The interferon-stimulated gene product HELZ2 destabilizes human LINE-1 RNA to inhibit LINE-1 retrotransposition and the associated type I interferon response

Ahmad Luqman Bin Abdul Fatah

The interferon-stimulated gene product HELZ2 destabilizes human LINE-1 RNA to inhibit LINE-1 retrotransposition and the associated type I interferon response

By

Ahmad Luqman Bin Abdul Fatah

A dissertation submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

in

Life Sciences

in the Graduate School of Biostudies

of

Kyoto University

January 2023

Abstract

Interferon-stimulated genes (ISGs) are upregulated in response to pathogen infection as a defense mechanism; several ISG-encoded proteins were previously reported to inhibit mobilization of human Long Interspersed Element-1 (LINE-1 or L1) retrotransposons. Differential analyses of immunoprecipitation of L1 ORF1-encoded protein (ORF1p) followed by liquid chromatography-tandem mass spectrometry between the wild type and an RNA-binding mutant ORF1p complexes revealed an ISG proteins network associating with L1 ribonucleoprotein complexes. Among the ISGs, I focused on three proteins: 2'-5'-oligoadenylate synthetase-like (OASL), HECT and RLD domain containing E3 ubiquitin-protein ligase 5 (HERC5), and helicase with zinc finger 2 (HELZ2), and found that these proteins inhibit L1 retrotransposition at different stages of the L1 replicating cycle. OASL showed impairment in the ORF1p cytoplasmic foci formation and HERC5 reduced ORF1p steady state levels, while HELZ2 recognizes structures and/or sequences within the L1 5' untranslated region (UTR) to reduce L1 RNA, ORF1p, and ORF1p cytoplasmic foci levels. In agreement with previous studies, my results suggest that type I interferon expression is induced by overexpression of WT and reverse-transcriptase deficient L1s, which was abolished upon HELZ2 overexpression. Notably, type I interferon expression is enhanced when the ORF1p RNA-binding mutant was overexpressed, suggesting that the ORF1p RNA-binding shields L1 RNA from "triggering" the type I IFN upregulation. Taken together, these results suggest a negative feedback regulation of L1 retrotransposition by ISG proteins through different mechanisms.

Acknowledgement I

This thesis is based on the materials contained in the following scholarly paper:

Luqman-Fatah, A., Watanabe, Y., Ishikawa, F., Moran, J. V., and Miyoshi, T. The interferon stimulated gene-encoded protein HELZ2 inhibits human LINE-1 retrotransposition and LINE-1 RNA-mediated type I interferon induction. Nature Communications (In press).

All of the figures (except Figs. 1.1 and 1.2), tables and source data, including the respective legends were modified from this paper.

Acknowledgements II

Genggam bara api, biar sampai menjadi arang

七転び八起き

Writing the acknowledgement part of this thesis marks the finishing line in my journey as an apprentice researcher/scientist. As I am overcome with a mixture of emotions from relief, joy, and fulfilment, looking back, it has been quite a long and arduous journey to get to this point, filled with unforgettable moments. As such, I would like to thank the people around me who have been nothing but supportive and encouraging to me in this endeavor.

Firstly, Miyoshi-sensei, who has been the biggest part in this journey; as the mentor, he went through the extra miles to provide help and guidance in both life and the scientific works. His energy and passion are contagious as he is working tirelessly to unravel the mechanism behind LINE-1 retrotransposition and regulation. Before arriving in Kyoto University, I never expected that my supervisor would actively help me in doing the experiments, but Miyoshi-sensei is always eager to lend a helping hand with finishing up some experiments; I really appreciate the kind gesture.

Next, I would like to thank my brother-in-arm Theventhiran, for being a close friend and with whom I've always had enjoyable, and insightful discussions, especially during our lunches when the discussed topics are the most random.

Many thanks to Ishikawa-sensei for his guidance and all the lab members, especially the L1 group, who have been good companions throughout my time in the lab, in particular, Makino-san, Sugino-san, and Nishimori-san. My deepest gratitude towards Tsuda-san, Hirata-san, Nishimoto-san, and the rest of Graduate School of Biostudies staff for all the help given.

Thank you to the Malaysian community that add colour to my life in Kyoto. Shoutout to Azly, Aizat, Saliza, Al, L, and Thefye. Special thanks to my love and wife, Azmina, for the constant encouragement and emotional support throughout the end of this journey.

Special appreciation to my parents, who have always been understanding and supportive of my decision to pursue my passion in science.

Credits to JASSO and MEXT (SGU) for providing me with the stipend to allow me focusing on the scientific works.

I will forever be indebted for all the guidance, support and help that I have received throughout this journey.

List of abbreviations

Abbreviations	Definitions
a.a.	amino acid
ACTB	actin beta
ADAR	double-stranded RNA-specific adenosine deaminase
AGS	Aicardi-Goutières syndrome
AID	activation-induced cytidine deaminase
APOBEC	apolipoprotein B mRNA editing enzyme, catalytic polypeptide
CMV	cytomegalovirus immediate-early
CTD	C-terminal domain
DDX60	probable ATP-dependent RNA helicase DDX60
DDX60L	probable ATP-dependent RNA helicase DDX60-like
DHX58	ATP-dependent RNA helicase DHX58
DNA	deoxyribonucleic acid
dRNase	deactivated RNB ribonuclease
dsRNA	double-stranded RNA
EIF2AK2	interferon-induced, double-stranded RNA-activated protein kinase
EN	endonuclease
ERV	endogenous retrovirus
FBXO44	F-box protein 44
Fluc	firefly luciferase
GO	gene ontology
GSEA	gene set enrichment analysis
HELZ2	helicase with zinc finger 2
HERC5	HECT and RLD domain containing E3 ubiquitin-protein ligase 5
IFI16	gamma-interferon-inducible protein 16
IFIT	interferon-induced protein with tetratricopeptide repeats

