the soma: a field jumping ahead. Mob DNA *9*, 22.

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

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

Feschotte, C., and Pritham, E.J. (2007). DNA transposons and the evolution of eukaryotic genomes. Annu Rev Genet *41*, 331-368.

Flasch, D.A., Macia, A., Sanchez, L., Ljungman, M., Heras, S.R., Garcia-Perez, J.L., Wilson, T.E., and Moran, J.V. (2019). Genome-wide de novo L1 Retrotransposition Connects Endonuclease Activity with Replication. Cell *177*, 837-851 e828.

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

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

Golovnina, K., Blinov, A., Akhmametyeva, E.M., Omelyanchuk, L.V., and Chang, L.S. (2005). Evolution and origin of merlin, the product of the Neurofibromatosis type 2 (NF2) tumor-suppressor gene. BMC Evol Biol *5*, 69.

Gonzalez-Agosti, C., Wiederhold, T., Herndon, M.E., Gusella, J., and Ramesh, V. (1999). Interdomain interaction of merlin isoforms and its influence on intermolecular binding to NHE-RF. J Biol Chem *274*, 34438-34442.

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

Gronholm, M., Sainio, M., Zhao, F., Heiska, L., Vaheri, A., and Carpen, O. (1999). Homotypic and heterotypic interaction of the neurofibromatosis 2 tumor suppressor protein merlin and the ERM protein ezrin. J Cell Sci *112 (Pt 6)*, 895-904.

Gutmann, D.H., Geist, R.T., Xu, H., Kim, J.S., and Saporito-Irwin, S. (1998). Defects in neurofibromatosis 2 protein function can arise at multiple levels. Hum Mol Genet *7*, 335-345.

Gutmann, D.H., Wright, D.E., Geist, R.T., and Snider, W.D. (1995). Expression of the neurofibromatosis 2 (NF2) gene isoforms during rat embryonic development. Hum Mol Genet *4*, 471-478.

Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of cancer: the next generation. Cell *144*, 646-674.

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

Holmes, S.E., Singer, M.F., and Swergold, G.D. (1992). Studies on p40, the leucine zipper motif-containing protein encoded by the first open reading frame of an active human LINE-1 transposable element. J Biol Chem *267*, 19765-19768.

Hong, A.W., Meng, Z., Plouffe, S.W., Lin, Z., Zhang, M., and Guan, K.L. (2020). Critical roles of phosphoinositides and NF2 in Hippo pathway regulation. Genes Dev *34*, 511-525.

Jin, H., Sperka, T., Herrlich, P., and Morrison, H. (2006). Tumorigenic transformation by CPI-17 through inhibition of a merlin phosphatase. Nature *442*, 576-579.

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

Kazazian, H.H., Jr. (2004). Mobile elements: drivers of genome evolution. Science *303*, 1626-1632.

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

Kazazian, H.H., Jr., Wong, C., Youssoufian, H., Scott, A.F., Phillips, D.G., and Antonarakis, S.E. (1988). Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. Nature *332*, 164-166.

Kim, H.A., DeClue, J.E., and Ratner, N. (1997). cAMP-dependent protein kinase A is required for Schwann cell growth: interactions between the cAMP and neuregulin/tyrosine kinase pathways. J Neurosci Res *49*, 236-247.

Kim, K., Heo, K., Choi, J., Jackson, S., Kim, H., Xiong, Y., and An, W. (2012). Vpr-binding protein antagonizes p53-mediated transcription via direct interaction with H3 tail. Mol Cell Biol *32*, 783-796.

Kim, K., Kim, J.M., Kim, J.S., Choi, J., Lee, Y.S., Neamati, N., Song, J.S., Heo, K., and An, W. (2013). VprBP has intrinsic kinase activity targeting histone H2A and represses gene transcription. Mol Cell *52*, 459-467.

Kopera, H.C., Larson, P.A., Moldovan, J.B., Richardson, S.R., Liu, Y., and Moran, J.V. (2016). LINE-1 Cultured Cell Retrotransposition Assay. Methods Mol Biol *1400*, 139-156.

Kopera, H.C., Moldovan, J.B., Morrish, T.A., Garcia-Perez, J.L., and Moran, J.V. (2011). Similarities between long interspersed element-1 (LINE-1) reverse transcriptase and telomerase. Proc Natl Acad Sci U S A *108*, 20345-20350.

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

Kulpa, D.A., and Moran, J.V. (2005). Ribonucleoprotein particle formation is necessary but not sufficient for LINE-1 retrotransposition. Hum Mol Genet *14*, 3237-3248.

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

Lallemand, D., Curto, M., Saotome, I., Giovannini, M., and McClatchey, A.I. (2003). NF2 deficiency promotes tumorigenesis and metastasis by destabilizing adherens junctions. Genes Dev *17*, 1090-1100.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W.*, et al.* (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860-921.

Laulajainen, M., Muranen, T., Carpen, O., and Gronholm, M. (2008). Protein kinase A-mediated phosphorylation of the NF2 tumor suppressor protein merlin at serine 10 affects the actin cytoskeleton. Oncogene *27*, 3233-3243.

Laulajainen, M., Muranen, T., Nyman, T.A., Carpen, O., and Gronholm, M. (2011). Multistep phosphorylation by oncogenic kinases enhances the degradation of the NF2 tumor suppressor merlin. Neoplasia *13*, 643-652.

Leibold, D.M., Swergold, G.D., Singer, M.F., Thayer, R.E., Dombroski, B.A., and Fanning, T.G. (1990). Translation of LINE-1 DNA elements in vitro and in human cells. Proc Natl Acad Sci U S A *87*, 6990-6994.

Lewinski, M.K., and Bushman, F.D. (2005). Retroviral DNA integration--mechanism and consequences. Adv Genet *55*, 147-181.

Li, W., You, L., Cooper, J., Schiavon, G., Pepe-Caprio, A., Zhou, L., Ishii, R., Giovannini, M., Hanemann, C.O., Long, S.B.*, et al.* (2010). Merlin/NF2 suppresses tumorigenesis by inhibiting the E3 ubiquitin ligase CRL4(DCAF1) in the nucleus. Cell *140*, 477-490.

Lima, A.F., May, G., Diaz-Colunga, J., Pedreiro, S., Paiva, A., Ferreira, L., Enver, T., Iborra, F.J., and Pires das Neves, R. (2018). Osmotic modulation of chromatin impacts on efficiency and kinetics of cell fate modulation. Sci Rep *8*, 7210.

Luan, D.D., Korman, M.H., Jakubczak, J.L., and Eickbush, T.H. (1993). Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. Cell *72*, 595-605.

Martin, S.L., Cruceanu, M., Branciforte, D., Wai-Lun Li, P., Kwok, S.C., Hodges, R.S., and Williams, M.C. (2005). LINE-1 retrotransposition requires the nucleic acid chaperone activity of the ORF1 protein. J Mol Biol *348*, 549-561.

McClatchey, A.I., Saotome, I., Ramesh, V., Gusella, J.F., and Jacks, T. (1997). The Nf2 tumor suppressor gene product is essential for extraembryonic development immediately prior to gastrulation. Genes Dev *11*, 1253-1265.

McClatchey, A.I., and Yap, A.S. (2012). Contact inhibition (of proliferation) redux. Curr Opin Cell Biol *24*, 685-694.

Meng, J.J., Lowrie, D.J., Sun, H., Dorsey, E., Pelton, P.D., Bashour, A.M., Groden, J., Ratner, N., and Ip, W. (2000). Interaction between two isoforms of the NF2 tumor suppressor protein, merlin, and between merlin and ezrin, suggests modulation of ERM proteins by merlin. J Neurosci Res *62*, 491-502.

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

Morrish, T.A., Gilbert, N., Myers, J.S., Vincent, B.J., Stamato, T.D., Taccioli, G.E., Batzer, M.A., and Moran, J.V. (2002). DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. Nat Genet *31*, 159-165.

Morrison, H., Sherman, L.S., Legg, J., Banine, F., Isacke, C., Haipek, C.A., Gutmann, D.H., Ponta, H., and Herrlich, P. (2001). The NF2 tumor suppressor gene product, merlin, mediates contact inhibition of growth through interactions with CD44. Genes Dev *15*, 968-980.

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

Nguyen, R., Reczek, D., and Bretscher, A. (2001). Hierarchy of merlin and ezrin N- and C-terminal domain interactions in homo- and heterotypic associations and their

relationship to binding of scaffolding proteins EBP50 and E3KARP. J Biol Chem *276*, 7621-7629.

Okada, T., Lopez-Lago, M., and Giancotti, F.G. (2005). Merlin/NF-2 mediates contact inhibition of growth by suppressing recruitment of Rac to the plasma membrane. J Cell Biol *171*, 361-371.

Ostertag, E.M., Prak, E.T., DeBerardinis, R.J., Moran, J.V., and Kazazian, H.H., Jr. (2000). Determination of L1 retrotransposition kinetics in cultured cells. Nucleic Acids Res *28*, 1418-1423.

Pearson, M.A., Reczek, D., Bretscher, A., and Karplus, P.A. (2000). Structure of the ERM protein moesin reveals the FERM domain fold masked by an extended actin binding tail domain. Cell *101*, 259-270.

Perrimon, N., Pitsouli, C., and Shilo, B.Z. (2012). Signaling mechanisms controlling cell fate and embryonic patterning. Cold Spring Harb Perspect Biol *4*, a005975.

Petrilli, A.M., and Fernandez-Valle, C. (2016). Role of Merlin/NF2 inactivation in tumor biology. Oncogene *35*, 537-548.

Ran, F.A., Hsu, P.D., Lin, C.Y., Gootenberg, J.S., Konermann, S., Trevino, A.E., Scott, D.A., Inoue, A., Matoba, S., Zhang, Y*., et al.* (2013a). Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. Cell *154*, 1380-1389.

Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013b). Genome engineering using the CRISPR-Cas9 system. Nat Protoc *8*, 2281-2308.

Richardson, S.R., Doucet, A.J., Kopera, H.C., Moldovan, J.B., Garcia-Perez, J.L., and Moran, J.V. (2015). The Influence of LINE-1 and SINE Retrotransposons on Mammalian Genomes. Microbiol Spectr *3*, MDNA3-0061-2014.

Richardson, S.R., Gerdes, P., Gerhardt, D.J., Sanchez-Luque, F.J., Bodea, G.O., Munoz-Lopez, M., Jesuadian, J.S., Kempen, M.H.C., Carreira, P.E., Jeddeloh, J.A*., et al.* (2017). Heritable L1 retrotransposition in the mouse primordial germline and early embryo. Genome Res *27*, 1395-1405.

Rong, R., Surace, E.I., Haipek, C.A., Gutmann, D.H., and Ye, K. (2004). Serine 518 phosphorylation modulates merlin intramolecular association and binding to critical effectors important for NF2 growth suppression. Oncogene *23*, 8447-8454.

Rouleau, G.A., Merel, P., Lutchman, M., Sanson, M., Zucman, J., Marineau, C., Hoang-Xuan, K., Demczuk, S., Desmaze, C., Plougastel, B*., et al.* (1993). Alteration in a new gene encoding a putative membrane-organizing protein causes neuro-fibromatosis type 2. Nature *363*, 515-521.

Sarraf, S., Tejada, R., Abawi, M., Oberst, M., Dennis, T., Simon, K.C., and Blancato, J. (2005). The human ovarian teratocarcinoma cell line PA-1 demonstrates a single

translocation: analysis with fluorescence in situ hybridization, spectral karyotyping, and bacterial artificial chromosome microarray. Cancer Genet Cytogenet *161*, 63-69.

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

Schorn, A.J., Gutbrod, M.J., LeBlanc, C., and Martienssen, R. (2017). LTR-Retrotransposon Control by tRNA-Derived Small RNAs. Cell *170*, 61-71 e11.

Scott, E.C., and Devine, S.E. (2017). The Role of Somatic L1 Retrotransposition in Human Cancers. Viruses *9*.

Shaw, R.J., McClatchey, A.I., and Jacks, T. (1998). Regulation of the neurofibromatosis type 2 tumor suppressor protein, merlin, by adhesion and growth arrest stimuli. J Biol Chem *273*, 7757-7764.

Shaw, R.J., Paez, J.G., Curto, M., Yaktine, A., Pruitt, W.M., Saotome, I., O'Bryan, J.P., Gupta, V., Ratner, N., Der, C.J.*, et al.* (2001). The Nf2 tumor suppressor, merlin, functions in Rac-dependent signaling. Dev Cell *1*, 63-72.

Sherman, L., Xu, H.M., Geist, R.T., Saporito-Irwin, S., Howells, N., Ponta, H., Herrlich, P., and Gutmann, D.H. (1997). Interdomain binding mediates tumor growth suppression by the NF2 gene product. Oncogene *15*, 2505-2509.

Shimizu, T., Seto, A., Maita, N., Hamada, K., Tsukita, S., Tsukita, S., and Hakoshima, T. (2002). Structural basis for neurofibromatosis type 2. Crystal structure of the merlin FERM domain. J Biol Chem *277*, 10332-10336.

Skowronski, J., and Singer, M.F. (1985). Expression of a cytoplasmic LINE-1 transcript is regulated in a human teratocarcinoma cell line. Proc Natl Acad Sci U S A *82*, 6050-6054.

Smit, A.F. (1996). The origin of interspersed repeats in the human genome. Curr Opin Genet Dev *6*, 743-748.

Smith, T.W. (1989). The fundamental mechanism of inotropic action of digitalis. Therapie *44*, 431-435.

Solyom, S., Ewing, A.D., Rahrmann, E.P., Doucet, T., Nelson, H.H., Burns, M.B., Harris, R.S., Sigmon, D.F., Casella, A., Erlanger, B.*, et al.* (2012). Extensive somatic L1 retrotransposition in colorectal tumors. Genome Res *22*, 2328-2338.

Sperger, J.M., Chen, X., Draper, J.S., Antosiewicz, J.E., Chon, C.H., Jones, S.B., Brooks, J.D., Andrews, P.W., Brown, P.O., and Thomson, J.A. (2003). Gene expression patterns in human embryonic stem cells and human pluripotent germ cell tumors. Proc Natl Acad Sci U S A *100*, 13350-13355.

Stamenkovic, I., and Yu, Q. (2010). Merlin, a "magic" linker between extracellular cues and intracellular signaling pathways that regulate cell motility, proliferation, and survival. Curr Protein Pept Sci *11*, 471-484.

Stokowski, R.P., and Cox, D.R. (2000). Functional analysis of the neurofibromatosis type 2 protein by means of disease-causing point mutations. Am J Hum Genet *66*, 873-891.

Surace, E.I., Haipek, C.A., and Gutmann, D.H. (2004). Effect of merlin phosphorylation on neurofibromatosis 2 (NF2) gene function. Oncogene *23*, 580-587.

Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell *126*, 663-676.

Tang, X., Jang, S.W., Wang, X., Liu, Z., Bahr, S.M., Sun, S.Y., Brat, D., Gutmann, D.H., and Ye, K. (2007). Akt phosphorylation regulates the tumour-suppressor merlin through ubiquitination and degradation. Nat Cell Biol *9*, 1199-1207.

Trofatter, J.A., MacCollin, M.M., Rutter, J.L., Murrell, J.R., Duyao, M.P., Parry, D.M., Eldridge, R., Kley, N., Menon, A.G., Pulaski, K.*, et al.* (1993). A novel moesin-, ezrin-, radixin-like gene is a candidate for the neurofibromatosis 2 tumor suppressor. Cell *72*, 791-800.

Wei, W., Gilbert, N., Ooi, S.L., Lawler, J.F., Ostertag, E.M., Kazazian, H.H., Boeke, J.D., and Moran, J.V. (2001). Human L1 retrotransposition: cis preference versus trans complementation. Mol Cell Biol *21*, 1429-1439.

Xiao, G.H., Beeser, A., Chernoff, J., and Testa, J.R. (2002). p21-activated kinase links Rac/Cdc42 signaling to merlin. J Biol Chem *277*, 883-886.

Xiao, G.H., Gallagher, R., Shetler, J., Skele, K., Altomare, D.A., Pestell, R.G., Jhanwar, S., and Testa, J.R. (2005). The NF2 tumor suppressor gene product, merlin, inhibits cell proliferation and cell cycle progression by repressing cyclin D1 expression. Mol Cell Biol *25*, 2384-2394.

Xu, H.M., and Gutmann, D.H. (1998). Merlin differentially associates with the microtubule and actin cytoskeleton. J Neurosci Res *51*, 403-415.

Zeuthen, J., Norgaard, J.O., Avner, P., Fellous, M., Wartiovaara, J., Vaheri, A., Rosen, A., and Giovanella, B.C. (1980). Characterization of a human ovarian teratocarcinoma-derived cell line. Int J Cancer *25*, 19-32.

Zhang, A., Dong, B., Doucet, A.J., Moldovan, J.B., Moran, J.V., and Silverman, R.H. (2014). RNase L restricts the mobility of engineered retrotransposons in cultured human cells. Nucleic Acids Res *42*, 3803-3820.

Zoch, A., Mayerl, S., Schulz, A., Greither, T., Frappart, L., Rubsam, J., Heuer, H., Giovannini, M., and Morrison, H. (2015). Merlin Isoforms 1 and 2 Both Act as Tumour Suppressors and Are Required for Optimal Sperm Maturation. PLoS One *10*, e0129151.