IFN	interferon
IFNAR	interferon-α/β receptor
IP	immunoprecipitation
ISG	interferon-stimulated genes
LGALS3BP	galectin-3-binding protein
LINE-1 or L1	long interspersed element-1
LTR	long terminal repeat
MS	mass spectrometry
Mya	million years ago
OAS	2’-5’-oligoadenylate synthetase
OASL	2’-5’-oligoadenylate synthetase-like protein
ORF	open reading frame
PARP	poly(ADP-ribose) polymerase
PEG	paternally expressed gene
ppt	polypurine tract
PRR	pattern recognition receptor
qPCR	quantitative polymerase chain reaction
RAG	recombination activating gene
RBM	RNA-binding mutant
RC	retrotransposition-competent
REP	retrotransposition indicator cassette
RNA	ribonucleic acid
RNB	exoribonuclease RNase II/R
RNP	ribonucleoprotein particle
RRM	RNA recognition motif
RT	reverse transcriptase
SAMHD1	deoxynucleoside triphosphate triphosphohydrolase SAMHD1

SETDB1	SET domain bifurcated histone lysine methyltransferase 1
siRNA	small interfering RNA
SLE	systemic lupus erythematosus
SS	Sjögren syndrome
ssDNA	single-stranded DNA
ssRNA	single-stranded RNA
SVA	SINE-VNTR-Alu
TE	transposable element
TREX1	three-prime repair exonuclease 1
TRIM	tripartite motif-containing protein
TUT	terminal uridylyl transferase
UTR	untranslated region
WA	walker A
WB	walker B
ZAP	zinc-finger antiviral protein ZC3HAV1
γ-H2AX	gamma-H2A histone family member X

Table of contents

Figures, Source Data, and Tables list

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 1

General Introduction

1.1 Transposable elements in the human genome

Initial report of human genome sequencing revealed an intriguing fact that the coding regions are extremely small in comparison to the non-coding ones; the coding regions only cover approximately ~1.5% of the human genome, and are estimated to consist of ~20,000 genes (Lander et al., 2001; Venter et al., 2001). The rest of the genome is non-coding and initially viewed as ‘’junk DNA’’. The relatively huge fraction of the non-coding regions begs the question of how many of them are physiologically relevant. Over the years, it was shown that these non-coding regions house critical regulatory elements including promoters, enhancers, introns, and untranslated regions (UTRs) for their respective mRNAs. Besides regulating the transcription of mRNAs, some of the non-coding DNAs are transcribed into standalone non-coding RNAs. Well-studied examples are ribosomal and transfer RNAs; however, less characterized non-coding RNAs are being discovered such as micro RNAs and long non-coding RNAs (Lander, 2011; Kaikkonen and Adelman, 2018). Recently, the complete human genome sequence has been released (Nurk et al., 2022). The percentage of DNA fraction shown in this thesis is based on the initial report of human genome sequencing.

Amongst the non-coding regions, repetitive transposable elements (TEs) was found to cover almost half of the human genome (Lander et al., 2001), confirming previous estimate based on the seminal DNA melting and reassociation experiments performed over 50 years ago (Waring and Britten, 1966; Britten and Kohne, 1968). Although this finding regarding DNA mobility and TEs abundance are currently well-known and accepted, it was groundbreaking when Barbara McClintock first proposed this idea in 1950 (McClintock, 1950). Her work shifted the paradigm that the genome is static, to the one that is more flexible and changing (Kazazian, 2011). In addition, she proposed that TEs play regulatory roles in gene expression. Her early foresight is proven true with more evidence found on TEs contribution in regulating gene expression e.g., DNA methylation and histone modifications on TE sequences can rewire gene transcriptional network, reviewed in (Friedli and Trono,

2015). Besides regulatory roles, TEs have also been proposed to be one of the main drivers in evolution. One example is gene domestication of an endogenous retrovirus *syncytin-1*, where the gene which originally was used by viruses to form the envelope is repurposed for placental formation in mammals (Mi et al., 2000). Besides placental formation, the adaptive immune system VD(J) recombination system in jawed vertebrates depends on the activity of another repurposed TEs: recombination activating gene (RAG)1 and RAG2 that belongs to *Transib* DNA transposon family (Agrawal et al., 1998; Kapitonov and Jurka, 2005; Zhang et al., 2019), and recently reviewed in (Liu et al., 2022).

TEs are also perhaps involved in speciation as TEs can sometimes cause reproductive isolation, reviewed in (Serrato-Capuchina and Matute, 2018). In addition, the rate of speciation in mammals was shown to be positively correlated with TE activity (Ricci et al., 2018); platypus, which is from the monotremata order (egg-laying mammals) that was thought to be the most ancient living order of mammals have the least amount of recently inserted TEs (Jurka et al., 2011; Ricci et al., 2018). In parallel with this hypothesis, intriguingly, two TEs, *paternally expressed 10* (*PEG10*) and *paternally expressed 11* (*PEG11*) that are conserved in eutherians (placental mammals) are not found in monotremes, while marsupials (mammals with pockets to carry the newborns) have only *PEG10* in their genome; *PEG10* and *PEG11* are involved in placental development and perhaps the evolution of viviparous (retaining fertilized egg in a parent's body) reproduction system (Ono et al., 2006; Sekita et al., 2008; Kaneko-Ishino and Ishino, 2010, 2012). These evidence hints at TEs contribution towards speciation. In sum, contrary to the dubbed name ''junk DNA'', the findings mentioned above clearly suggest the importance of TEs in the genome as we are starting to uncover more treasures among the "junk" (Kazazian, 2011).

1.2 Classifications of human transposable elements

TEs are generally categorized into class I retrotransposons that mobilize using RNA intermediates, colloquially known as ‘’copy-and-paste’’ mechanism, and class II DNA transposons that mobilize without an intermediate through a “cut-and-paste’’ mechanism. DNA transposons are usually flanked by inverted terminal repeats (ITR) that serve as recognition sites for the transposase enzymes that are responsible for DNA transposon activities. When DNA transposons mobilize, the sequence will be removed from the original sites by transposases to be inserted into new sites in the genome, sometimes leaving behind DNA excision marks, reviewed in (Hickman and Dyda, 2015). As most DNA transposons mobilize without copying its sequence, it is perplexing how these elements manage to greatly amplify their copy number in the genome. Nevertheless, DNA transposons somehow manage to have proliferation bursts in an anthropoid primate ancestor during ~80 million years ago (Mya) to ~37 Mya. For unknown reason(s), DNA transposons stopped mobilizing in most primates at ~37 Mya. It was suggested that the ancestor perhaps gained potent inhibitor(s) against horizontal transfer (HT) of DNA transposons as HT is thought to be the main mode of human DNA transposons amplification. During the last 40 Mya, a single lineage of primate-specific retrotransposon Long Interspersed Element-1 (LINE-1 or L1) L1PAs were amplifying their sequences to remain active in the human genome until now (Khan et al., 2006; Pace and Feschotte, 2007). As DNA transposons became inactive, retrotransposons became the main TEs in the primate ancestor. Currently, in human, DNA transposons make up ~3% of the genome (Lander et al., 2001), which mostly came from the burst amplification period (Pace and Feschotte, 2007).

Retrotransposons are classified based on the presence of flanking long terminal repeat (LTR) sequence(s), known as LTR retrotransposons and non-LTR retrotransposons, respectively. Human LTR retrotransposons, which are also known as endogenous retroviruses (ERVs) make up ~8% of the human genome and are not able to retrotranspose due to accumulated mutations such as point mutations, truncations, and internal

recombination, which sometimes result in having only a single LTR (solo LTR) (Lander et al., 2001) (**Fig. 1.1**). Despite this, some human-specific ERVs (HERVs) are still transcribed (Schmitt et al., 2013; Grow et al., 2015; Bannert et al., 2018; Tokuyama et al., 2018) and some of the HERVs still have intact ORFs (Tokuyama et al., 2018; Ueda et al., 2020). HERVs are known for their contribution in rewiring transcriptional network during embryogenesis (Grow et al., 2015). The most transcriptionally active and studied HERV is HERV-K (HML-2), which has been extensively reviewed in (Garcia-Montojo et al., 2018; Xue et al., 2020).

Non-LTR retrotransposons make up the majority of human TEs, with L1 covering ~16.9% of the human genome, followed by Alu (~10.6%), and a tiny fraction of SINE-VNTR-Alu (SVA) and processed pseudogenes (<1% each). Among the non-LTR retrotransposons, Alu has the most copy number (~1.1 million copies), which is approximately double of L1's (~516,000 copies); processed pseudogenes and SVA have ~8000 and ~2,700 copies, respectively (Lander et al., 2001). The full-length L1 is significantly longer than other non-LTR retrotransposons at approximately 6 kb (Scott et al., 1987), followed by SVA and Alu at ~2 kb and ~300 bases, respectively (**Fig. 1.1**). Alu and SVA details are extensively reviewed in (Hancks and Kazazian, 2010, 2016; Richardson et al., 2015). Among the non-LTR retrotransposons, only L1 is autonomous and is capable of retrotransposition using its own encoded proteins, while Alu and SVA rely on L1-encoded reverse transcriptase for *trans* mobilization (Esnault et al., 2000; Dewannieux et al., 2003; Hancks et al., 2011; Raiz et al., 2012).