Chapter 4

Conclusions

Overview

My dissertation research has focused on elucidating the mechanism(s) of L1-delivered reporter gene silencing (L1-REPEL) in human embryonic-derived cells known to accommodate endogenous L1 expression. Previous studies hypothesized that host factors recognize L1 TPRT intermediates to epigenetically silence L1-delivered reporter genes either during or immediately after retrotransposition (Garcia-Perez et al., 2010). In Chapter 2, we demonstrated that genome-wide CRISPR/Cas9-based gene knockout screens, in conjunction with validation assays, can efficiently lead to the identification of candidate genes that may be involved in L1-REPEL. We optimized the GeCKO CRISPR/Cas9 knockout system in PA-1 cells and established methods to prioritize candidate genes necessary for L1-REPEL. We identified 489 putative candidate genes that may affect L1-REPEL, including 20 that were highly enriched in our screens, and arrived at our top candidate gene, *NF2* (Neurofibromin 2). We also established a transient sgRNA/Cas9 plasmid-based assay to validate several candidate L1-REPEL genes. Thus, the successful application of CRISPR/Cas9 technology uncovered putative host factors responsible for L1-REPEL.

In Chapter 3, we investigated the role of the tumor suppressor protein NF2/merlin in L1-REPEL. We demonstrated that clonal- and population-based knockout of *NF2* in PA-1 cells attenuated L1-REPEL, suggesting that the NF2/merlin protein is necessary for efficient L1-REPEL. We further demonstrated that expression of the NF2/merlin isoform 1 cDNA efficiently restored L1-REPEL in *NF2* knockout cells, indicating NF2/merlin expression is sufficient to re-establish L1-REPEL. In contrast, knocking out *NF2* in two clonal PA-1 derivative clonal cell lines (pk5 and pc39) that contain silenced L1 retrotransposition events was not sufficient to alleviate L1-REPEL, implying that NF2/merlin may be necessary for the initiation, but not the maintenance step, of L1-REPEL. Finally, we found that L1-REPEL was less efficient in *NF2* knockout cells cultured in differentiation media *vs.* FBS-containing media, suggesting that *NF2* knockout and culturing cells in differentiation medium may act independently or in combination to attenuate L1-REPEL. The above results indicate that NF2/merlin is necessary to establish efficient L1-REPEL in PA-1 hECs and suggest that a known tumor suppressor gene that plays a role in human disease may act to silence endogenous L1 retrotransposition events in cell lines that serve as proxies for early stages of embryonic development. Below, I discuss the significance of the data presented in this dissertation and suggest possible future directions for ensuing research.

PA-1 cells exhibit L1-REPEL

Endogenous human LINE-1 elements are expressed at relatively high levels in human embryonic carcinoma-derived cells lines (hESCs) when compared to other immortalized cell lines (Garcia-Perez et al., 2007; Hohjoh and Singer, 1996; Skowronski

and Singer, 1985). However, despite their high levels of expression, reporter genes integrated into the genome using engineered L1s are efficiently and stably silenced upon retrotransposition in hECs (Garcia-Perez et al., 2010) by a mechanism termed L1-REPEL. Importantly, L1-REPEL appears to be peculiar to the TPRT mechanism used by non-LTR retrotransposons (see below), leading us to speculate that L1-REPEL may represent a mechanism to silence endogenous *de novo* LINE-1 retrotransposition events that occur during early human development (Garcia-Perez et al., 2010).

We confirmed previous reports that various L1-delivered reporter genes are subject to L1-REPEL in PA-1 embryonic carcinoma cells (Figure 2.1), but not somatic HeLa or HAP1 cells (Figure 3.2) (Garcia-Perez et al., 2010). We further confirmed that treating PA-1 cells with histone deacetylase inhibitors efficiently reactivates retrotransposition events subject to L1-REPEL and that subsequent removal of histone deacetylase inhibitors reestablishes L1-REPEL (Figure 2.1 B and C) (Garcia-Perez et al., 2010).

These results demonstrate that L1-REPEL is likely an epigenetic mechanism that is both mitotically stable and reversible. Our L1-REPEL working model posits that both initiation and maintenance steps are required for efficient epigenetic silencing of L1-delivered reporter genes (Figure 2.1 D). Given that L1-REPEL appears to be specific to TPRT-mediated retrotransposition (Garcia-Perez et al., 2010), we hypothesize that cellular factors recognize TPRT intermediates leading to the establishment of an epigenetic mark required for L1-REPEL in PA-1 cells. Moreover, we hypothesize that the same or different cellular factors are required to both initiate and maintain L1-REPEL in PA-1 cells.

Genome-wide CRISPR/Cas9-based gene knockout screens identify candidate L1-REPEL factors

We designed and implemented a forward genetic screen utilizing a lentiviral-delivered CRISPR/Cas9-based system (GeCKO) to knockout genes on a genome-wide scale (Sanjana et al., 2014; Shalem et al., 2014), then performed L1 retrotransposition assays to identify genes necessary for L1-REPEL (Figure 2.2 B). These L1-REPEL screens yielded a candidate gene list comprising 489 genes, including 20 highly enriched candidates (Table 2.1). We then validated a subset of our candidate genes, including *NF2*, *XPO7*, *TADA2B*, and *ZC3H12B*, using a plasmid-based gene knockout assay (Figure 2.4). Subsequent validation assays revealed *DDB2*, *NCBP2, TP53, NCAPG, MPHOSPH8*, *ATF7IP, DOT1L, EHMT1*, *and TAF5L* as additional genes necessary for efficient L1-REPEL (Figure A.2). Together, our screening and validation strategies identified several L1-REPEL factors that might be involved in the initiation and/or maintenance of L1-REPEL. Future experiments are necessary to distinguish factors that mediate initiation vs maintenance phases of L1-REPEL.

Is TPRT recognized by cellular factors?

During Target-Primed Reverse Transcription (TPRT), ORF2p EN makes a single-strand endonucleolytic nick in DNA, liberating a 3' hydroxyl group that can be used as a primer by ORF2p RT to initiate L1 cDNA synthesis (see Chapter 1). Although the downstream steps of TPRT are not completely understood, the completion of TPRT likely requires cellular factors that facilitate retrotransposition. For example, the single-strand endonucleolytic nick induced by ORF2p EN can trigger the recruitment and activation of poly(ADP-ribose) polymerase 2 (PARP2) leading to the subsequent

recruitment of the replication protein A (RPA) complex to facilitate retrotransposition (Miyoshi et al., 2019). Additionally, core members of the nucleotide excision repair (NER) pathway, including XPC, can recognize 3' DNA flap structures generated during TPRT, limiting L1 retrotransposition (Servant et al., 2017). Together, these results demonstrate that cellular factors can recognize TPRT.

We propose that host factors necessary to initiate L1-REPEL may be recruited to the site of L1 integration via direct interaction with ORF2p and/or may recognize DNA structures generated during TPRT, resulting in the direct or indirect establishment of an epigenetic mark and transcriptional repression of the L1-delivered retrotransposition indicator cassette and perhaps sequences within the *de novo* engineered L1 insertion (Figure 4.1 A). For example, *DDB2* encodes a protein that recognizes DNA damage when complexed with DDB1, CUL4A, and PARP1 (Appendix) (Luijsterburg et al., 2007; Pines et al., 2012). Interestingly, DDB2 interacts with histone deacetylases (HDAC1 and HDAC2) to decrease H3K56 acetylation after DNA damage (Zhu et al., 2015). Thus, knockout of *DDB2* might hinder the DNA damage response at sites of TPRT, resulting in escape from L1-REPEL. Intriguingly, NF2/merlin also interacts with DDB1 and CUL4A as part of an E3 ubiquitin ligase complex, suggesting a possible connection between NF2/merlin and DDB2.

Host factors necessary to maintain L1-REPEL may be recruited to the site of L1 integration via the epigenetic mark established during the initiation phase. These factors likely mediate HDAC-containing corepressor complex formation and transcriptional repression of the L1-delivered retrotransposition indicator cassette (Figure 4.1 B). For example, *MPHOSPH8* and *ATF7IP* encode proteins that are members of the

heterochromatin-forming HUSH complex, which previously was implicated in retrotransposon silencing (Liu et al., 2018; Robbez-Masson et al., 2018; Seczynska et al., 2021; Zhu et al., 2018). The HUSH complex spreads heterochromatin via deposition of H3K9me3, acting to maintain transcriptional repression (Tchasovnikarova et al., 2015). Thus, knockout of *MPHOSPH8* or *ATF7IP* may result in the loss of H3K9 methylation at sites of L1 retrotransposition, resulting in escape from L1-REPEL. Intriguingly, chaetocin, an inhibitor of the H3K9 histone methyltransferase SU(VAR)3-9 (Cherblanc et al., 2013; Greiner et al., 2005), also led to the reactivation of EGFP expression in pk5 and pc39 cells at levels comparable to TSA-treatment (Figure A.3, see Appendix). Hence, *MPHOSPH8* and *ATF7IP* may promote a stable and reversible epigenetic mark (H3K9 methylation) that promotes L1-REPEL.

NF2/merlin is necessary for efficient L1-REPEL in PA-1 cells

In Chapter 3, we established that knockout of *NF2* in PA-1 cells attenuated L1-REPEL (Figure 3.5). We further demonstrated that expression of NF2/merlin isoform 1 efficiently restored L1-REPEL in clonal *NF2* knockout cells (Figure 3.8). Our working model posits that L1-REPEL occurs through a two-step initiation and maintenance process to efficiently and stably silence L1-delivered reporter genes in PA-1 cells. Maintenance assays revealed that knockout of *NF2* in pc39 cells did not influence L1-REPEL, implying that NF2/merlin may be necessary to initiate L1-REPEL (Figure 3.10). Lastly, we found that L1-REPEL was less efficient in *NF2* knockout cells cultured in differentiation media, suggesting that L1-REPEL is further attenuated during cellular differentiation (Figure 3.11).

The above results have led to a working model where I propose that NF2/merlin regulates the availability of embryonic cellular host factors that are necessary to initiate L1-REPEL in PA-1 cells (Figure 4.2). These factors likely recognize DNA structures generated by ORF2p during TPRT, resulting in the direct or indirect establishment of an epigenetic mark, leading to transcriptional silencing of the L1-delivered reporter gene (Figure 4.2 A and B). Based on my and previous studies (Garcia-Perez et al., 2010), I posit that transcriptional repression could be achieved through the recruitment of corepressor complexes containing histone deacetylase and histone methyltransferase activity (Figure 4.2 B). In this model, cells containing functional NF2/merlin exhibit efficient L1-REPEL (Figure 4.2 B: left and right panels), whereas cells lacking functional NF2/merlin exhibit downregulation of cellular factors necessary for efficient L1-REPEL (Figure 4.2 B: middle panel). One question remains, how does NF2/merlin regulate the availability of cellular factors that mediate L1-REPEL?

Does E3 ubiquitin ligase activity regulate the availability of L1-REPEL factors?

Intra-molecular association of NF2/merlin promotes a “closed” conformation, which allows its translocation to the nucleus (Rong et al., 2004; Shimizu et al., 2002). Nuclear NF2/merlin can bind DCAF1 (VPRBP) and modulate $CRL4^{DCAF1}$ E3 ubiquitin ligase activity (Li et al., 2010). Polyubiquitination typically marks proteins for proteasomal degradation, whereas addition of single ubiquitin molecule often modifies protein localization and function (Rape, 2018). These data suggest that NF2/merlin can alter $CRL4^{DCAF1}$ E3 ubiquitin ligase activity in the nucleus. Intriguingly, a recent report demonstrated that retroviral silencing is regulated by a similar mechanism involving E3 ubiquitin ligase activity (Wang and Goff, 2017).

In embryonic cells, ZFP809 binds the retroviral tRNA binding site and blocks viral DNA synthesis. In differentiated cells, ZFP809 is modified by polyubiquitin chains, resulting in proteasomal degradation of the ZFP809 protein, which allows tRNA priming of viral DNA synthesis. The E3 ubiquitin ligase TRIM28 was found to promote the degradation of ZFP809 in differentiated cells (Wang and Goff, 2017). Furthermore, these findings illustrate that the proteasomal degradation pathway can regulate the availability of cellular factors in the embryonic state, and that E3 ubiquitin ligase activity coincides with cellular differentiation.

These findings have led me to speculate that NF2/merlin controls the availability of L1-REPEL factors through regulation of $CRL4^{DCAF1}$ E3 ubiquitin ligase activity (Figure 4.3). Activation of NF2/merlin promotes its nuclear translocation and binding to DCAF1. NF2/merlin binding to DCAF1 inhibits $CRL4^{DCAF1}$ E3 ubiquitin ligase activity, preventing degradation of DCAF1 substrates that promote L1-REPEL (Figure 4.3). In the absence of NF2/merlin, $CRL4^{DCAF1}$ can ubiquitinate DCAF1 substrates, resulting in protein degradation. Thus, in *NF2* knockout cells, $CRL4^{DCAF1}$-mediated degradation of L1-REPEL factors results in attenuation of L1-REPEL (Figure 4.3). For example, TP53 is a $CRL4^{DCAF1}$ substrate that is highly expressed in embryonic cells, then undergoes downregulation during cellular differentiation (Klijn et al., 2015; Lutzker and Levine, 1996). Therefore, NF2/merlin may inhibit $CRL4^{DCAF1}$-mediated degradation of TP53 in embryonic cells, allowing TP53 to mediate L1-REPEL. However, during cellular differentiation NF2/merlin would be confined to the cytoplasm allowing $CRL4^{DCAF1}$-mediated degradation of TP53 and attenuation of L1-REPEL. Ultimately, all $CRL4^{DCAF1}$

substrates that are downregulated in *NF2* knockout cells would be primary L1-REPEL candidates.

Does Hippo signaling influence L1-REPEL?

Previous reports suggest that loss of NF2/merlin results in CRL4DCAF1-mediated ubiquitination of core Hippo proteins LATS1 and LATS2, thereby inhibiting the Hippo signaling pathway (Li et al., 2014a). Originally identified in Drosophila, Hippo signaling is an evolutionarily conserved pathway implicated in the regulation of stem cell self-renewal, cell proliferation, and cell fate determination (Zheng and Pan, 2019). The mammalian Hippo pathway comprises a kinase cascade that negatively regulates the downstream effector Yes-associated protein (YAP). Canonical Hippo signaling in mammals is initiated by TAO kinase (TAOK1-3), which directly phosphorylates and activates MST1/2, the mammalian homolog of Drosophila hippo kinase (Boggiano et al., 2011). Activated MST1/2 subsequently phosphorylates and activates LATS1/2. Phosphorylated LATS1/2 then mediates the phosphorylation and inactivation of the Hippo effector YAP. Inactivated YAP is sequestered to the cytoplasm by 14-3-3 proteins, inhibiting downstream regulation of YAP-target genes (Zhao et al., 2007). When Hippo signaling is active, or “on,” YAP-mediated signaling is repressed. When Hippo is inactive, or “off,” YAP translocates to the nucleus and associates with a diverse set of cofactors involved in transcriptional activation and repression (Zhao et al., 2007).

Hippo signaling can mediate a variety of cellular responses; thus, elucidating the upstream regulators of Hippo signaling is of primary interest. Many reports have implicated NF2/merlin as a biochemical and biomechanical sensor, promoting Hippo pathway activation (Cooper and Giancotti, 2014; Zheng and Pan, 2019). Furthermore,

PTM of NF2/merlin acts as a switch to control its tumor suppressor activity and subcellular localization. In addition to phosphorylation (see Chapter 3), sumoylation, and ubiquitination have been implicated in NF2/merlin regulation. For example, NEDD4L-mediated ubiquitination of NF2/merlin was reported to activate Hippo signaling (Wei et al., 2020), which was dependent upon serine 518 dephosphorylation in response to calcium (Liu et al., 2019) or cell detachment (Zhao et al., 2012). These data demonstrate that NF2/merlin can act as a versatile upstream regulator of Hippo signaling in response to cellular stimuli.

We postulate that NF2/merlin-mediated activation of the Hippo pathway may be necessary for L1-REPEL in PA-1 cells. We reasoned that if Hippo pathway activation is necessary for L1-REPEL, factors mediating Hippo signaling would be enriched in the L1-REPEL screens. Indeed, within our 20 highly enriched candidate genes, we identified TAOK1, a serine/threonine protein kinase that directly phosphorylates MST1/2 and activates Hippo signaling (Boggiano et al., 2011). Like NF2/merlin, TAOK1 has been implicated in the cellular stress response and cytoskeleton regulation (Fang et al., 2020). We also identified several moderately enriched candidates, including the core Hippo members LATS1 and LATS2 (Table 4.1).

Given these findings, we propose a model where NF2/merlin regulates Hippo-dependent availability of L1-REPEL factors in PA-1 cells (Figure 4.4). Cells lacking NF2/merlin cannot activate Hippo signaling in response to cellular stimuli or cell stress. This lack of Hippo signaling promotes nuclear translocation of YAP, where it can associate with co-repressor complexes, such as NuRD (Nucleosome Remodeling and Deacetylase) and SWI/SNF (SWItch/Sucrose Non-Fermentable), resulting in

transcriptional silencing of YAP-target genes (Chang et al., 2018; Hillmer and Link, 2019).

Future Directions

Candidate gene validation assays

Our validations assays assume that we have a high knockout efficiency in a population of cells. Candidate gene knockout vectors contain a puromycin selectable marker, which should promote efficient editing within the puromycin resistant cell population. However, to arrive at rigorous conclusions regarding the necessity of cellular factors in maintaining L1-REPEL, we need to ensure that a functional protein is significantly diminished or absent in our cell populations by western blot analyses. Quantification of endogenous protein would then serve as a proxy for overall editing efficiency.