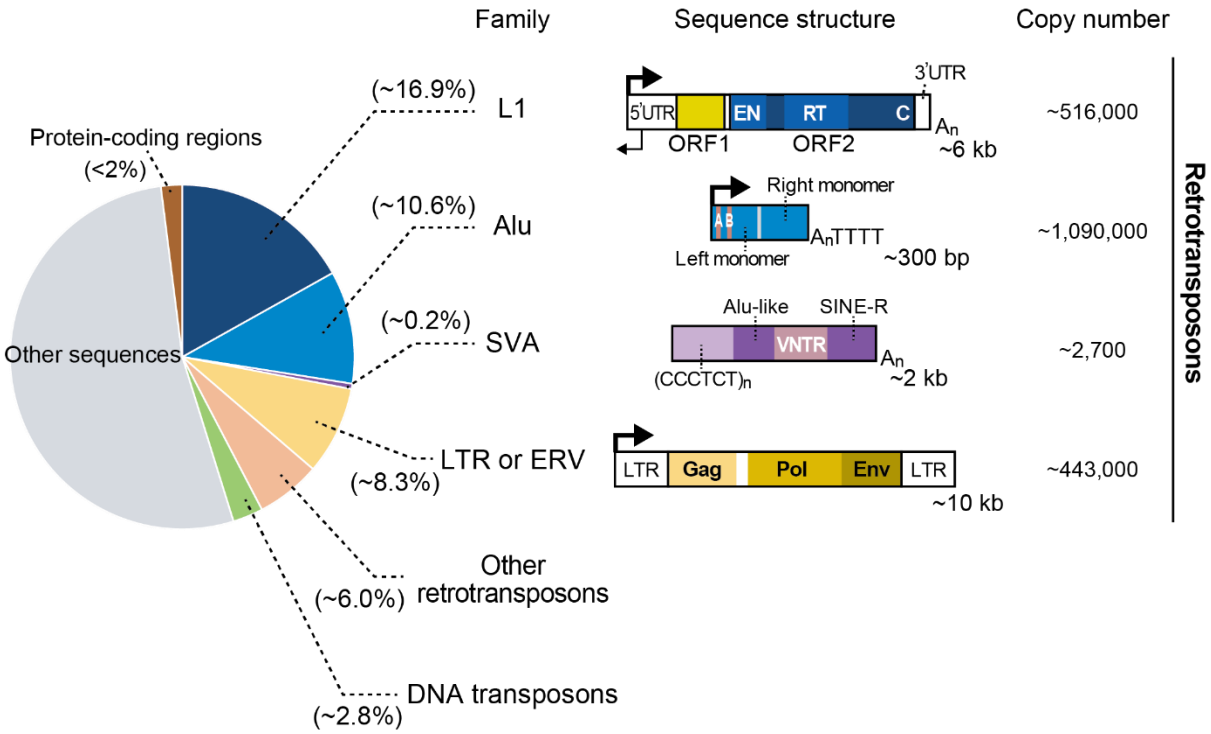


Fig 1.1: *TE classification and distribution in the human genome* (modified from (Luqman-Fatah and Miyoshi, 2022)). **Left**: The pie chart displays individual TE percentage in the human genome. Less than 2% of the human genome is covered by the protein-coding regions, in comparison to TEs that cover almost half of the human genome. Non-LTR retrotransposons make up the largest fraction amongst transposable elements by fraction i.e., L1 (~16.9%), Alu (~10.6%), and SVA (~0.2%). LTR retrotransposons/endogenous retroviruses (ERVs) cover ~8.3% of the genome, while DNA transposons cover ~2.8% and other retrotransposons cover ~6.0% of the genome. **Right**: Retrotransposons' respective sequence structures, copy numbers, and nucleotide lengths. The full-length retrotransposition competent L1 sequence is approximately 6 kb in length, with a bidirectional RNA polymerase II promoter in the 5' UTR, ORF1 (yellow box), a short intergenic spacer region followed by ORF2 (blue box) that contains endonuclease (EN), reverse transcriptase (RT), and cysteine-rich domains (C). There are approximately 516,000 L1 copies, which is nearly half of the Alu copy number (~1,090,000). The Alu sequence is ~300 bp in length and made up of two monomers (left and right monomer). The left monomer contains an RNA polymerase III promoter with well-conserved A and B boxes. SVA is the least abundant non-LTR retrotransposon with only ~2,700 copies with a length of ~2 kb. SVA is made up of a hexameric repeat $(CCCTCT)_n$, Alu-like, variable number of tandem repeat (VNTR), and SINE-R regions. An LTR retrotransposon or ERV is approximately 10 kb in length with flanking 5' and 3' LTR regions. It encodes Gag, Pol, and Env

proteins, like typical retroviruses, albeit truncated in some LTRs. The expression of LTR retrotransposon is driven by an RNA polymerase II promoter within the 5' LTR sequence.

1.3 Human L1 expression and retrotransposition activity

Despite being the only autonomous TE in human, most L1 copies are rendered inactive due to mutational processes e.g., base substitutions, 5' (or 3') truncation, or rearrangements during evolution (Grimaldi et al., 1984; Richardson et al., 2015). Among the ~516,000 L1 copies, only ~80-100 copies of full-length L1 retain the ability to retrotranspose (Sassaman et al., 1997; Brouha et al., 2003). Only a handful of highly active retrotransposition competent L1 (RC-L1) contribute to the majority of L1 insertion in human-lineage specific insertion (Skowronski et al., 1988; Sassaman et al., 1997; Myers et al., 2002; Brouha et al., 2003; Boissinot et al., 2004; Beck et al., 2010; Ewing and Kazazian, 2010; Huang et al., 2010; Philippe et al., 2016; Deininger et al., 2017). In primates, the evolution of L1 subfamilies follow a replacement cycle, in which a more retrotransposition-competent younger L1 emerged to replace an L1 that recently became inactive e.g., L1PA9 replaced L1PA10, followed by L1PA9 replacement by L1PA8, and so on; hence, primate L1 subfamilies form a ladder-like phylogeny. This is suggested to happen due to an arms race between L1 and Krüppel-associated box zinc finger proteins (KZFPs) that recognize and suppress specific L1 family expression (Jacobs et al., 2014; Wells and Feschotte, 2020). Currently, the L1 human-specific (L1Hs or L1PA1) family makes up the majority of active L1s in humans and is widely used in L1-related studies.

In most cells, full-length L1 expression is epigenetically repressed by DNA methylation (Woodcock et al., 1997; Bourc'his and Bestor, 2004; Coufal et al., 2009) and histone modifications (Montoya-Durango et al., 2009, 2016; Garcia-Perez et al., 2010; Rowe et al., 2010; Jacobs et al., 2014; Van Meter et al., 2014; Liu et al., 2018; Seczynska et al., 2022). Nevertheless, L1 expression was reported in the germline (Ostertag et al., 2002; Richardson et al., 2017), during early embryonic development (Garcia-Perez et al., 2007; van den Hurk

et al., 2007; Kano et al., 2009; Jachowicz et al., 2017; Richardson et al., 2017; Percharde et al., 2018; Feusier et al., 2019), in neural progenitor cells (Muotri et al., 2005; Coufal et al., 2009; Faulkner and Garcia-Perez, 2017; Sanchez-Luque et al., 2019), post-mitotic neurons (Macia et al., 2017), and several cancers (e.g., (Iskow et al., 2010; Lee et al., 2012; Solyom et al., 2012; Shukla et al., 2013; Helman et al., 2014; Tubio et al., 2014; Rodić et al., 2015; Scott et al., 2016; Rodriguez-Martin et al., 2020). L1 retrotransposition can alter the genome to generate inter- and intra-individual genetic variation (Gilbert et al., 2002, 2005; Symer et al., 2002; Beck et al., 2011; Richardson et al., 2015), and also cause *de novo* disease-producing insertional mutations (Hancks and Kazazian, 2012, 2016; Kazazian and Moran, 2017). Hence, L1 retrotransposition can have a range of impact on human individual health and human genetic varieties in general.

1.4 L1 retrotransposition mechanism

An ~6 kb full-length L1 contains a 5' UTR, two ORFs (ORF1 and ORF2) that encode ORF1p and ORF2p respectively (Scott et al., 1987; Dombroski et al., 1991; Moran et al., 1996), a small (~50 bp) intergenic region in between the ORFs, and a 3' UTR with a weak polyadenylation signal (Holmes et al., 1994; Moran et al., 1999; Goodier et al., 2000; Pickeral et al., 2000) (**Fig. 1.1**). L1 is transcribed by RNA polymerase II that depends on the promoter activity in the 5' UTR which exhibits both sense and anti-sense activities (Swergold, 1990; Speek, 2001; Olovnikov et al., 2007; Alexandrova et al., 2012). Multiple transcription factors were reported to promote L1 expression, with YY1 being the most characterized (Becker et al., 1993; Athanikar et al., 2004). Other transcription factors include RUNX family transcription factor 3 (Yang et al., 2003), Ets proto-oncogene family members (Yang et al., 1998), specificity protein 1 (Yang et al., 1998), and SRY-box transcription factor family members (Tchenio et al., 2000; Muotri et al., 2005).

Following L1 RNA transcription and cytoplasmic export, the L1-encoded proteins (ORF1p and ORF2p) are translated and bind to their own encoding RNAs in *cis* to form L1 ribonucleoprotein particles (RNPs); although ORF1p and ORF2p preferentially retrotranspose L1 RNA (Esnault et al., 2000; Wei et al., 2001; Kulpa and Moran, 2006), the proteins could also act in *trans* on non-autonomous retrotransposons i.e., Alu, SVA (Dewannieux et al., 2003; Hancks et al., 2011; Raiz et al., 2012), or other mRNAs as processed pseudogenes (Esnault et al., 2000). L1 RNP formation is critical but insufficient for retrotransposition (Martin, 1991; Hohjoh and Singer, 1996; Esnault et al., 2000; Wei et al., 2001; Kulpa and Moran, 2005, 2006; Doucet et al., 2010). L1 RNP can gain access into the nucleus without nuclear envelope breakdown e.g., in post-mitotic neuronal cells, suggesting an import mechanism involved (Kubo et al., 2006; Macia et al., 2017) (**Fig. 1.2**). However, disintegration of nuclear envelope during cell division was shown to increase L1 retrotransposition efficiency as cell division was shown to promote retrotransposition (Shi et al., 2007; Xie et al., 2013), and L1 retrotransposition frequency is highest during the S phase (Mita et al., 2018).

Endogenous ORF1p is readily detectable by western blotting and immunohistochemistry (Hohjoh and Singer, 1996; Garcia-Perez et al., 2007; Soper et al., 2008; Rodić et al., 2014; Doucet-O'Hare et al., 2015; Jachowicz et al., 2017; Payer and Burns, 2019), and is significantly more abundant than ORF2p, likely due to the extremely low translation efficiency of ORF2p; this happens as L1 RNA is bicistronic with no functional internal ribosome entry site (IRES) in the inter-ORF region, and ORF2p is translated through an unconventional reinitiation mechanism following ORF1p translation termination (Alisch et al., 2006; Doucet et al., 2010). ORF2p role in L1 retrotransposition is obvious as the ~150 kDa protein possesses endonuclease (EN) and reverse transcriptase (RT) activities (Mathias et al., 1991; Feng et al., 1996; Moran et al., 1996; Doucet et al., 2010). The EN activity introduces a single-stranded (ss) nick on the genome, which was observed *in vitro* (Feng et al., 1996; Cost and Boeke, 1998; Cost et al., 2002), followed by RT activity to generate L1

cDNA (Mathias et al., 1991; Dombroski et al., 1994). In contrast, although ORF1p is indispensable for L1 retrotransposition, the role ORF1p plays during retrotransposition remains unclear. However, it is known that the ~40 kDa protein has RNA-binding and nucleic acid chaperone activities (Hohjoh and Singer, 1996; Kolosha and Martin, 1997, 2003; Martin and Bushman, 2001; Martin et al., 2003, 2005; Kulpa and Moran, 2005; Khazina et al., 2011). It was suggested that the nucleic acid chaperone activity helps in strand exchange between L1 RNA and target site genomic DNA during the retrotransposition (Martin and Bushman, 2001). Interestingly, ORF1p is dispensable for Alu retrotransposition, and ORF2p alone is sufficient for Alu retrotransposition (Dewannieux et al., 2003).

The genomic integration of L1 cDNA is known as target-site primed reverse transcription (TPRT) (Luan et al., 1993; Feng et al., 1996; Cost et al., 2002). Once L1 RNP is in the nucleus, L1 ORF2p creates a ssDNA nick with a 3'-hydroxyl and 5'-phosphate group through cleaving at a 5'-TTTT/AA-3' sequence (the slash indicates the typical scission site by ORF2p EN) (Feng et al., 1996; Cost and Boeke, 1998; Cost et al., 2002), although the sequence might have some variations (Feng et al., 1996; Moran et al., 1996; Cost and Boeke, 1998; Cost et al., 2002; Flasch et al., 2019; Sultana et al., 2019). The liberated T-rich DNA anneals to the L1 RNA poly(A) tail and primes ORF2p reverse transcription of L1 cDNA (Kulpa and Moran, 2006; Monot et al., 2013; Doucet et al., 2015) (**Fig. 1.2**). This step is assisted by a few known host factors, i.e., PCNA recruits ORF2p to potential initiation sites or increasing ORF2p RT processivity (Taylor et al., 2013), and poly(ADP-ribose) polymerase 2 (PARP2) recognizes the nick generated by ORF2p EN to be activated and enhances poly(ADP-ribose) synthesis (Miyoshi et al., 2019). Following this, a ssDNA binding protein replication protein A (RPA) binds to L1 integration sites via poly(ADP-ribose) generated by PARP2 and may protect ssDNA intermediates generated during TPRT from cytidine deamination by the potent L1 inhibitor, APOBEC3A (A3A) or perhaps unscheduled nucleolytic attack by cellular nucleases. The position of the second nick on the opposite strand will determine whether target site duplication/deletion will be generated i.e., a downstream nick will cause target site

duplication and an upstream nick will cause target site deletion; however, mechanism of the second nick generation is still unknown (**Fig. 1.2**). The cDNA generated from reverse transcription will form DNA/RNA hybrid that depends on RNase H activity to generate a second complementary DNA strand; how this process occurs during TPRT requires elucidation as ORF2p does not exhibit any RNase H activity, but RNase H2 was suggested to compensate for this activity (Benitez-Guijarro et al., 2018).