To assess editing efficiency more accurately, genomic DNA from the puromycin resistant cell population could be subjected to PCR amplification using primers flanking the sgRNA-target site followed by high throughput sequencing. The percentage of reads containing mutations (*i.e.*, the number of indel reads *vs.* reference reads) could then be used to assess the resultant editing efficiency. Notably, the half-life of the candidate protein must be considered to account for protein turnover when generating “population knockout” cells. Therefore, validation assays should include collecting cells several days post transfection to assess editing efficiency and/or stable protein expression. An alternative, but perhaps more time consuming, approach to the population-based assay would be to generate independent knockout clones. Characterization of sgRNA-target alleles and endogenous protein expression in several knockout cell lines would then

provide a complementary and rigorous approach to determine whether a candidate factor is necessary for L1-REPEL in cells.

Knockout *LATS1/2* in PA-1 cells

To determine if Hippo signaling is necessary for L1-REPEL, population- and clonal-based knockouts of core Hippo components LATS1 and LATS2 would be characterized following the above guidelines, then subjected to L1 retrotransposition assays. Notably, I previously generated knockout vectors containing *LATS1* and *LATS2* sgRNAs.

RNA sequencing of *NF2* knockout cells

L1-REPEL is downregulated during cellular differentiation of PA-1 cells (Figure 3.11) (Garcia-Perez et al., 2010). We found that knockout of *NF2* in PA-1 cells similarly attenuated L1-REPEL (Figure 3.11). Therefore, it is possible that *NF2* knockout in PA-1 cells disrupts the embryonic cell state, leading to cellular differentiation and indirect attenuation of L1-REPEL. Intriguingly, we found that culturing *NF2* knockout cells in differentiation medium further attenuated L1-REPEL, however, these cells continued to express the pluripotency marker OCT3/4 (Figure 3.11).

To characterize the cell state and define the extent of cellular differentiation, RNA sequencing would be performed in *NF2* knockout cells. Expression of genes associated with cell stemness and/or differentiation would be analyzed to determine if *NF2* knockout cells are in a state of intermediate differentiation— *i.e.*, differentially expressing a subset of genes associated with cellular differentiation compared to PA-1 cells.

Inhibition of ubiquitin-dependent proteasomal degradation

To test whether E3 ubiquitin ligase activity influences L1-REPEL, *NF2* knockout cells would be subjected to the L1 retrotransposition assay in the presence of MG132, a potent membrane-permeable proteasome inhibitor. MG132 treatment would be carried out following a previously established protocol with minor modifications (Wang and Goff, 2017). *NF2* knockout cells would be treated with ~10 μM MG132 twenty-four hours prior to L1 transfection, then treated daily for three days or until L1-reporter drug selection commences.

An alternative approach to assess E3 ubiquitin ligase activity in L1-REPEL would be to inhibit specific components of the E3 ubiquitin ligase complexes. For example, multiple studies have shown that genetic perturbation of DCAF1 can abolish $CRL4^{DCAF1}$ substrate recognition (Li et al., 2010; Nakagawa et al., 2015). To determine if $CRL4^{DCAF1}$ activity effects L1-REPEL, CRISPR/Cas9-mediated knock-in of a DCAF1 mutant lacking the C-terminal substrate recognition domain would be generated in *NF2* knockout cells, then subjected to the L1 retrotransposition assay. An increase in L1-REPEL would suggest that $CRL4^{DCAF1}$ E3 ubiquitin ligase activity likely facilitates proteasomal degradation of cellular factors that mediate L1-REPEL.

ORF2p pulldown in PA-1 cells

ORF2p immunoprecipitation (IP)-coupled mass spectrometry would be carried out as previously described (Miyoshi et al., 2019) to identify L1 ORF2p interacting proteins. Our model proposes that L1-REPEL factors recognize DNA structures generated by ORF2p during TPRT. Thus, Flag-tagged ORF2p would be expressed (via pTMF3 and

pTMO2F3 vectors) and immunoprecipitated in PA-1 cells to identify factors associated with TPRT.

Is L1-REPEL specific to TPRT?

We hypothesize that L1-REPEL factors recognize DNA structures generated by ORF2p during TPRT, resulting in the establishment of an epigenetic mark and transcriptional silencing of the L1-delivered reporter gene. However, many DNA structures generated during TPRT are not unique to L1 integration. For example, naturally occurring single-strand and double-strand breaks in DNA can generate intermediates similar to ORF2p EN activity. Therefore, it is possible that the embryonic state promotes a global silencing mechanism at sites of DNA damage. To test this, reporter genes would be delivered into the genome of PA-1 cells using a variety of mechanisms. Then, PA-1 cells would be assessed for expression of the delivered reporter gene. Notably, previous reports established that retroviral delivered *EGFP* reporter genes are efficiently expressed in PA-1 cells (Garcia-Perez et al., 2010).

First, an *EGFP* reporter gene would be delivered by CRISPR/Cas9-mediated knock-in at a genomic location specific to the pk5 L1-EGFP insertion on chromosome 12. The Cas9 nuclease generates a double-strand break, allowing homologous recombination of a DNA donor template containing EGFP flanked by homology arms. If the double-strand break induced by Cas9 recruits silencing factors, we would expect transcriptional silencing of the *EGFP* reporter gene. This result would support a global silencing mechanism at sites of DNA damage.

Second, an *EGFP* reporter gene would be delivered by the Sleeping Beauty transposon system (see Chapter 1). The sleeping beauty system utilizes a transposase

enzyme to mediate genomic integration of a transposon containing genetic cargo, or the *EGFP* reporter gene. If transposase activity recruits silencing factors, we would expect transcriptional silencing of the *EGFP* reporter gene. This result would support a global silencing mechanism.

Lastly, a *Neo* reporter gene would be delivered by Alu retrotransposition. Alu retrotransposons steal ORF2p from L1 to facilitate TPRT (see Chapter 1). Therefore, we would expect transcriptional silencing of the *neo* reporter gene in PA-1 cells subjected to the Alu retrotransposition assay. PA-1 cells express high levels of endogenous L1s. Therefore, Alu retrotransposition assays would be done with or without cotransfection of ORF2p.

Summary

My thesis has examined an epigenetic silencing phenomenon that restricts the expression of L1-delivered reporter genes (L1-REPEL) in a cell line that serves as a proxy for early stages of human development. We designed and implemented a genome-wide CRISPR-Cas9-based genetic screen that identified cellular factors influencing L1-REPEL in PA-1 cells. We developed an adaptable transient plasmid-based workflow that efficiently allowed the validation of candidate genes mediating L1-REPEL. We demonstrated that NF2/merlin was necessary for efficient L1-REPEL by establishing and characterizing clonal *NF2* knockout cell lines and that reintroduction of NF2/merlin was sufficient to reestablish L1-REPEL in *NF2* knockout cells. We also generated preliminary data that should allow a future assessment of the role(s) of NF2/merlin in regulating $CRL4^{DCAF1}$ E3 ubiquitin ligase activity and/or Hippo signaling in relation to L1-REPEL. Finally, we utilized small molecule inhibitors to interrogate

epigenetic regulation of L1-REPEL. My preliminary results suggest that histone methyltransferase and DNA methyltransferase activity may possibly contribute to L1-REPEL. In sum, this work has led to a better understanding of how cellular factors and epigenetic regulation mediate the repression of L1-delivered reporter genes in PA-1 cells. We speculate that a similar mechanism acts to suppress *de novo* L1 retrotransposition events during human embryonic development.

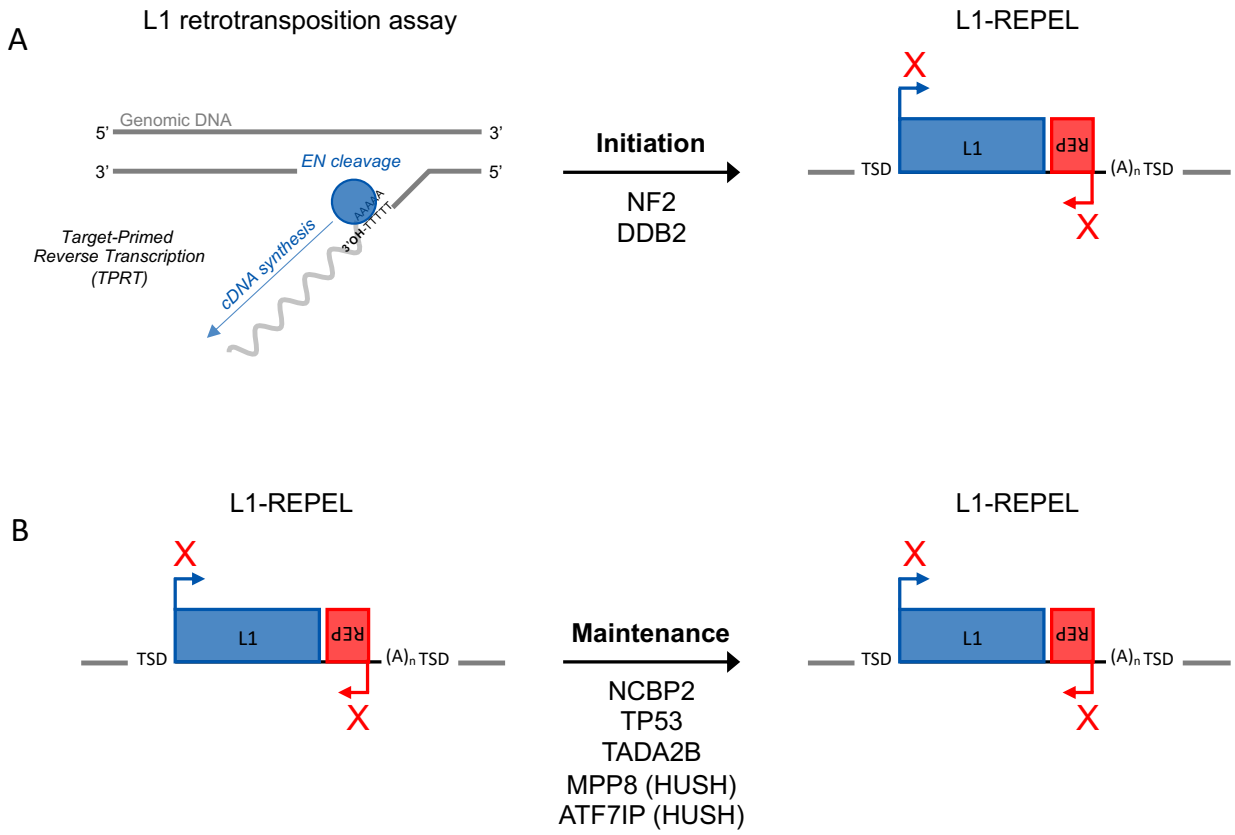


Figure 4.1: Candidate factors may affect different steps of L1-REPEL.

A schematic illustrating the two phases of L1-REPEL. TPRT-mediated genomic integration triggers L1-REPEL, resulting in the transcriptional repression of the L1-delived reporter gene and perhaps the retrotransposed L1 itself. (A) Indicated are candidate genes necessary for efficient (A) **initiation** or (B) **maintenance** of L1-REPEL. The blue circle represents ORF2p. The X represents transcriptional repression. TSD indicates target site duplications flanking the L1 insertions.

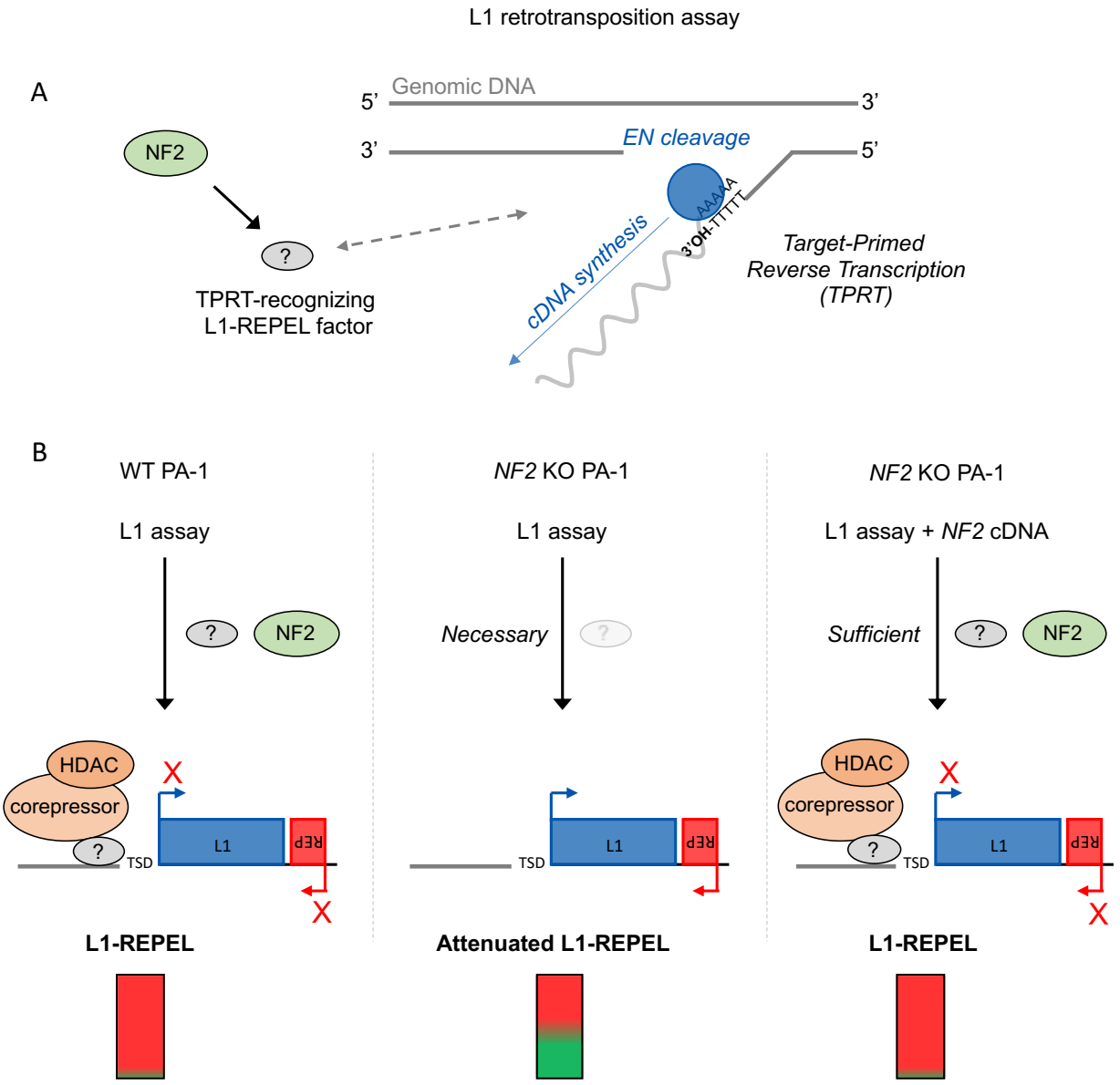


Figure 4.2: NF2/merlin is necessary for efficient L1-REPEL in PA-1 cells.

(A) A working model of L1-REPEL. NF2/merlin-dependent availability of factors that recognize TPRT determines L1-REPEL efficiency in PA-1 cells. The blue circle represents ORF2p. (B) Wild-type PA-1 cells expressing NF2/merlin exhibit efficient L1-REPEL (left panel). *NF2* knockout PA-1 cells lacking NF2/merlin expression exhibit attenuated L1-REPEL (middle panel). Transient expression of NF2/merlin is sufficient to reestablish L1-REPEL in *NF2* knockout cells (right panel). The X represents transcriptional repression via factor-dependent recruitment of co-repressor complexes

containing histone deacetylase (HDAC) activity. The gradient bars at the bottom of each panel represent L1-REPEL efficiency. TSD; target site duplications.

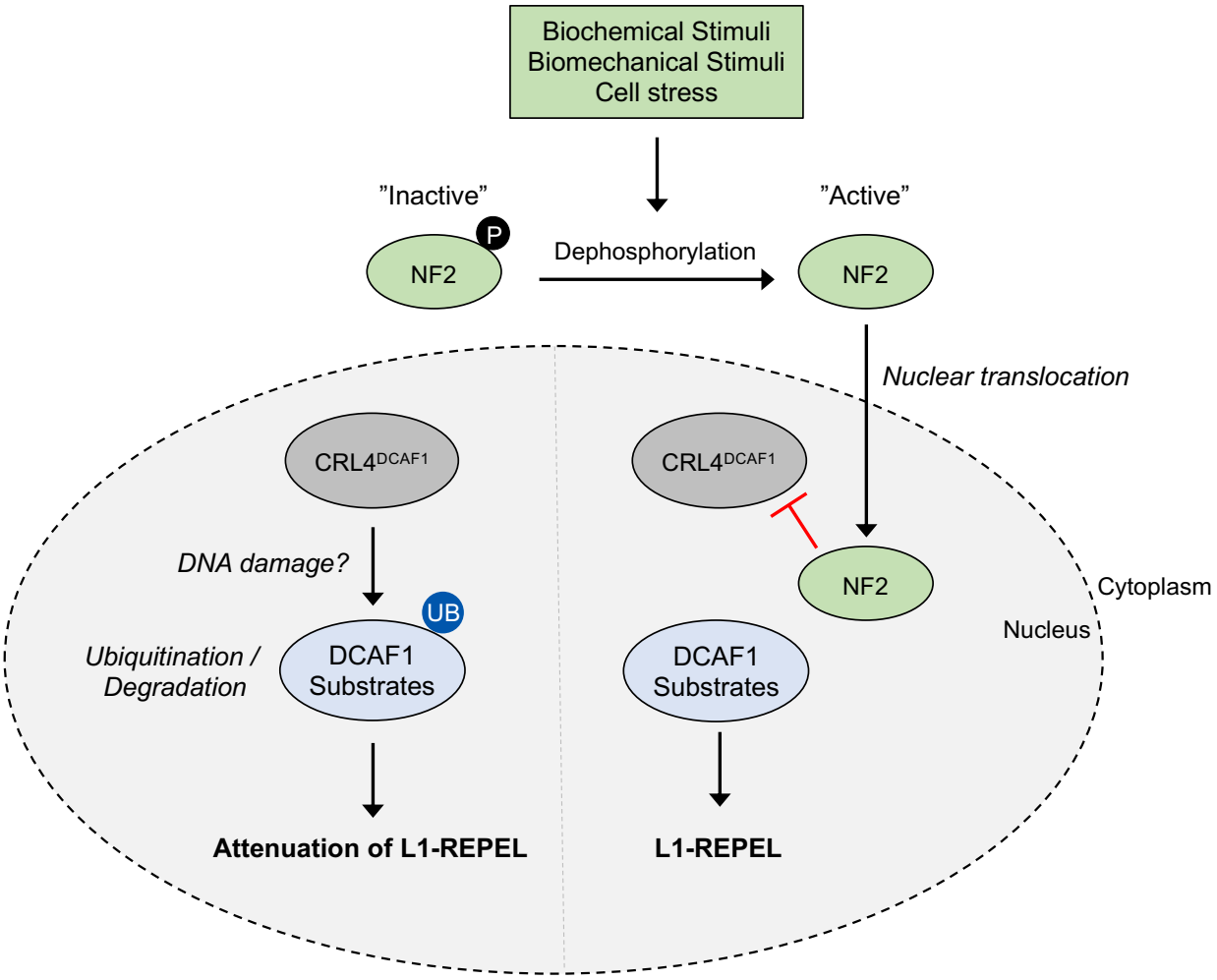


Figure 4.3: NF2/merlin regulates CRL4^{DCAF1}-dependent degradation of L1-REPEL factors in PA-1 cells.

Activation of NF2/merlin promotes its nuclear translocation and binding to DCAF1. NF2/merlin binding to DCAF1 inhibits CRL4^{DCAF1} E3 ubiquitin ligase activity, preventing degradation of DCAF1 substrates that promote L1-REPEL. In the absence of NF2/merlin, CRL4^{DCAF1} can ubiquitinate DCAF1 substrates. Ubiquitination can alter protein function or mark the protein for degradation. Thus, CRL4^{DCAF1}-mediated down-regulation of L1-REPEL factors results in attenuation of L1-REPEL. P indicates phosphorylation. UB indicates ubiquitination.

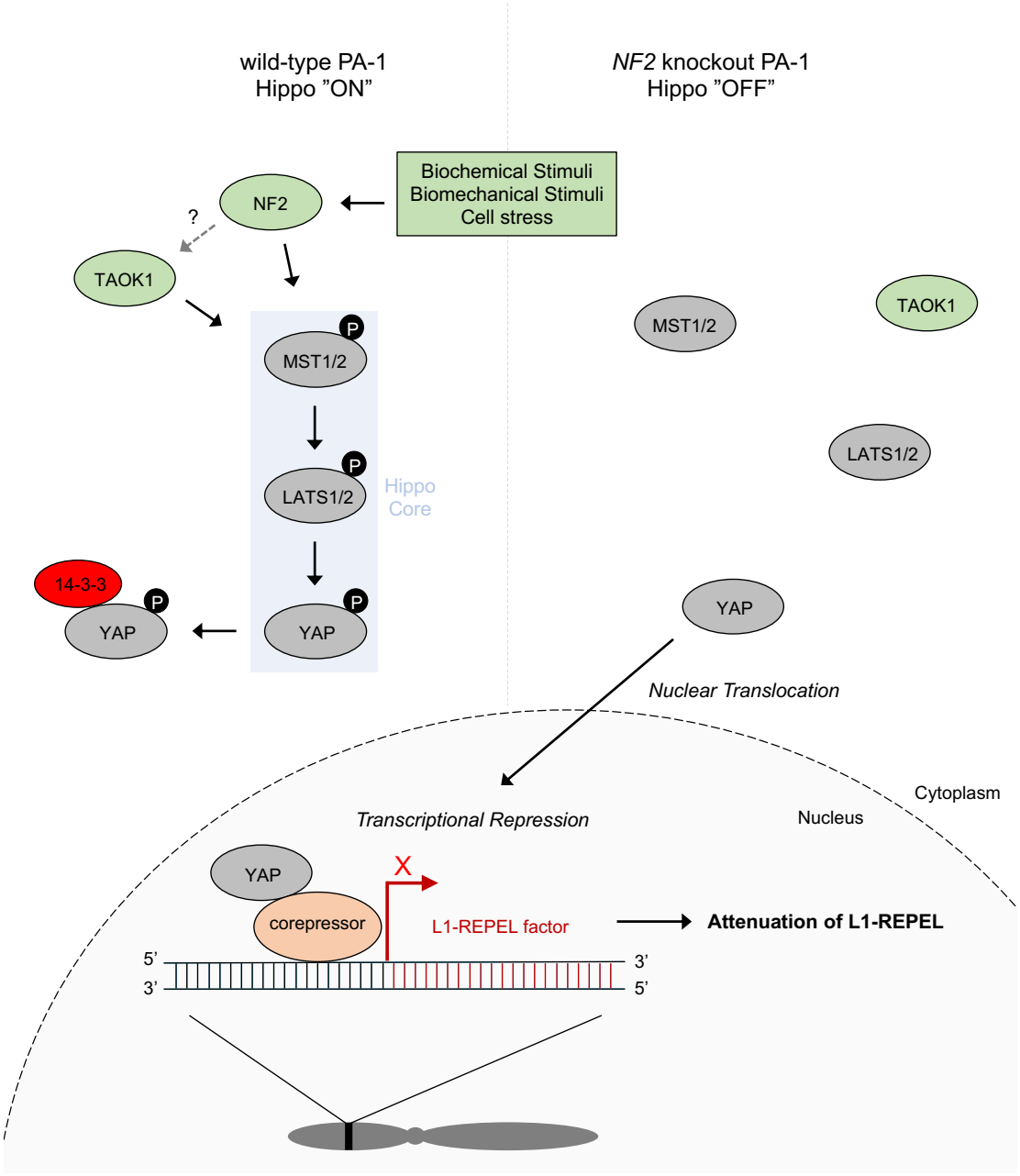


Figure 4.4: Hippo-dependent L1-REPEL in PA-1 cells.

NF2/merlin activates the Hippo kinase cascade in response to cellular stimuli, possibly through TAOK1-mediated phosphorylation of MST1/2. Hippo signaling results in sequestration of YAP to the cytoplasm by a 14-3-3 protein. In the absence of functional NF2/merlin, Hippo signaling is inhibited, promoting the nuclear translocation of YAP. Nuclear YAP associates with co-repressor complexes, mediating down-regulation of L1-REPEL factors. The X represents transcriptional repression.

Table 4.1: Hippo pathway genes identified in the L1-REPEL screen.

	# filtered sgRNAs		
Gene	GeCKO 1	GeCKO 2	Brunello
NF2	4	4	2
TAOK1	2	1	1
LATS2	1	2	1
LATS1	1	2	0
FAT4	1	1	1
PARD6G	0	1	2
PARD6A	1	1	1
TCF7	1	1	1
WWP2	2	0	1
PTPN14	1	1	0

References

Ahronowitz, I., Xin, W., Kiely, R., Sims, K., MacCollin, M., and Nunes, F.P. (2007). Mutational spectrum of the NF2 gene: a meta-analysis of 12 years of research and diagnostic laboratory findings. Hum Mutat *28*, 1-12.

Aksu, M., Pleiner, T., Karaca, S., Kappert, C., Dehne, H.J., Seibel, K., Urlaub, H., Bohnsack, M.T., and Gorlich, D. (2018). Xpo7 is a broad-spectrum exportin and a nuclear import receptor. J Cell Biol *217*, 2329-2340.

Bao, X.M., He, Q., Wang, Y., Huang, Z.H., and Yuan, Z.Q. (2017). The roles and mechanisms of the Hippo/YAP signaling pathway in the nervous system. Yi Chuan *39*, 630-641.

Billings, S.E., Pierzchalski, K., Butler Tjaden, N.E., Pang, X.Y., Trainor, P.A., Kane, M.A., and Moise, A.R. (2013). The retinaldehyde reductase DHRS3 is essential for preventing the formation of excess retinoic acid during embryonic development. FASEB J *27*, 4877-4889.

Blomen, V.A., Majek, P., Jae, L.T., Bigenzahn, J.W., Nieuwenhuis, J., Staring, J., Sacco, R., van Diemen, F.R., Olk, N., Stukalov, A*., et al.* (2015). Gene essentiality and synthetic lethality in haploid human cells. Science *350*, 1092-1096.

Boggiano, J.C., Vanderzalm, P.J., and Fehon, R.G. (2011). Tao-1 phosphorylates Hippo/MST kinases to regulate the Hippo-Salvador-Warts tumor suppressor pathway. Dev Cell *21*, 888-895.

Calses, P.C., Crawford, J.J., Lill, J.R., and Dey, A. (2019). Hippo Pathway in Cancer: Aberrant Regulation and Therapeutic Opportunities. Trends Cancer *5*, 297-307.

Cantu, I., van de Werken, H.J.G., Gillemans, N., Stadhouders, R., Heshusius, S., Maas, A., Esteghamat, F., Ozgur, Z., van, I.W.F.J., Grosveld, F*., et al.* (2019). The mouse KLF1 Nan variant impairs nuclear condensation and erythroid maturation. PLoS One *14*, e0208659.

Chang, L., Azzolin, L., Di Biagio, D., Zanconato, F., Battilana, G., Lucon Xiccato, R., Aragona, M., Giulitti, S., Panciera, T., Gandin, A*., et al.* (2018). The SWI/SNF complex is a mechanoregulated inhibitor of YAP and TAZ. Nature *563*, 265-269.

Chen, H., Zhang, Z., Jiang, S., Li, R., Li, W., Zhao, C., Hong, H., Huang, X., Li, H., and Bo, X. (2020). New insights on human essential genes based on integrated analysis

and the construction of the HEGIAP web-based platform. Brief Bioinform *21*, 1397-1410.

Cheon, Y., Kim, H., Park, K., Kim, M., and Lee, D. (2020). Dynamic modules of the coactivator SAGA in eukaryotic transcription. Exp Mol Med *52*, 991-1003.

Cherblanc, F.L., Chapman, K.L., Brown, R., and Fuchter, M.J. (2013). Chaetocin is a nonspecific inhibitor of histone lysine methyltransferases. Nat Chem Biol *9*, 136-137.

Cockburn, K., Biechele, S., Garner, J., and Rossant, J. (2013). The Hippo pathway member Nf2 is required for inner cell mass specification. Curr Biol *23*, 1195-1201.

Cooper, J., and Giancotti, F.G. (2014). Molecular insights into NF2/Merlin tumor suppressor function. FEBS Lett *588*, 2743-2752.

Correia-Melo, C., Marques, F.D., Anderson, R., Hewitt, G., Hewitt, R., Cole, J., Carroll, B.M., Miwa, S., Birch, J., Merz, A.*, et al.* (2016). Mitochondria are required for pro-ageing features of the senescent phenotype. EMBO J *35*, 724-742.

Costanzi, C., and Pehrson, J.R. (1998). Histone macroH2A1 is concentrated in the inactive X chromosome of female mammals. Nature *393*, 599-601.

Cui, Y., Groth, S., Troutman, S., Carlstedt, A., Sperka, T., Riecken, L.B., Kissil, J.L., Jin, H., and Morrison, H. (2019). The NF2 tumor suppressor merlin interacts with Ras and RasGAP, which may modulate Ras signaling. Oncogene *38*, 6370-6381.

Dai, L., Taylor, M.S., O'Donnell, K.A., and Boeke, J.D. (2012). Poly(A) binding protein C1 is essential for efficient L1 retrotransposition and affects L1 RNP formation. Mol Cell Biol *32*, 4323-4336.

Douet, J., Corujo, D., Malinverni, R., Renauld, J., Sansoni, V., Posavec Marjanovic, M., Cantarino, N., Valero, V., Mongelard, F., Bouvet, P.*, et al.* (2017). MacroH2A histone variants maintain nuclear organization and heterochromatin architecture. J Cell Sci *130*, 1570-1582.

Fang, C.Y., Lai, T.C., Hsiao, M., and Chang, Y.C. (2020). The Diverse Roles of TAO Kinases in Health and Diseases. Int J Mol Sci *21*.

Flasch, D.A., Macia, A., Sanchez, L., Ljungman, M., Heras, S.R., Garcia-Perez, J.L., Wilson, T.E., and Moran, J.V. (2019). Genome-wide de novo L1 Retrotransposition Connects Endonuclease Activity with Replication. Cell *177*, 837-851 e828.

Fleming, A.B., Kao, C.F., Hillyer, C., Pikaart, M., and Osley, M.A. (2008). H2B ubiquitylation plays a role in nucleosome dynamics during transcription elongation. Mol Cell *31*, 57-66.

Fu, M., and Blackshear, P.J. (2017). RNA-binding proteins in immune regulation: a focus on CCCH zinc finger proteins. Nat Rev Immunol *17*, 130-143.

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

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

Garcia-Rendueles, M.E., Ricarte-Filho, J.C., Untch, B.R., Landa, I., Knauf, J.A., Voza, F., Smith, V.E., Ganly, I., Taylor, B.S., Persaud, Y.*, et al.* (2015). NF2 Loss Promotes Oncogenic RAS-Induced Thyroid Cancers via YAP-Dependent Transactivation of RAS Proteins and Sensitizes Them to MEK Inhibition. Cancer Discov *5*, 1178-1193.

Golovnina, K., Blinov, A., Akhmametyeva, E.M., Omelyanchuk, L.V., and Chang, L.S. (2005). Evolution and origin of merlin, the product of the Neurofibromatosis type 2 (NF2) tumor-suppressor gene. BMC Evol Biol *5*, 69.

Gonzalez-Agosti, C., Wiederhold, T., Herndon, M.E., Gusella, J., and Ramesh, V. (1999). Interdomain interaction of merlin isoforms and its influence on intermolecular binding to NHE-RF. J Biol Chem *274*, 34438-34442.

Goodier, J.L. (2016). Restricting retrotransposons: a review. Mob DNA *7*, 16.

Goodier, J.L., Cheung, L.E., and Kazazian, H.H., Jr. (2012). MOV10 RNA helicase is a potent inhibitor of retrotransposition in cells. PLoS Genet *8*, e1002941.

Goodier, J.L., Cheung, L.E., and Kazazian, H.H., Jr. (2013). Mapping the LINE1 ORF1 protein interactome reveals associated inhibitors of human retrotransposition. Nucleic Acids Res *41*, 7401-7419.

Goodier, J.L., Pereira, G.C., Cheung, L.E., Rose, R.J., and Kazazian, H.H., Jr. (2015). The Broad-Spectrum Antiviral Protein ZAP Restricts Human Retrotransposition. PLoS Genet *11*, e1005252.

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

Gorgoulis, V., Adams, P.D., Alimonti, A., Bennett, D.C., Bischof, O., Bishop, C., Campisi, J., Collado, M., Evangelou, K., Ferbeyre, G.*, et al.* (2019). Cellular Senescence: Defining a Path Forward. Cell *179*, 813-827.

Greiner, D., Bonaldi, T., Eskeland, R., Roemer, E., and Imhof, A. (2005). Identification of a specific inhibitor of the histone methyltransferase SU(VAR)3-9. Nat Chem Biol *1*, 143-145.

Gronholm, M., Sainio, M., Zhao, F., Heiska, L., Vaheri, A., and Carpen, O. (1999). Homotypic and heterotypic interaction of the neurofibromatosis 2 tumor suppressor protein merlin and the ERM protein ezrin. J Cell Sci *112 (Pt 6)*, 895-904.

Gutmann, D.H., Wright, D.E., Geist, R.T., and Snider, W.D. (1995). Expression of the neurofibromatosis 2 (NF2) gene isoforms during rat embryonic development. Hum Mol Genet *4*, 471-478.

Halder, G., and Johnson, R.L. (2011). Hippo signaling: growth control and beyond. Development *138*, 9-22.

Hamaratoglu, F., Willecke, M., Kango-Singh, M., Nolo, R., Hyun, E., Tao, C., Jafar-Nejad, H., and Halder, G. (2006). The tumour-suppressor genes NF2/Merlin and Expanded act through Hippo signalling to regulate cell proliferation and apoptosis. Nat Cell Biol *8*, 27-36.

Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of cancer: the next generation. Cell *144*, 646-674.

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

Hart, T., Brown, K.R., Sircoulomb, F., Rottapel, R., and Moffat, J. (2014). Measuring error rates in genomic perturbation screens: gold standards for human functional genomics. Mol Syst Biol *10*, 733.

Hart, T., Chandrashekhar, M., Aregger, M., Steinhart, Z., Brown, K.R., MacLeod, G., Mis, M., Zimmermann, M., Fradet-Turcotte, A., Sun, S.*, et al.* (2015). High-Resolution CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities. Cell *163*, 1515-1526.

Harvey, K.F., Zhang, X., and Thomas, D.M. (2013). The Hippo pathway and human cancer. Nat Rev Cancer *13*, 246-257.