1.5 L1 retrotransposition inhibitors

As aberrant L1 expression can cause *de novo* insertional mutagenesis, genomic alteration, and induce an innate immune response, the host has developed a range of defense mechanisms against L1 expression. The interaction between L1 and the host is often described as an arms race whereas the host gain inhibitors against L1 and at the same time L1 mutates to evade suppression. I will mainly focus on L1 post-translational inhibitors here as immunoprecipitation-coupled mass spectrometry (IP-MS) of L1 RNP performed in this study discovered a group of antiviral proteins that inhibit L1 retrotransposition. Other L1 inhibitors have been extensively reviewed in (Ariumi, 2016; Goodier, 2016; Luqman-Fatah and Miyoshi, 2022).

Intriguingly, most of the post-translational inhibitors reported to inhibit L1 retrotransposition are interferon-stimulated genes (ISGs). ISGs are a group of proteins induced upon immunogen (e.g., DNA/RNA) sensing due to type I interferon (IFN) upregulation. Many ISGs are antiviral proteins and L1 interaction with ISGs is reminiscent of a host-viral interaction. Type I IFN was shown to antagonize L1 retrotransposition (Goodier et al., 2015; Yu et al., 2015), in which ISG proteins are probably the mediators of L1 retrotransposition inhibition. The list of ISG proteins observed to inhibit L1 is as follows (**Fig. 1.2**):

(1) Three-prime repair exonuclease 1 (TREX1), is an exonuclease domain-containing protein that recognizes cytoplasmic ssDNAs for degradation, which include retrotransposons such as L1s. TREX1 mutations were found in patients with a congenital autoimmune disease, Aicardi-Goutières syndrome (AGS) (Stetson et al., 2008). An upregulation of type I IFN response that was partly mediated by L1 ssDNA was observed in a specific TREX1 knockout in neurons (Thomas et al., 2017); TREX1 is also able to reduce the ORF1p levels independent of its exonuclease activity (Li et al., 2017).

(2) Another AGS-linked protein is deoxynucleoside triphosphate triphosphohydrolase SAMHD1 (SAMHD1), which has triphosphohydrolase activity that depletes dNTP and inhibits retrovirus infection (Goldstone et al., 2011). However, this triphophohydrolase activity is not involved in L1 retrotransposition inhibition (Zhao et al., 2013); rather, SAMHD1 may inhibit L1 retrotransposition through promoting stress granule formation (Hu et al., 2015) and/or through blocking the ORF2p RT activity (Zhao et al., 2013). Contrary with these reports, Herrmann et al. suggested that SAMHD1 affects neither stress granule formation nor ORF2p RT and the interaction between L1 RNP and SAMHD1 is necessary for the inhibition through local dNTP depletion (Herrmann et al., 2018). Recently, SAMHD1 was reported to generally reduce cellular ssRNA amount to cause an increase in the number of phase-separated condensates in both nucleus and cytoplasm, i.e., stress granules, nucleoli, PML bodies, and nuclear speckles to protect immunogenic RNAs such as dsRNA from being detected by the RNA sensors, thus reducing the innate immune response (Maharana et al., 2022); a relatively high amount of ssRNA was shown to suppress condensate (e.g., prion) formation (Maharana et al., 2018). Both Maharana *et al*., 2022 and Hu et al, 2015 suggest that SAMHD1 promotes stress granule formation that may cause L1 retrotransposition inhibition; however, Maharana *et al.*, 2022 showed that SAMHD1 inhibits the innate immune response that can antagonize L1 retrotransposition and the exact SAMHD1 mechanism of action in L1 retrotransposition inhibition remains controversial.

(3) Another controversial L1 inhibitors are RNase H2A and RNase H2B (also AGS associated proteins), as opposite results were reported: on the one hand, the RNases may promote L1 retrotransposition through degradation of L1 RNA from the DNA/RNA hybrid form after reverse transcription to release the first strand L1 cDNA for the second strand cDNA synthesis (Bartsch et al., 2017; Benitez-Guijarro et al., 2018); on the other hand, the same RNases may suppress L1s by interacting with the potent L1 inhibitor, MOV10 (Choi et al., 2018).

(4) Another AGS-linked protein, double-stranded RNA-specific adenosine deaminase 1 (ADAR1) that catalyzes the deamination of adenosine to produce inosine in dsRNA (George et al., 2014) was found to inhibit L1 retrotransposition (Orecchini et al., 2017); however, the mechanism is unclear as the catalytic activity is dispensable for the L1 inhibition and L1 RNA and ORF1p amounts are not affected by ADAR1 expression (Orecchini et al., 2017). ADAR2 was also reported to inhibit L1 retrotransposition independent of its deaminase activity (Frassinelli et al., 2021).

(5) Two papers reported that the zinc-finger antiviral protein ZC3HAV1 (ZAP) inhibits L1 retrotransposition in 2015 (Goodier et al., 2015; Moldovan and Moran, 2015). It was suggested that ZAP directly regulates the L1 RNP stability as the protein co-localizes with and destabilizes L1 RNA and ORF1p in cytoplasmic foci (Moldovan and Moran, 2015).

(6) An antiviral dsRNA degrading protein pair, the 2',5-oligoadenylate synthetase (OAS)-RNase L starts the surveillance with OAS detecting dsRNA that promotes RNase L dimerization, which in turn cleaves single-stranded regions of viral RNA (Dong and Silverman, 1995). The same mechanism was observed in L1 RNA degradation (Zhang et al., 2014), suggesting that L1s may form dsRNAs that trigger this antiviral system.