Hattangadi, S.M., Martinez-Morilla, S., Patterson, H.C., Shi, J., Burke, K., Avila-Figueroa, A., Venkatesan, S., Wang, J., Paulsen, K., Gorlich, D.*, et al.* (2014). Histones to the cytosol: exportin 7 is essential for normal terminal erythroid nuclear maturation. Blood *124*, 1931-1940.

He, M., Zhou, Z., Shah, A.A., Hong, Y., Chen, Q., and Wan, Y. (2016). New insights into posttranslational modifications of Hippo pathway in carcinogenesis and therapeutics. Cell Div *11*, 4.

Hillmer, R.E., and Link, B.A. (2019). The Roles of Hippo Signaling Transducers Yap and Taz in Chromatin Remodeling. Cells *8*.

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

Hong, A.W., Meng, Z., Plouffe, S.W., Lin, Z., Zhang, M., and Guan, K.L. (2020). Critical roles of phosphoinositides and NF2 in Hippo pathway regulation. Genes Dev *34*, 511-525.

Hu, X., Xing, Y., Fu, X., Yang, Q., Ren, L., Wang, Y., Li, Q., Li, J., and Zhang, L. (2020). NCAPG Dynamically Coordinates the Myogenesis of Fetal Bovine Tissue by Adjusting Chromatin Accessibility. Int J Mol Sci *21*.

Huggler, K.S., Rossiter, N.J., Flickinger, K.M., and Cantor, J.R. (2022). CRISPR/Cas9 Screening to Identify Conditionally Essential Genes in Human Cell Lines. Methods Mol Biol *2377*, 29-42.

Innes, A.J., Sun, B., Wagner, V., Brookes, S., McHugh, D., Pombo, J., Porreca, R.M., Dharmalingam, G., Vernia, S., Zuber, J.*, et al.* (2021). XPO7 is a tumor suppressor regulating p21(CIP1)-dependent senescence. Genes Dev *35*, 379-391.

Jain, A.K., Allton, K., Iacovino, M., Mahen, E., Milczarek, R.J., Zwaka, T.P., Kyba, M., and Barton, M.C. (2012). p53 regulates cell cycle and microRNAs to promote differentiation of human embryonic stem cells. PLoS Biol *10*, e1001268.

Jolma, A., Zhang, J., Mondragon, E., Morgunova, E., Kivioja, T., Laverty, K.U., Yin, Y., Zhu, F., Bourenkov, G., Morris, Q.*, et al.* (2020). Binding specificities of human RNA-binding proteins toward structured and linear RNA sequences. Genome Res *30*, 962-973.

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res *44*, D457-462.

Kawano, K., Doucet, A.J., Ueno, M., Kariya, R., An, W., Marzetta, F., Kuroki, M., Turelli, P., Sukegawa, S., Okada, S.*, et al.* (2018). HIV-1 Vpr and p21 restrict LINE-1 mobility. Nucleic Acids Res *46*, 8454-8470.

Kim, K., Heo, K., Choi, J., Jackson, S., Kim, H., Xiong, Y., and An, W. (2012). Vpr-binding protein antagonizes p53-mediated transcription via direct interaction with H3 tail. Mol Cell Biol *32*, 783-796.

Kim, K., Kim, J.M., Kim, J.S., Choi, J., Lee, Y.S., Neamati, N., Song, J.S., Heo, K., and An, W. (2013). VprBP has intrinsic kinase activity targeting histone H2A and represses gene transcription. Mol Cell *52*, 459-467.

Klijn, C., Durinck, S., Stawiski, E.W., Haverty, P.M., Jiang, Z., Liu, H., Degenhardt, J., Mayba, O., Gnad, F., Liu, J.*, et al.* (2015). A comprehensive transcriptional portrait of human cancer cell lines. Nat Biotechnol *33*, 306-312.

Lallemand, D., Curto, M., Saotome, I., Giovannini, M., and McClatchey, A.I. (2003). NF2 deficiency promotes tumorigenesis and metastasis by destabilizing adherens junctions. Genes Dev *17*, 1090-1100.

Lallemand, D., Saint-Amaux, A.L., and Giovannini, M. (2009). Tumor-suppression functions of merlin are independent of its role as an organizer of the actin cytoskeleton in Schwann cells. J Cell Sci *122*, 4141-4149.

Laulajainen, M., Muranen, T., Carpen, O., and Gronholm, M. (2008). Protein kinase A-mediated phosphorylation of the NF2 tumor suppressor protein merlin at serine 10 affects the actin cytoskeleton. Oncogene *27*, 3233-3243.

Levine, A.J., and Berger, S.L. (2017). The interplay between epigenetic changes and the p53 protein in stem cells. Genes Dev *31*, 1195-1201.

Li, W., Cooper, J., Karajannis, M.A., and Giancotti, F.G. (2012). Merlin: a tumour suppressor with functions at the cell cortex and in the nucleus. EMBO Rep *13*, 204-215.

Li, W., Cooper, J., Zhou, L., Yang, C., Erdjument-Bromage, H., Zagzag, D., Snuderl, M., Ladanyi, M., Hanemann, C.O., Zhou, P.*, et al.* (2014a). Merlin/NF2 loss-driven tumorigenesis linked to CRL4(DCAF1)-mediated inhibition of the hippo pathway kinases Lats1 and 2 in the nucleus. Cancer Cell *26*, 48-60.

Li, W., and Giancotti, F.G. (2010). Merlin's tumor suppression linked to inhibition of the E3 ubiquitin ligase CRL4 (DCAF1). Cell Cycle *9*, 4433-4436.

Li, W., Xu, H., Xiao, T., Cong, L., Love, M.I., Zhang, F., Irizarry, R.A., Liu, J.S., Brown, M., and Liu, X.S. (2014b). MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. Genome Biol *15*, 554.

Li, W., You, L., Cooper, J., Schiavon, G., Pepe-Caprio, A., Zhou, L., Ishii, R., Giovannini, M., Hanemann, C.O., Long, S.B.*, et al.* (2010). Merlin/NF2 suppresses tumorigenesis by inhibiting the E3 ubiquitin ligase CRL4(DCAF1) in the nucleus. Cell *140*, 477-490.

Li, Y., Zhou, H., Li, F., Chan, S.W., Lin, Z., Wei, Z., Yang, Z., Guo, F., Lim, C.J., Xing, W.*, et al.* (2015). Angiomotin binding-induced activation of Merlin/NF2 in the Hippo pathway. Cell Res *25*, 801-817.

Lian, I., Kim, J., Okazawa, H., Zhao, J., Zhao, B., Yu, J., Chinnaiyan, A., Israel, M.A., Goldstein, L.S., Abujarour, R.*, et al.* (2010). The role of YAP transcription coactivator in regulating stem cell self-renewal and differentiation. Genes Dev *24*, 1106-1118.

Lima, A.F., May, G., Diaz-Colunga, J., Pedreiro, S., Paiva, A., Ferreira, L., Enver, T., Iborra, F.J., and Pires das Neves, R. (2018). Osmotic modulation of chromatin impacts on efficiency and kinetics of cell fate modulation. Sci Rep *8*, 7210.

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

Liu, Y., and Kulesz-Martin, M. (2001). p53 protein at the hub of cellular DNA damage response pathways through sequence-specific and non-sequence-specific DNA binding. Carcinogenesis *22*, 851-860.

Liu, Z., Wei, Y., Zhang, L., Yee, P.P., Johnson, M., Zhang, X., Gulley, M., Atkinson, J.M., Trebak, M., Wang, H.G.*, et al.* (2019). Induction of store-operated calcium entry (SOCE) suppresses glioblastoma growth by inhibiting the Hippo pathway transcriptional coactivators YAP/TAZ. Oncogene *38*, 120-139.

Luijsterburg, M.S., Goedhart, J., Moser, J., Kool, H., Geverts, B., Houtsmuller, A.B., Mullenders, L.H., Vermeulen, W., and van Driel, R. (2007). Dynamic in vivo interaction of DDB2 E3 ubiquitin ligase with UV-damaged DNA is independent of damage-recognition protein XPC. J Cell Sci *120*, 2706-2716.

Lutzker, S.G., and Levine, A.J. (1996). A functionally inactive p53 protein in teratocarcinoma cells is activated by either DNA damage or cellular differentiation. Nat Med *2*, 804-810.

Ma, S., Meng, Z., Chen, R., and Guan, K.L. (2019). The Hippo Pathway: Biology and Pathophysiology. Annu Rev Biochem *88*, 577-604.

Mair, B., Aregger, M., Tong, A.H.Y., Chan, K.S.K., and Moffat, J. (2022). A Method to Map Gene Essentiality of Human Pluripotent Stem Cells by Genome-Scale CRISPR Screens with Inducible Cas9. Methods Mol Biol *2377*, 1-27.

Mao, B., Gao, Y., Bai, Y., and Yuan, Z. (2015). Hippo signaling in stress response and homeostasis maintenance. Acta Biochim Biophys Sin (Shanghai) *47*, 2-9.

McClatchey, A.I., and Giovannini, M. (2005). Membrane organization and tumorigenesis--the NF2 tumor suppressor, Merlin. Genes Dev *19*, 2265-2277.

McClatchey, A.I., and Yap, A.S. (2012). Contact inhibition (of proliferation) redux. Curr Opin Cell Biol *24*, 685-694.

Meng, J.J., Lowrie, D.J., Sun, H., Dorsey, E., Pelton, P.D., Bashour, A.M., Groden, J., Ratner, N., and Ip, W. (2000). Interaction between two isoforms of the NF2 tumor suppressor protein, merlin, and between merlin and ezrin, suggests modulation of ERM proteins by merlin. J Neurosci Res *62*, 491-502.

Miyoshi, T., Makino, T., and Moran, J.V. (2019). Poly(ADP-Ribose) Polymerase 2 Recruits Replication Protein A to Sites of LINE-1 Integration to Facilitate Retrotransposition. Mol Cell *75*, 1286-1298 e1212.

Moldovan, J.B., and Moran, J.V. (2015). The Zinc-Finger Antiviral Protein ZAP Inhibits LINE and Alu Retrotransposition. PLoS Genet *11*, e1005121.

Mori, T., Gotoh, S., Shirakawa, M., and Hakoshima, T. (2014). Structural basis of DDB1-and-Cullin 4-associated Factor 1 (DCAF1) recognition by merlin/NF2 and its implication in tumorigenesis by CD44-mediated inhibition of merlin suppression of DCAF1 function. Genes Cells *19*, 603-619.

Moroishi, T., Hansen, C.G., and Guan, K.L. (2015). The emerging roles of YAP and TAZ in cancer. Nat Rev Cancer *15*, 73-79.

Morrison, H., Sherman, L.S., Legg, J., Banine, F., Isacke, C., Haipek, C.A., Gutmann, D.H., Ponta, H., and Herrlich, P. (2001). The NF2 tumor suppressor gene product, merlin, mediates contact inhibition of growth through interactions with CD44. Genes Dev *15*, 968-980.

Mota, M., and Shevde, L.A. (2020). Merlin regulates signaling events at the nexus of development and cancer. Cell Commun Signal *18*, 63.

Naxerova, K., Di Stefano, B., Makofske, J.L., Watson, E.V., de Kort, M.A., Martin, T.D., Dezfulian, M., Ricken, D., Wooten, E.C., Kuroda, M.I.*, et al.* (2021). Integrated loss- and gain-of-function screens define a core network governing human embryonic stem cell behavior. Genes Dev *35*, 1527-1547.

Okada, T., Lopez-Lago, M., and Giancotti, F.G. (2005). Merlin/NF-2 mediates contact inhibition of growth by suppressing recruitment of Rac to the plasma membrane. J Cell Biol *171*, 361-371.

Pan, D. (2010). The hippo signaling pathway in development and cancer. Dev Cell *19*, 491-505.

Pearson, M.A., Reczek, D., Bretscher, A., and Karplus, P.A. (2000). Structure of the ERM protein moesin reveals the FERM domain fold masked by an extended actin binding tail domain. Cell *101*, 259-270.

Perrimon, N., Pitsouli, C., and Shilo, B.Z. (2012). Signaling mechanisms controlling cell fate and embryonic patterning. Cold Spring Harb Perspect Biol *4*, a005975.

Pijuan-Galito, S., Tamm, C., and Anneren, C. (2014). Serum Inter-alpha-inhibitor activates the Yes tyrosine kinase and YAP/TEAD transcriptional complex in mouse embryonic stem cells. J Biol Chem *289*, 33492-33502.

Pines, A., Vrouwe, M.G., Marteijn, J.A., Typas, D., Luijsterburg, M.S., Cansoy, M., Hensbergen, P., Deelder, A., de Groot, A., Matsumoto, S.*, et al.* (2012). PARP1 promotes nucleotide excision repair through DDB2 stabilization and recruitment of ALC1. J Cell Biol *199*, 235-249.

Plouffe, S.W., Hong, A.W., and Guan, K.L. (2015). Disease implications of the Hippo/YAP pathway. Trends Mol Med *21*, 212-222.

Qi, Q., Liu, X., Brat, D.J., and Ye, K. (2014). Merlin sumoylation is required for its tumor suppressor activity. Oncogene *33*, 4893-4903.

Ran, F.A., Hsu, P.D., Lin, C.Y., Gootenberg, J.S., Konermann, S., Trevino, A.E., Scott, D.A., Inoue, A., Matoba, S., Zhang, Y*., et al.* (2013a). Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. Cell *154*, 1380-1389.

Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013b). Genome engineering using the CRISPR-Cas9 system. Nat Protoc *8*, 2281-2308.

Rape, M. (2018). Ubiquitylation at the crossroads of development and disease. Nat Rev Mol Cell Biol *19*, 59-70.

Richardson, S.R., Narvaiza, I., Planegger, R.A., Weitzman, M.D., and Moran, J.V. (2014). APOBEC3A deaminates transiently exposed single-strand DNA during LINE-1 retrotransposition. Elife *3*, e02008.

Robbez-Masson, L., Tie, C.H.C., Conde, L., Tunbak, H., Husovsky, C., Tchasovnikarova, I.A., Timms, R.T., Herrero, J., Lehner, P.J., and Rowe, H.M. (2018). The HUSH complex cooperates with TRIM28 to repress young retrotransposons and new genes. Genome Res *28*, 836-845.

Robellet, X., Vanoosthuyse, V., and Bernard, P. (2017). The loading of condensin in the context of chromatin. Curr Genet *63*, 577-589.

Rong, R., Surace, E.I., Haipek, C.A., Gutmann, D.H., and Ye, K. (2004). Serine 518 phosphorylation modulates merlin intramolecular association and binding to critical effectors important for NF2 growth suppression. Oncogene *23*, 8447-8454.

Rouleau, G.A., Merel, P., Lutchman, M., Sanson, M., Zucman, J., Marineau, C., Hoang-Xuan, K., Demczuk, S., Desmaze, C., Plougastel, B.*, et al.* (1993). Alteration in a new gene encoding a putative membrane-organizing protein causes neuro-fibromatosis type 2. Nature *363*, 515-521.

Sanjana, N.E., Shalem, O., and Zhang, F. (2014). Improved vectors and genome-wide libraries for CRISPR screening. Nat Methods *11*, 783-784.

Schabla, N.M., Mondal, K., and Swanson, P.C. (2019). DCAF1 (VprBP): emerging physiological roles for a unique dual-service E3 ubiquitin ligase substrate receptor. J Mol Cell Biol *11*, 725-735.

Seczynska, M., Bloor, S., Cuesta, S.M., and Lehner, P.J. (2021). Genome surveillance by HUSH-mediated silencing of intronless mobile elements. Nature.

Shalem, O., Sanjana, N.E., Hartenian, E., Shi, X., Scott, D.A., Mikkelson, T., Heckl, D., Ebert, B.L., Root, D.E., Doench, J.G.*, et al.* (2014). Genome-scale CRISPR-Cas9 knockout screening in human cells. Science *343*, 84-87.

Sherman, L., Xu, H.M., Geist, R.T., Saporito-Irwin, S., Howells, N., Ponta, H., Herrlich, P., and Gutmann, D.H. (1997). Interdomain binding mediates tumor growth suppression by the NF2 gene product. Oncogene *15*, 2505-2509.

Shilatifard, A. (2006). Chromatin modifications by methylation and ubiquitination: implications in the regulation of gene expression. Annu Rev Biochem *75*, 243-269.

Shimizu, T., Seto, A., Maita, N., Hamada, K., Tsukita, S., Tsukita, S., and Hakoshima, T. (2002). Structural basis for neurofibromatosis type 2. Crystal structure of the merlin FERM domain. J Biol Chem *277*, 10332-10336.

Skowronski, J., and Singer, M.F. (1985). Expression of a cytoplasmic LINE-1 transcript is regulated in a human teratocarcinoma cell line. Proc Natl Acad Sci U S A *82*, 6050-6054.

Smith, T.W. (1989). The fundamental mechanism of inotropic action of digitalis. Therapie *44*, 431-435.

Stamenkovic, I., and Yu, Q. (2010). Merlin, a "magic" linker between extracellular cues and intracellular signaling pathways that regulate cell motility, proliferation, and survival. Curr Protein Pept Sci *11*, 471-484.

Stokowski, R.P., and Cox, D.R. (2000). Functional analysis of the neurofibromatosis type 2 protein by means of disease-causing point mutations. Am J Hum Genet *66*, 873-891.

Takahashi, A., Ohtani, N., Yamakoshi, K., Iida, S., Tahara, H., Nakayama, K., Nakayama, K.I., Ide, T., Saya, H., and Hara, E. (2006). Mitogenic signalling and the p16INK4a-Rb pathway cooperate to enforce irreversible cellular senescence. Nat Cell Biol *8*, 1291-1297.

Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell *126*, 663-676.

Tamm, C., Bower, N., and Anneren, C. (2011). Regulation of mouse embryonic stem cell self-renewal by a Yes-YAP-TEAD2 signaling pathway downstream of LIF. J Cell Sci *124*, 1136-1144.

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

Taylor, M.S., LaCava, J., Dai, L., Mita, P., Burns, K.H., Rout, M.P., and Boeke, J.D. (2016). Characterization of L1-Ribonucleoprotein Particles. Methods Mol Biol *1400*, 311-338.

Taylor, M.S., LaCava, J., Mita, P., Molloy, K.R., Huang, C.R., Li, D., Adney, E.M., Jiang, H., Burns, K.H., Chait, B.T.*, et al.* (2013). Affinity proteomics reveals human host factors implicated in discrete stages of LINE-1 retrotransposition. Cell *155*, 1034-1048.

Tchasovnikarova, I.A., Timms, R.T., Matheson, N.J., Wals, K., Antrobus, R., Gottgens, B., Dougan, G., Dawson, M.A., and Lehner, P.J. (2015). GENE SILENCING. Epigenetic silencing by the HUSH complex mediates position-effect variegation in human cells. Science *348*, 1481-1485.

Thadani, R., Uhlmann, F., and Heeger, S. (2012). Condensin, chromatin crossbarring and chromosome condensation. Curr Biol *22*, R1012-1021.

Tiwari, B., Jones, A.E., Caillet, C.J., Das, S., Royer, S.K., and Abrams, J.M. (2020). p53 directly represses human LINE1 transposons. Genes Dev *34*, 1439-1451.

Toufektchan, E., and Toledo, F. (2018). The Guardian of the Genome Revisited: p53 Downregulates Genes Required for Telomere Maintenance, DNA Repair, and Centromere Structure. Cancers (Basel) *10*.

Trofatter, J.A., MacCollin, M.M., Rutter, J.L., Murrell, J.R., Duyao, M.P., Parry, D.M., Eldridge, R., Kley, N., Menon, A.G., Pulaski, K.*, et al.* (1993). A novel moesin-, ezrin-, radixin-like gene is a candidate for the neurofibromatosis 2 tumor suppressor. Cell *72*, 791-800.

Victorelli, S., and Passos, J.F. (2017). Telomeres and Cell Senescence - Size Matters Not. EBioMedicine *21*, 14-20.

Wang, C., and Goff, S.P. (2017). Differential control of retrovirus silencing in embryonic cells by proteasomal regulation of the ZFP809 retroviral repressor. Proc Natl Acad Sci U S A *114*, E922-E930.

Wang, T., Birsoy, K., Hughes, N.W., Krupczak, K.M., Post, Y., Wei, J.J., Lander, E.S., and Sabatini, D.M. (2015). Identification and characterization of essential genes in the human genome. Science *350*, 1096-1101.

Wang, T., Wei, J.J., Sabatini, D.M., and Lander, E.S. (2014). Genetic screens in human cells using the CRISPR-Cas9 system. Science *343*, 80-84.

Ward, J.R., Crawford, B., Nock, S., Frisbie, T.J., Moran, J.V., and Longworth, M.S. (in preparation). LINE-1 induces cytoplasmic Super Condensin complex formation and condensin-mediated Type I IFN expression. In preparation.

Ward, J.R., Vasu, K., Deutschman, E., Halawani, D., Larson, P.A., Zhang, D., Willard, B., Fox, P.L., Moran, J.V., and Longworth, M.S. (2017). Condensin II and GAIT

complexes cooperate to restrict LINE-1 retrotransposition in epithelial cells. PLoS Genet *13*, e1007051.

Warkocki, Z., Krawczyk, P.S., Adamska, D., Bijata, K., Garcia-Perez, J.L., and Dziembowski, A. (2018). Uridylation by TUT4/7 Restricts Retrotransposition of Human LINE-1s. Cell *174*, 1537-1548 e1529.

Wei, Y., Yee, P.P., Liu, Z., Zhang, L., Guo, H., Zheng, H., Anderson, B., Gulley, M., and Li, W. (2020). NEDD4L-mediated Merlin ubiquitination facilitates Hippo pathway activation. EMBO Rep *21*, e50642.

Wolf, D., and Goff, S.P. (2007). TRIM28 mediates primer binding site-targeted silencing of murine leukemia virus in embryonic cells. Cell *131*, 46-57.

Wolf, D., and Goff, S.P. (2009). Embryonic stem cells use ZFP809 to silence retroviral DNAs. Nature *458*, 1201-1204.

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

Xiao, G.H., Gallagher, R., Shetler, J., Skele, K., Altomare, D.A., Pestell, R.G., Jhanwar, S., and Testa, J.R. (2005). The NF2 tumor suppressor gene product, merlin, inhibits cell proliferation and cell cycle progression by repressing cyclin D1 expression. Mol Cell Biol *25*, 2384-2394.

Xiao, Y., Hill, M.C., Zhang, M., Martin, T.J., Morikawa, Y., Wang, S., Moise, A.R., Wythe, J.D., and Martin, J.F. (2018). Hippo Signaling Plays an Essential Role in Cell State Transitions during Cardiac Fibroblast Development. Dev Cell *45*, 153-169 e156.

Yan, F., Qian, M., He, Q., Zhu, H., and Yang, B. (2020). The posttranslational modifications of Hippo-YAP pathway in cancer. Biochim Biophys Acta Gen Subj *1864*, 129397.

Yu, F.X., and Guan, K.L. (2013). The Hippo pathway: regulators and regulations. Genes Dev *27*, 355-371.

Zhang, A., Dong, B., Doucet, A.J., Moldovan, J.B., Moran, J.V., and Silverman, R.H. (2014a). RNase L restricts the mobility of engineered retrotransposons in cultured human cells. Nucleic Acids Res *42*, 3803-3820.

Zhang, N., Bai, H., David, K.K., Dong, J., Zheng, Y., Cai, J., Giovannini, M., Liu, P., Anders, R.A., and Pan, D. (2010). The Merlin/NF2 tumor suppressor functions through the YAP oncoprotein to regulate tissue homeostasis in mammals. Dev Cell *19*, 27-38.

Zhang, X., Wang, H., Han, Y., Zhu, M., Song, Z., Zhan, D., and Jia, J. (2020). NCAPG Induces Cell Proliferation in Cardia Adenocarcinoma via PI3K/AKT Signaling Pathway. Onco Targets Ther *13*, 11315-11326.

Zhang, Z.N., Chung, S.K., Xu, Z., and Xu, Y. (2014b). Oct4 maintains the pluripotency of human embryonic stem cells by inactivating p53 through Sirt1-mediated deacetylation. Stem Cells *32*, 157-165.

Zhao, B., Li, L., Wang, L., Wang, C.Y., Yu, J., and Guan, K.L. (2012). Cell detachment activates the Hippo pathway via cytoskeleton reorganization to induce anoikis. Genes Dev *26*, 54-68.

Zhao, B., Wei, X., Li, W., Udan, R.S., Yang, Q., Kim, J., Xie, J., Ikenoue, T., Yu, J., Li, L.*, et al.* (2007). Inactivation of YAP oncoprotein by the Hippo pathway is involved in cell contact inhibition and tissue growth control. Genes Dev *21*, 2747-2761.

Zheng, Y., and Pan, D. (2019). The Hippo Signaling Pathway in Development and Disease. Dev Cell *50*, 264-282.

Zhu, Q., Battu, A., Ray, A., Wani, G., Qian, J., He, J., Wang, Q.E., and Wani, A.A. (2015). Damaged DNA-binding protein down-regulates epigenetic mark H3K56Ac through histone deacetylase 1 and 2. Mutat Res *776*, 16-23.

Zhu, Y., Wang, G.Z., Cingoz, O., and Goff, S.P. (2018). NP220 mediates silencing of unintegrated retroviral DNA. Nature *564*, 278-282.

Zoch, A., Mayerl, S., Schulz, A., Greither, T., Frappart, L., Rubsam, J., Heuer, H., Giovannini, M., and Morrison, H. (2015). Merlin Isoforms 1 and 2 Both Act as Tumour Suppressors and Are Required for Optimal Sperm Maturation. PLoS One *10*, e0129151.

Appendix

The data presented in the Appendix remain preliminary and require follow-up experiments. Optimization of viral transduction efficiency for the L1-REPEL maintenance screen was performed in collaboration with Maria Virgilio. Preliminary small molecule screens were performed by Dr. Aurelien Doucet and Dr. Peter Larson. I performed all experiments and analyses discussed in the Appendix.

A GeCKO-based Screen to Identify L1-REPEL Maintenance Factors

To identify factors necessary to maintain L1-REPEL, we performed a preliminary GeCKO screen in pk5 cells for L1-*GFP* reactivation (Figure A.1). We reasoned that knockout of factors required to maintain L1-REPEL would result in expression of the EGFP reporter gene in pk5 cells. Consequently, genes necessary to initiate L1-REPEL might not be identified in the L1-REPEL maintenance screen. We sequenced EGFP expressing pk5 cells and used MAGeCK software to generate a ranked list of candidate genes (Figure A.1: C). Using a simplified count-based analysis, we identified nine candidate genes that each had two sgRNAs in multiple replicate samples (Figure A.1: D). Of the nine candidate genes, none were identified within multiple replicates of the GeCKO or Brunello screens. We also identified 417 candidates that had at least two different sgRNAs targeting a single gene across all four L1-REPEL maintenance experiments. Of the 417 genes, 32 genes were also identified within the GeCKO or

Brunello L1-REPEL screens (Figure A.1: E). Intriguingly, we did not find *NF2* in the L1-REPEL maintenance screen. This may suggest that *NF2* is necessary to initiate L1-REPEL in PA-1 cells. Also, since pk5 cells are derived from PA-1 cells, these results strongly argue against the false-positive identification of *NF2* in the L1-REPEL screens due to a cell-type specific phenotype caused by *NF2* knockout, such as a proliferative advantage.

There were several technical issues that contributed to low sorting efficiency within the L1-REPEL maintenance screen, including: (1) a high background of EGFP expressing cells; (2) cell deterioration and damage due to the lengthy sorting process; and (3) inefficient genomic DNA yield from the low numbers of sorted EGFP-positive cells. Consequently, the L1-REPEL maintenance screen lacked adequate power to identify robust candidate genes.

Future studies should aim to generate a cell line harboring a silenced, drug-selectable, L1-REPEL event, which could then be used to screen for L1-REPEL maintenance factors. To generate this cell line, it may be necessary to design an engineered L1 containing a dual function *EGFP*/*neomycin*^{R} reporter gene (L1-*neo/GFP*). PA-1 cells would be transfected with the engineered L1-*neo/GFP* construct to generate a population of cells containing silenced L1-*neo/GFP* retrotransposition events. Cells would then be treated with TSA to reverse L1-REPEL, and EGFP-positive cells would be FACS sorted into plates for clonal outgrowth. After clonal cell lines containing silenced *L1-neo/GFP* retrotransposition events are established, the cells would be transduced with lentiviral vectors with either the GeCKOv2 or Brunello sgRNA libraries. Once efficiently edited, the resultant cells would be subjected to G418-

selection to identify cells that are stably expressing the L1-delivered reporter gene. Notably, G418-resistant cells should also express the EGFP reporter gene. This approach addresses many of the technical issues hampering the initial L1-REPEL maintenance screen in pk5 cells.

L1-REPEL Candidate Gene Validation Assays

To validate other candidate genes identified in the GeCKO-based L1-REPEL screens (see Chapter 2), we cloned ~80 additional sgRNA-Cas9 knockout vectors and performed population "knockout" validation assays (Figure A.2). Population "knockout" of *DDB2*, *NCBP2, TP53, NCAPG, MPHOSPH8*, and *ATF7IP* promoted drug-resistant colony formation above background levels in the L1 retrotransposition assay, suggesting that some of these candidate genes may be necessary for efficient L1-REPEL (Figure A.2: B and C). Notably, these *NF2* validation assays utilized an sgRNA targeting exon 11 within the helical domain, whereas the previous sgRNAs, N19 and N60, target exon 4 within the FERM domain (see Figure 3.1). As knockout efficiency was not determined for these preliminary assays, this negative result is likely due to an inefficient sgRNA. It will be important to determine the editing efficiency and include N19 or N60 sgRNAs in future validation assays as a positive control (see below).

Furthermore, we also performed L1-REPEL maintenance assays in pc39 cells (Figure A.2: D). Population "knockout" of *MPHOSPH8*, *ATF7IP, NCBP2, TP53, DOT1L, EHMT1*, *and TAF5L* reactivated EGFP expression in pc39 cells above background levels, suggesting a role for these genes in maintaining L1-REPEL (Figure A.2: E). Together, these preliminary validation assays have identified several L1-REPEL factors that might be involved in the initiation and/or maintenance of L1-REPEL. A brief

description of the documented functions of each of the genes/proteins is provided below:

MPP8 (*MPHOSPH8)* and *ATF7IP*: encode proteins that are members of the heterochromatin-forming HUSH complex, which previously was implicated in retrotransposon silencing (Liu et al., 2018; Robbez-Masson et al., 2018; Seczynska et al., 2021; Zhu et al., 2018). The HUSH complex spreads heterochromatin via deposition of H3K9me3, acting to maintain transcriptional repression (Tchasovnikarova et al., 2015). As expected, “population knockout” of *MPHOSPH8* and *ATF7IP* in pc39 cells reactivated EGFP expression above background levels, suggesting a possible role in the maintenance of L1-REPEL (Figure A.2: E). Intriguingly, chaetocin, an inhibitor of the H3K9 histone methyltransferase SU(VAR)3-9 (Cherblanc et al., 2013; Greiner et al., 2005), also led to the reactivation of EGFP expression in pk5 and pc39 cells at levels comparable to TSA-treatment (see below).

DNA damage-binding protein 2 (*DDB2*): encodes a protein that recognizes DNA damage when complexed with DDB1, CUL4A, and PARP1 (Luijsterburg et al., 2007; Pines et al., 2012). Interestingly, DDB2 interacts with histone deacetylases (HDAC1 and HDAC2) to decrease H3K56 acetylation after UV-induced DNA damage (Zhu et al., 2015). Thus, the loss of DDB2 might hinder the DNA damage response at sites of TPRT, resulting in escape from L1-REPEL. Intriguingly, NF2/merlin also interacts with DDB1 and CUL4A, suggesting a possible interaction with DDB2.

Nuclear cap binding protein subunit 2 (*NCBP2*): encodes a member of the cap-binding complex (CBC), which co-transcriptionally binds to the 5’-m^7G cap of pre-mRNAs and is involved in splicing, translation, nonsense-mediated decay (NMD), miRNA-mediated gene silencing, and mRNA nuclear export. If L1-REPEL is dependent upon RNA-mediated gene silencing, it is plausible that loss of NCBP2 could impede efficient L1-REPEL in PA-1 cells. Interestingly, NCBP2 was previously identified as an L1 ORF1p-interacting protein by LC-MS/MS (Moldovan and Moran, 2015).

Condensin complex subunit 3 (*NCAPG*): encodes a subunit of condensin I, which mediates chromosome condensation (Robellet et al., 2017; Thadani et al., 2012). Recent work suggests that NCAPG acts as an oncogene by regulating the PI3K/AKT signaling pathway (Zhang et al., 2020). Intriguingly, condensin I and condensing II cooperate with the GAIT complex to restrict L1 retrotransposition in human colon adenocarcinoma cells (Ward et al., in preparation; Ward et al., 2017). Furthermore, studies using fetal bovine tissue revealed that NCAPG is involved in myoblast differentiation and chromatin accessibility (Hu et al., 2020). Thus, the loss of NCAPG may disrupt chromatin organization and accessibility, mitigating L1-REPEL in PA-1 cells.

Tumor protein 53 (*TP53*): encodes a tumor suppressor protein and is considered to be the “guardian of the genome” due to its roles in DNA damage repair, the cellular stress response, cell cycle progression, genomic stability, and stem cell maintenance (Toufektchan and Toledo, 2018). Embryonic carcinoma cells, including PA-1, express the TP53 protein several-fold higher than differentiated cells (Klijn et al., 2015; Lutzker and Levine, 1996). TP53 regulation is vital to maintain an epigenetic environment that promotes stem cell renewal (Levine and Berger, 2017). Notably, when retinoic acid is added to embryonic stem cell cultures, differentiation coincides with the acetylation of TP53 lysine residues (Jain et al., 2012; Zhang et al., 2014b). These data highlight TP53 as a possible L1-REPEL candidate because TP53 is highly expressed in multipotent embryonic carcinoma cells and undergoes down-regulation and/or modification in differentiated cells.

Intriguingly, tp53 knockout in zebrafish increased the retrotransposition efficiency of engineered L1s (Wylie et al., 2016). Moreover, tp53 knockout increased endogenous ORF1p expression and decreased H3K9me3 at endogenous L1 loci. More recently, TP53 was demonstrated to repress human L1s through direct interactions with the L1 5’ UTR (Tiwari et al., 2020). TP53 mutation also correlated with increased expression of endogenous L1s in human colon cancer tissues (Wylie et al., 2016). Intriguingly, TP53 exhibits both sequence-dependent

and sequence-independent DNA-binding activities, suggesting that it may be recruited to structures associated with DNA damage and repair (Liu and Kulesz-Martin, 2001). Thus, it is plausible that TP53 may be recruited to TPRT intermediates, which might be recognized as DNA lesions, to promote L1-REPEL in PA-1 cells. Future studies should determine whether TP53 influences L1 expression in human embryonic cells.