(7) One of the most potent L1 retrotransposition inhibitor is MOV10 RNA helicase (Arjan-Odedra et al., 2012; Goodier et al., 2012), which has two modes of L1 inhibition, through ORF2p RT inhibition and L1 RNA degradation. In conjunction with MOV10, Terminal Uridylyl

Transferase 7 (TUT7)-dependent uridylation of L1 RNA inhibits the ORF2p RT activity, whilst the destabilization of L1 RNA is mediated by TUT4-dependent uridylation and perhaps their co-localization in the cytoplasmic foci (Warkocki et al., 2018).

(8) A group of proteins known as the activation-induced cytidine deaminase (AID)/APOBEC are cytidine deaminases that function in C-to-U deamination on nascent ssDNA produced during viral infections (Harris et al., 2003; Vieira and Soares, 2013); this group of proteins are known to inhibit both exogenous retroviruses as well as retrotransposons (Turelli et al., 2004; Bogerd et al., 2006a, 2006b; Chen et al., 2006; Chiu et al., 2006; Muckenfuss et al., 2006; Stenglein and Harris, 2006; Hulme et al., 2007; Kinomoto et al., 2007; Niewiadomska et al., 2007; MacDuff et al., 2009; Ikeda et al., 2011; Wissing et al., 2011; Metzner et al., 2012; Lindič et al., 2013; Richardson et al., 2014). AID inhibits L1 retrotransposition through translational block by directly binding to L1 RNA (MacDuff et al., 2009; Metzner et al., 2012). APOBEC3A and APOBEC3B strongly reduce both L1 and Alu retrotransposition efficiency (Bogerd et al., 2006b; Chen et al., 2006; Muckenfuss et al., 2006; Stenglein and Harris, 2006; Niewiadomska et al., 2007; Wissing et al., 2011; Richardson et al., 2014), while APOBEC3C and APOBEC3F also inhibit L1 retrotransposition, albeit less potently than APOBEC3A and APOBEC3B (Bogerd et al., 2006b; Muckenfuss et al., 2006; Niewiadomska et al., 2007); however, APOBEC3G and APOBEC3H did not show any reduction in the L1 retrotransposition efficiency, but significantly inhibited Alu retrotransposition instead (Turelli et al., 2004; Chiu et al., 2006; Hulme et al., 2007). It is important to note that APOBEC3B, APOBEC3C, and APOBEC3F inhibit L1 through a deamination-independent mechanism (Stenglein and Harris, 2006; Horn et al., 2014); this is similar in AID inhibition of L1 retrotransposition (MacDuff et al., 2009) as well as APOBEC1 (Ikeda et al., 2011). Only APOBEC3A inhibits L1 retrotransposition by deamination where the deamination of C to U on the L1s was observed using RNase H *in vitro* or a uracil DNA glycosylase-deficient cell line *in vivo*, suggesting that APOBEC3A acts on ss L1 cDNAs where the resultant deaminated strand may be immediately cleaved by a

cellular nuclease during TPRT (Richardson et al., 2014). Although the exact mechanism of APOBEC3B, APOBEC3C and APOBEC3F remains to be elucidated, co-localization of APOBEC3s with retroviruses and retrotransposons in stress granules or processing bodies was suggested to also sequester L1 RNPs and/or mediate L1 RNP degradation (Wichroski et al., 2006; Gallois-Montbrun et al., 2007; Horn et al., 2014; Goodier, 2016). Moreover, APOBEC3C may inhibit ORF2p reverse transcription (Horn et al., 2014). Thus, the reported ISG proteins that inhibit L1 suggest a viral-host-like interaction between L1 and the host cells.