In sum, we performed a series of L1-REPEL validation assays that identified several high-confidence L1-REPEL factors involved in the initiation and/or maintenance of L1-REPEL. Our working model posits that L1-REPEL occurs through a two-step initiation and maintenance process to efficiently and stably silence L1-delivered reporter genes in PA-1 cells. Notably, our maintenance assay assumes that we are efficiently "knocking out" the candidate gene in a population of pk5 or pc39 cells. It is notable that we exploited a puromycin marker to select for cells expressing sgRNAs and Cas9, which should promote efficient editing within the puromycin resistant cell populations. However, to arrive at rigorous conclusions regarding the necessity of cellular factors in maintaining L1-REPEL, we need to ensure that functional protein is significantly diminished or absent in our pk5 or pc39 cell populations by western blot analyses. Quantification of endogenous protein would then serve as a proxy for editing efficiency.

Alternatively, genomic DNA from the puromycin resistant population of cells could be subjected to PCR amplification using primers flanking the sgRNA-target site followed by next-generation sequencing. The percentage of indel vs reference reads would represent the editing efficiency. Notably, the half-life of the candidate protein must be considered to account for protein turnover when generating "population knockout" pk5 or pc39 cells. An alternative, but perhaps more time consuming, approach to the population-based assay would be to generate independent pk5 or pc39 knockout

clones. Assessing EGFP expression in several knockout pk5 or pc39 cell lines then would provide a complementary and rigorous approach to determine whether a candidate factor is necessary to maintain L1-REPEL in PA-1 cells.

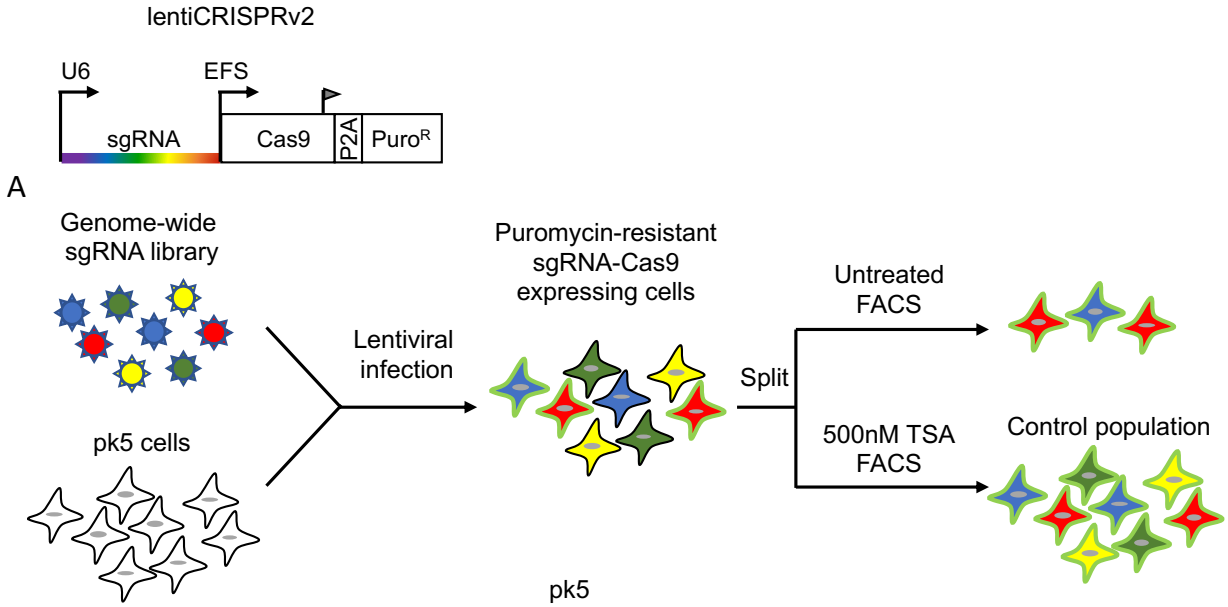
Epigenetic Marks Associated with L1-REPEL

To elucidate epigenetic marks associated with L1-REPEL, we performed a small molecule screen with drugs that specifically target epigenetic modifications associated with transcriptional regulation (Figure A.3). Pk5 and pc39 cells were treated with a variety of compounds that inhibit marks associated with transcriptional repression (Figure A.3: A). For example, histone deacetylation and methylation are two modifications associated with transcriptional repression. Histone deacetylase (HDAC) and histone methyltransferase (HMT) inhibitors are drugs that inhibit these repressive histone marks. DNA methylation at gene promoters is linked to transcriptional repression. DNA methyltransferase (DNMT) inhibitors block the deposition of DNA methylation. Thus, pk5 and pc39 cells were treated with HDAC, HMT, and DNMT inhibitors, then analyzed for EGFP expression (Figure A.3: A and B).

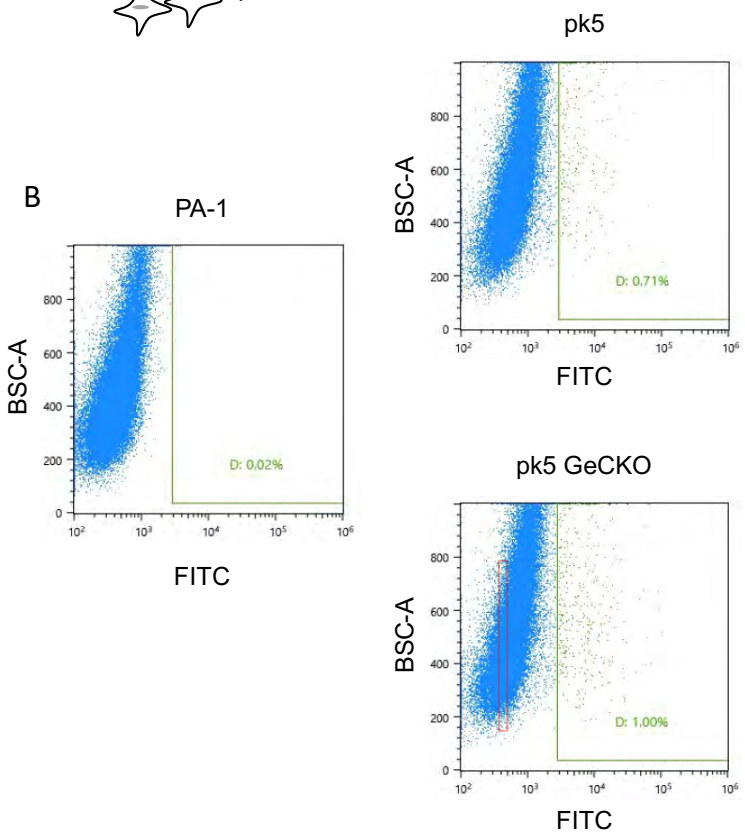
Pk5 and pc39 cells treated with HDAC inhibitors trichostatin A (TSA), valproic acid (VPA), and sodium butyrate (NaB) efficiently expressed EGFP, suggesting that class I HDAC activity may play a role in L1-REPEL (Figure A.3: B). Pk5 and pc39 treated with the HMT inhibitor chaetocin (Cha) expressed EGFP at levels similar to cells treated with HDAC inhibitors, suggesting that H3K9 histone methylation may play a role in L1-REPEL (See Chapter 4). Pk5 and pc39 cells treated with DNMT inhibitor 5-Azacytidine, but not 5-aza-2'-deoxycytidine, expressed EGFP in a dose dependent manner. The cytidine analog 5-azacytidine incorporates into RNA, whereas 5-aza-2'-deoxycytidine

incorporates into DNA. These results may suggest an RNA-dependent component to L1-REPL. In sum, these results suggest that redundant epigenetic silencing mechanisms may act to maintain transcriptional repression of L1-delivered reporter genes.

A



B



C

pk5 GeCKO screen

Rank	Gene	# sgRNAs
1	VAV2	2
2	CSPG5	1
3	ANGPTL4	1
4	SCLY	1
5	PCDHGB5	2
6	MS4A4A	2
7	mir-3186	1
8	NPPB	3
9	ATP6V0C	2
10	MYO5B	2
11	TRABD2A	3
12	SAC3D1	1
13	ZNF227	1
14	CAMK2N2	3
15	FAM129A	3
16	MON1B	2
17	METTL14	1
18	SLC35A3	1
19	CHSY3	3
20	USP28	2
21	FAM209A	2
22	ZNF747	3
23	TMEM14B	1
24	SEMA3D	1
25	MBL2	2

D

Genes with 2+ sgRNAs in 2+ replicates

Gene	sgRNAs	Replicates
SALL4	2	2
USP28	2	2
ATP6V0C	2	2
C8orf48	2	2
ECE2	2	2
MLL4	2	2
MS4A4A	2	2
POU3F1	2	2
PPIL4	2	2

E

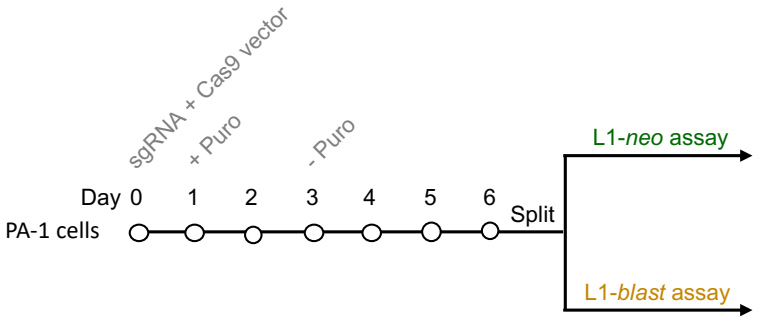
Genes with 2+ sgRNAs in 2+ screens

Rank	Gene	pk5 GeCKO	PA-1 GeCKO 1	PA-1 GeCKO 2	PA-1 Brunello
1	TMC6	2	2	1	1
2	RD3L	2	1	1	2
3	NMBR	2	1	0	2
4	SAP25	2	2	0	1
5	PIP4K2B	2	0	2	1
6	THSD7B	2	0	2	1
7	CHSY3	3	2	1	0
8	ARIH2	2	2	1	0
9	SEPTIN6	2	2	1	0
10	PTAR1	2	2	1	0
11	THBS1	2	2	1	0
12	KLK8	2	1	2	0
13	NUCB1	2	1	2	0
14	KRTAP4-2	2	1	2	0
15	ZFP62	2	1	2	0
16	PKDCC	2	1	2	0
17	ANKRD32	2	1	2	0
18	HBA1	2	1	2	0
19	GRIK2	2	2	0	0
20	KCNK1	2	2	0	0
21	SH3TC1	2	2	0	0
22	LPAR2	2	2	0	0
23	GGH	2	2	0	0
24	TTC19	2	2	0	0
25	SNRNP200	2	0	2	0
26	ST6GALNAC1	2	0	2	0
27	MANF	2	0	0	2
28	GAS2L3	2	0	0	2
29	ALKBH4	2	0	0	2
30	FBLN7	2	0	0	2
31	RTBDN	2	0	0	2
32	GAREM	2	0	0	2

Figure A.1: **GeCKO-based L1-REPEL maintenance screen in pk5 cells.**

(A) Schematic of the L1-REPEL maintenance screen. Day 21 lentiCRISPRv2-GeCKO transduced pk5 cells were FACS sorted for cells expressing the L1-*GFP* reporter. (B) Single, DAPI negative, cells were sorted based on EGFP expression in wild-type PA-1 cells (left), pk5 cells (middle) or pk5 lentiCRISPRv2-GeCKO infected cells (right). Green box: sorted EGFP (+) cells. Red box: sorted EGFP (-) cells. (C) MAGeCK-based identification of candidate genes from four independent L1-REPEL maintenance experiments. The top 25 genes are shown with the indicated number of enriched sgRNAs. The GeCKOv2 sgRNA library has 6 sgRNAs per gene. (D) Simplified-based identification of genes with 2 or more sgRNAs represented in two or more biological replicates. The number of sgRNAs and biological replicates is indicated (E) Simplified-based identification of genes with two or more sgRNAs in two or more screens. The number of identified sgRNAs is indicated.

A

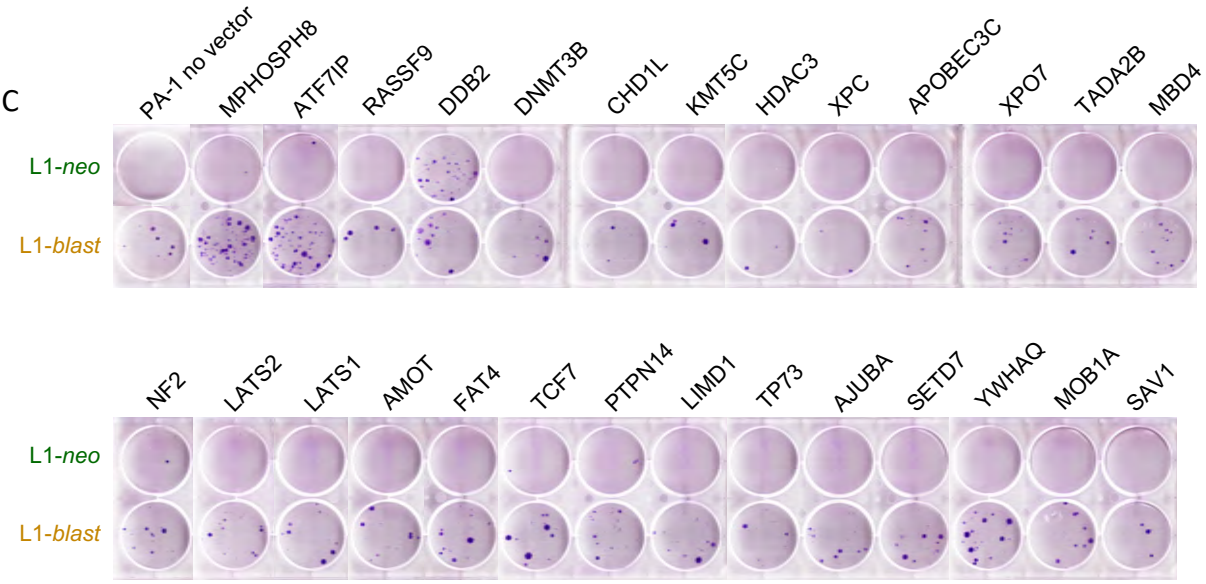


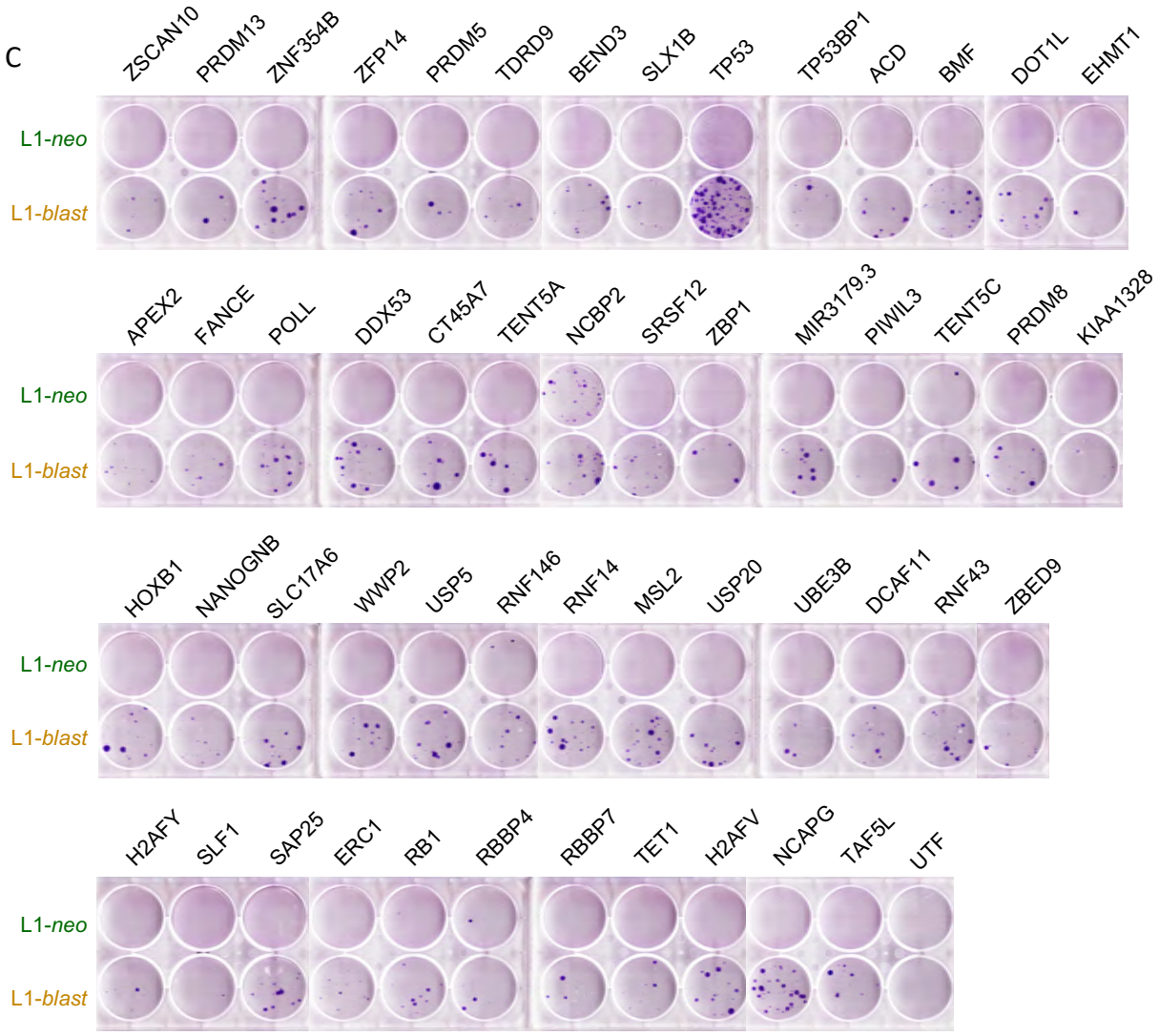
B

Candidate gene summary

MPHOSPH8	XPC	AMOT	YWHAQ	TDRD9	EHMT1	SRSF12	NANOGNB	UBE3B	RB1
ATF7IP	APOBEC3C	FAT4	MOB1A	BEND3	APEX2	ZBP1	SLC17A6	DCAF11	RBBP4
RASSF9	XPO7	TCF7	SAV1	SLX1B	FANCE	MIR3179.3	WWP2	RNF43	RBBP7
DDB2	TADA2B	PTPN14	ZSCAN10	TP53	POLL	PIWIL3	USP5	ZBED9	TET1
DNMT3B	MBD4	LIMD1	PRDM13	TP53BP1	DDX53	TENT5C	RNF146	H2AFY	H2AFV
CHD1L	NF2	TP73	ZNF354B	ACD	CT45A7	PRDM8	RNF14	SLF1	NCAPG
KMT5C	LATS2	AJUBA	ZFP14	BMF	TENT5A	KIAA1328	MSL2	SAP25	TAF5L
HDAC3	LATS1	SETD7	PRDM5	DOT1L	NCBP2	HOXB1	USP20	ERC1	

C




C
ZSCAN10
PRDM13
ZNF354B
ZFP14
PRDM5
TDRD9
BEND3
SLX1B
TP53
TP53BP1
ACD
BMF
DOT1L
EHMT1
L1-neo
L1-blast
APEX2
FANCE
POLL
DDX53
CT45A7
TENT5A
NCBP2
SRSF12
ZBP1
MIR3179.3
PIWIL3
TENT5C
PRDM8
KIAA1328
L1-neo
L1-blast
HOXB1
NANOGNB
SLC17A6
WWP2
USP5
RNF146
RNF14
MSL2
USP20
UBE3B
DCAF11
RNF43
ZBED9
L1-neo
L1-blast
H2AFY
SLF1
SAP25
ERC1
RB1
RBBP4
RBBP7
TET1
H2AFV
NCAPG
TAF5L
UTF
L1-neo
L1-blast

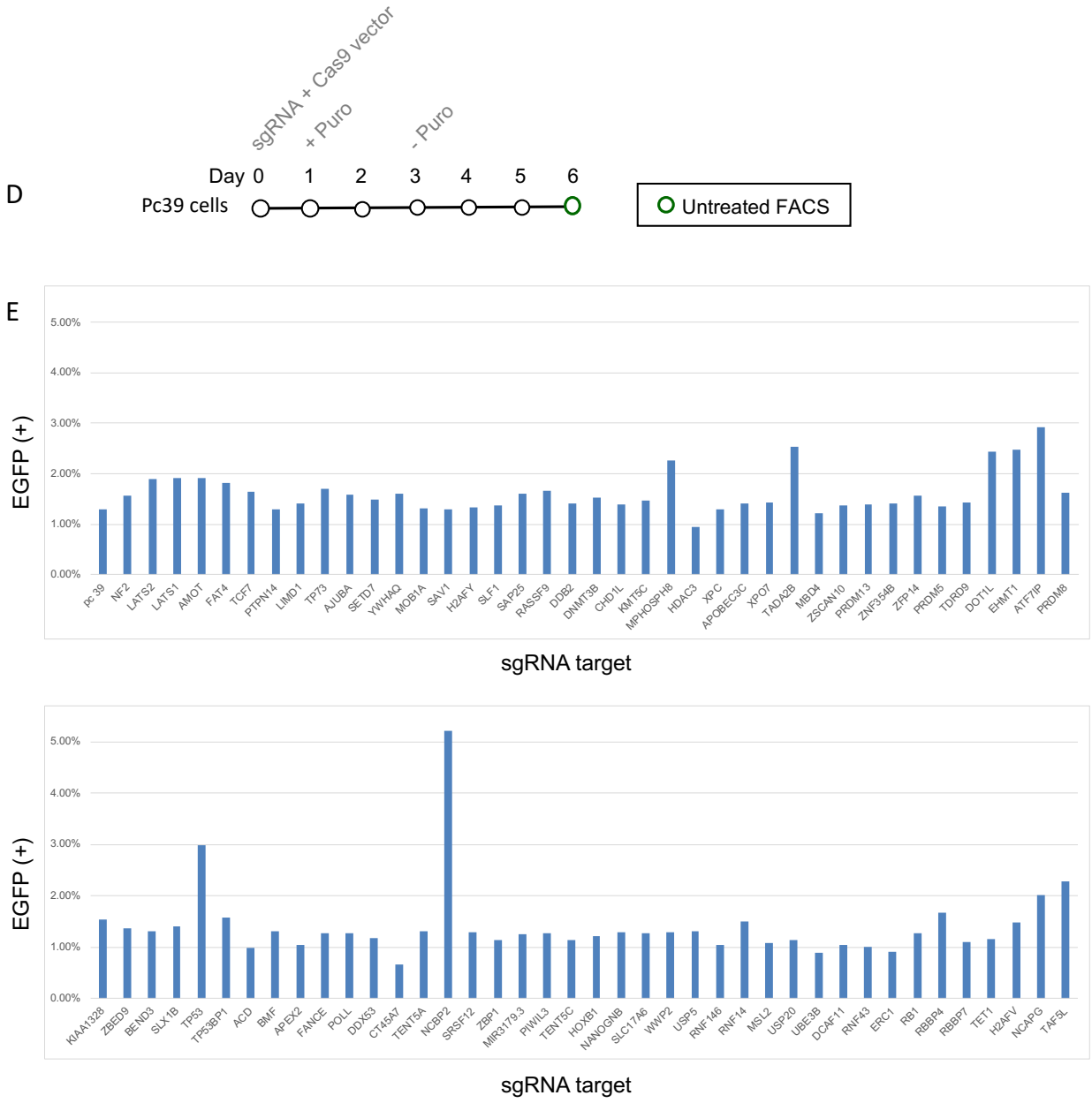


Figure A.2: Candidate gene validation assays.

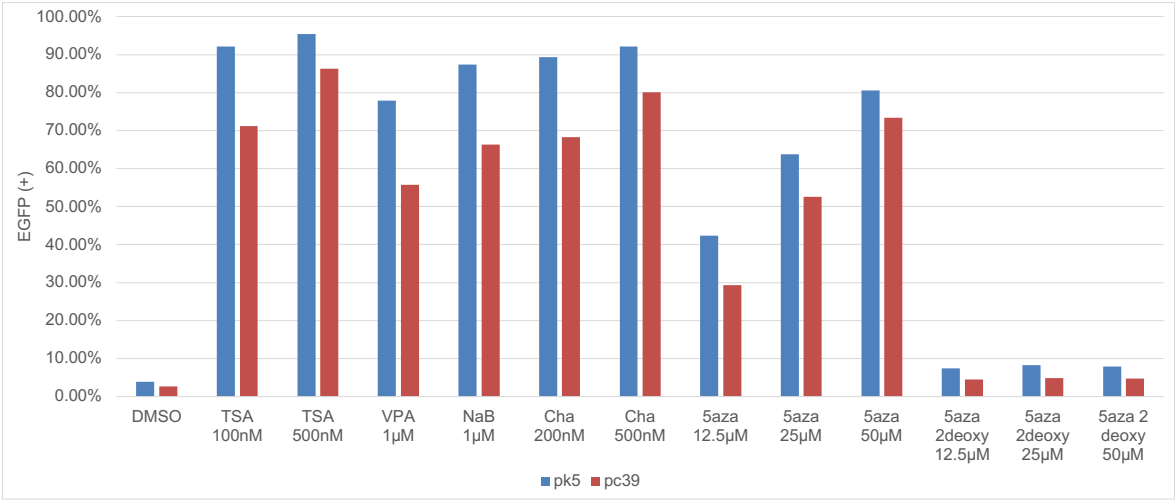
(A) Schematic of the candidate gene validation strategy. PA-1 cells were transfected with a Cas9 + sgRNA expressing vector targeting a specific candidate gene. Population "knockout" cells were then subjected to the L1 retrotransposition assay to determine L1-REPEL efficiency. (B) Summary table showing all the tested candidate genes. Population "knockout" of highlighted candidate genes exhibited a clear increase in neomycin-resistant or blasticidin-resistant foci formation. C) Results of the L1-*neo* (top)

and L1-*blast* (bottom) retrotransposition assays. Cells subjected to the L1-*neo* assay were transfected with JM101/L1.3. Cells subjected to the L1-*blast* assay were transfected with JJ101/L1.3. (D) Schematic of the candidate gene L1-REPEL maintenance assay. pc39 cells were transfected with a Cas9 + sgRNA expressing vector targeting a specific candidate gene. “Population knockout” cells were then subjected to flow cytometry to assess L1-*GFP* reporter gene expression. (E) Results of the candidate gene L1-REPEL maintenance assay. The y-axis indicates the percentage of EGFP-expressing cells. The x-axis indicates the sgRNA-targeted candidate gene.

A

	Compound	Activity
TSA	Trichostatin A	HDAC inhibitor (I & II)
VPA	Valproic acid	HDAC inhibitor
NaB	Sodium butyrate	HDAC inhibitor
Cha	Chaetocin	HMT inhibitor (H3K9)
5aza	5-Azacytidine	DNMT inhibitor cytidine analog (RNA)
5aza 2deoxy	5-aza-2′-deoxycytidine	DNMT inhibitor cytidine analog (DNA)

B



C

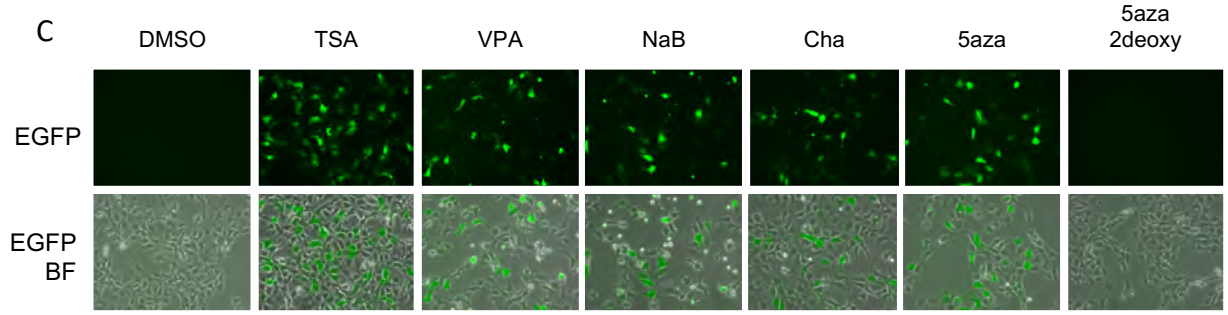


Figure A.3: Small molecule compounds that reverse L1-REPEL.

Small molecule compounds that inhibit histone deacetylase (HDAC), histone methyltransferase (HMT), and DNA methyltransferase (DNMT) activity were added to culture media for 18 hours. (A) Table describing small molecule compounds and their associated activities. (B) EGFP reactivation in pk5 and pc39 cells after 18 hours of treatment with the indicated compound. EGFP expression was determined by flow cytometry. (C) 20x images of pk5 cells after 18 hours of treatment with the indicated compound.

References

Cherblanc, F.L., Chapman, K.L., Brown, R., and Fuchter, M.J. (2013). Chaetocin is a nonspecific inhibitor of histone lysine methyltransferases. Nat Chem Biol *9*, 136-137.

Greiner, D., Bonaldi, T., Eskeland, R., Roemer, E., and Imhof, A. (2005). Identification of a specific inhibitor of the histone methyltransferase SU(VAR)3-9. Nat Chem Biol *1*, 143-145.

Hu, X., Xing, Y., Fu, X., Yang, Q., Ren, L., Wang, Y., Li, Q., Li, J., and Zhang, L. (2020). NCAPG Dynamically Coordinates the Myogenesis of Fetal Bovine Tissue by Adjusting Chromatin Accessibility. Int J Mol Sci *21*.

Jain, A.K., Allton, K., Iacovino, M., Mahen, E., Milczarek, R.J., Zwaka, T.P., Kyba, M., and Barton, M.C. (2012). p53 regulates cell cycle and microRNAs to promote differentiation of human embryonic stem cells. PLoS Biol *10*, e1001268.

Klijn, C., Durinck, S., Stawiski, E.W., Haverty, P.M., Jiang, Z., Liu, H., Degenhardt, J., Mayba, O., Gnad, F., Liu, J.*, et al.* (2015). A comprehensive transcriptional portrait of human cancer cell lines. Nat Biotechnol *33*, 306-312.

Levine, A.J., and Berger, S.L. (2017). The interplay between epigenetic changes and the p53 protein in stem cells. Genes Dev *31*, 1195-1201.

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

Liu, Y., and Kulesz-Martin, M. (2001). p53 protein at the hub of cellular DNA damage response pathways through sequence-specific and non-sequence-specific DNA binding. Carcinogenesis *22*, 851-860.

Luijsterburg, M.S., Goedhart, J., Moser, J., Kool, H., Geverts, B., Houtsmuller, A.B., Mullenders, L.H., Vermeulen, W., and van Driel, R. (2007). Dynamic in vivo interaction of DDB2 E3 ubiquitin ligase with UV-damaged DNA is independent of damage-recognition protein XPC. J Cell Sci *120*, 2706-2716.

Lutzker, S.G., and Levine, A.J. (1996). A functionally inactive p53 protein in teratocarcinoma cells is activated by either DNA damage or cellular differentiation. Nat Med *2*, 804-810.

Moldovan, J.B., and Moran, J.V. (2015). The Zinc-Finger Antiviral Protein ZAP Inhibits LINE and Alu Retrotransposition. PLoS Genet *11*, e1005121.

Pines, A., Vrouwe, M.G., Marteijn, J.A., Typas, D., Luijsterburg, M.S., Cansoy, M., Hensbergen, P., Deelder, A., de Groot, A., Matsumoto, S.*, et al.* (2012). PARP1 promotes nucleotide excision repair through DDB2 stabilization and recruitment of ALC1. J Cell Biol *199*, 235-249.

Robbez-Masson, L., Tie, C.H.C., Conde, L., Tunbak, H., Husovsky, C., Tchasovnikarova, I.A., Timms, R.T., Herrero, J., Lehner, P.J., and Rowe, H.M. (2018). The HUSH complex cooperates with TRIM28 to repress young retrotransposons and new genes. Genome Res *28*, 836-845.

Robellet, X., Vanoosthuyse, V., and Bernard, P. (2017). The loading of condensin in the context of chromatin. Curr Genet *63*, 577-589.

Seczynska, M., Bloor, S., Cuesta, S.M., and Lehner, P.J. (2021). Genome surveillance by HUSH-mediated silencing of intronless mobile elements. Nature.

Tchasovnikarova, I.A., Timms, R.T., Matheson, N.J., Wals, K., Antrobus, R., Gottgens, B., Dougan, G., Dawson, M.A., and Lehner, P.J. (2015). GENE SILENCING. Epigenetic silencing by the HUSH complex mediates position-effect variegation in human cells. Science *348*, 1481-1485.

Thadani, R., Uhlmann, F., and Heeger, S. (2012). Condensin, chromatin crossbarring and chromosome condensation. Curr Biol *22*, R1012-1021.

Tiwari, B., Jones, A.E., Caillet, C.J., Das, S., Royer, S.K., and Abrams, J.M. (2020). p53 directly represses human LINE1 transposons. Genes Dev *34*, 1439-1451.

Toufektchan, E., and Toledo, F. (2018). The Guardian of the Genome Revisited: p53 Downregulates Genes Required for Telomere Maintenance, DNA Repair, and Centromere Structure. Cancers (Basel) *10*.

Ward, J.R., Crawford, B., Nock, S., Frisbie, T.J., Moran, J.V., and Longworth, M.S. (in preparation). LINE-1 induces cytoplasmic Super Condensin complex formation and condensin-mediated Type I IFN expression. In preparation.

Ward, J.R., Vasu, K., Deutschman, E., Halawani, D., Larson, P.A., Zhang, D., Willard, B., Fox, P.L., Moran, J.V., and Longworth, M.S. (2017). Condensin II and GAIT complexes cooperate to restrict LINE-1 retrotransposition in epithelial cells. PLoS Genet *13*, e1007051.

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

Zhang, X., Wang, H., Han, Y., Zhu, M., Song, Z., Zhan, D., and Jia, J. (2020). NCAPG Induces Cell Proliferation in Cardia Adenocarcinoma via PI3K/AKT Signaling Pathway. Onco Targets Ther *13*, 11315-11326.

Zhang, Z.N., Chung, S.K., Xu, Z., and Xu, Y. (2014). Oct4 maintains the pluripotency of human embryonic stem cells by inactivating p53 through Sirt1-mediated deacetylation. Stem Cells *32*, 157-165.

Zhu, Q., Battu, A., Ray, A., Wani, G., Qian, J., He, J., Wang, Q.E., and Wani, A.A. (2015). Damaged DNA-binding protein down-regulates epigenetic mark H3K56Ac through histone deacetylase 1 and 2. Mutat Res *776*, 16-23.

Zhu, Y., Wang, G.Z., Cingoz, O., and Goff, S.P. (2018). NP220 mediates silencing of unintegrated retroviral DNA. Nature *564*, 278-282.