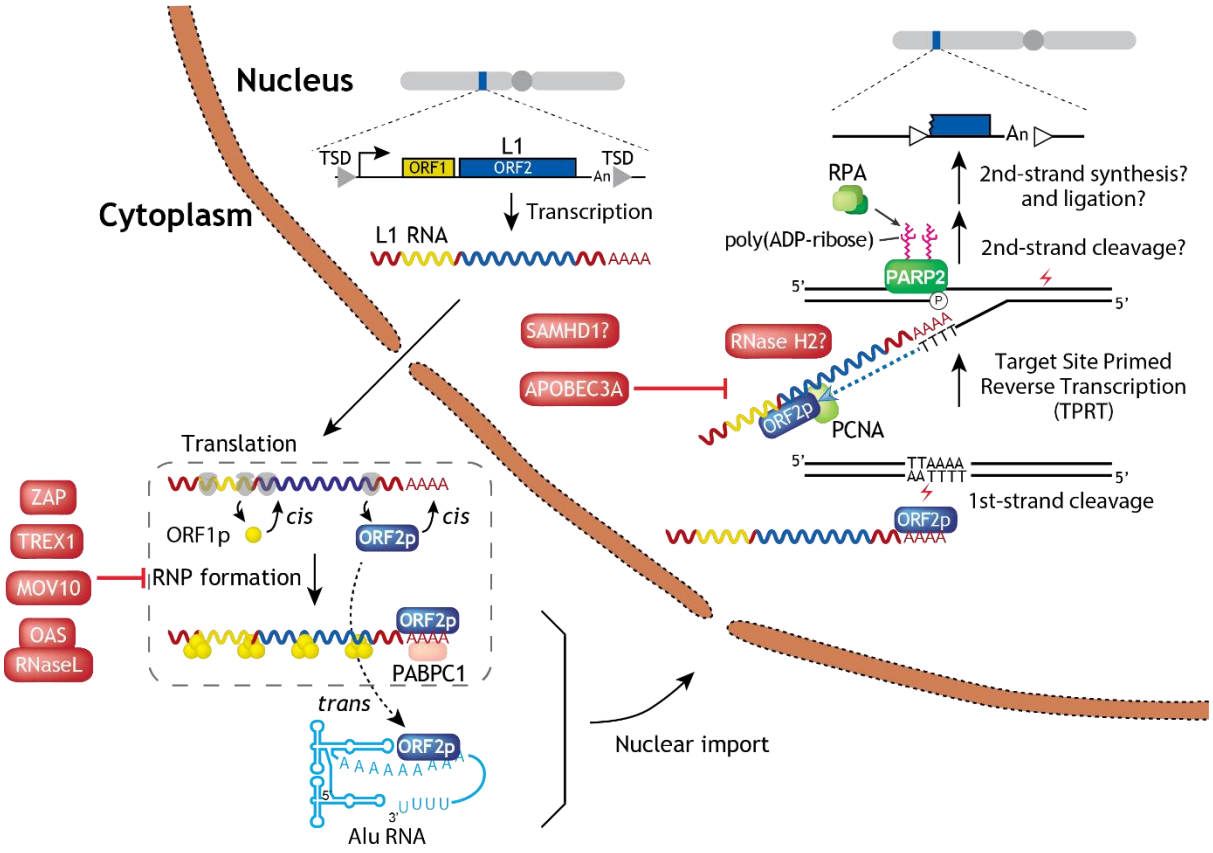


Fig 1.2: *L1 retrotransposition cycle and the regulators* (modified from (Luqman-Fatah and Miyoshi, 2022)). L1 RNA is exported into the cytoplasm after transcription where both ORF1p and ORF2p are translated. Both ORF1p and ORF2p preferentially associate with their encoding RNA in *cis* to form a ribonucleoprotein (RNP) complex that is indispensable for retrotransposition. PABPC1 is a poly(A) binding protein that binds to L1 RNA to stabilize the L1 RNP formation. L1-encoded proteins ORF1p and/or ORF2p can also retrotranspose other RNAs in *trans* including non-autonomous retrotransposons such as Alu. Contrary to ORF2p, ORF1p is dispensable for Alu retrotransposition. After L1 RNPs gain access to the nucleus, ORF2p EN cleavage at a degenerate consensus sequence,

e.g., 5′-TTTT/AA-3’ (ORF2p cleavage site is indicated by the slash) to initiate TPRT; this generates a free 3′-hydroxyl group that primes ORF2p RT reverse transcription, which may be facilitated by PCNA. ORF2p EN nucleolytic digestion causes PARP2 to be recruited to the site to increase the local concentration of poly(ADP-ribose) at the L1 integration site which increases the chance for RPA binding; RPA may protect (-) single-strand L1 cDNA from APOBEC3A deamination to facilitate L1 retrotransposition. It is unclear whether cellular RNase H such as RNase H2 may play a role in L1 RNA/cDNA hybrid clearance following ORF2p reverse transcription. Second-strand genomic DNA cleavage, (+) strand L1 cDNA synthesis, and L1 cDNA ligation to genomic DNA mechanisms remain unclear. ISG proteins that inhibit L1 retrotransposition are shown as red rectangles. ZAP, TREX1, MOV10, and OAS-RNaseL inhibit L1 RNP components (L1 ORF1p and/or RNA), denoted in the gray-dotted box. RNase H2 and SAMHD1 mechanisms in L1 retrotransposition inhibition are controversial and shown with question marks.

1.6 L1 retrotransposition-associated diseases

As L1 genomic integration can cause genetic mutation, it is only natural to assume that some L1 retrotransposition might drive genetic diseases. Indeed, the first report on L1 retrotransposition-driven genetic diseases is in two patients with hemophilia A, who were found to have L1 sequence insertion into an exon of the blood clotting coagulation *Factor VIII* gene (Kazazian et al., 1988). The first cancer driver mutation by L1 retrotransposition was found in colon cancer cells where L1 was inserted into a tumor suppressor gene *APC* (Miki et al., 1992). More than 100 cases of mutagenic L1, Alu, SVA, and processed pseudogene insertions associated with human diseases have been reported since; it is now established that *de novo* L1 insertions can drive or contribute to sporadic cases of genetic diseases. L1 contribution to human genetic diseases has been extensively reviewed in (Hancks and Kazazian, 2016; Burns, 2017, 2020; Kazazian and Moran, 2017; Terry and Devine, 2019). It is evident that L1 exonic insertions will cause insertion–deletion (indel) and frameshift mutations, whilst insertions in introns and untranslated regions could affect splicing and RNA stability (e.g., intron inclusion and exon skipping) (Beck et al., 2011). The

Pan-Cancer Analysis of Whole Genomes project revealed that L1 insertions potentially contribute to chromosomal rearrangements, including duplications, inversions, translocations, and 3' transductions in various types of cancer cells which have been underappreciated (Rodriguez-Martin et al., 2020). Approximately half of all cancers have somatic L1 retrotransposition, suggesting that L1-mediated genomic alterations are more common in cancer cells than was previously assumed (Tubio et al., 2014; Rodriguez-Martin et al., 2020). Since high L1 expression is regarded as a hallmark of human cancer, L1 ORF1p has been proposed as a biomarker for cancer diagnosis due to the high abundance of ORF1p found in some cancer cells (Rodić et al., 2014; Ardeljan et al., 2020; Cohen et al., 2020).

An increase of γ-H2AX signals that signify DSBs was observed to occur due to L1 ORF2p EN activity (Gasior et al., 2006; Miyoshi et al., 2019); L1 retrotransposition is also found to cause replication stress (Flasch et al., 2019; Ardeljan et al., 2020; Mita et al., 2020), suggesting cytotoxic effects of L1 retrotransposition. L1-mediated DNA damage may lead to apoptosis (Belgnaoui et al., 2006), senescence (Wallace et al., 2008), and cell cycle arrest, due to replication stress overwhelming replication-coupled DNA repair factors (e.g., FA proteins). As a major "guardian" of genomic integrity, p53 plays a central role in the cellular response to L1-mediated genetic damage and rearrangements; not surprisingly, p53 depletion permits cell growth and L1 retrotransposition (Ardeljan et al., 2020). Another facet of an L1 impact on the cell that recently gains a lot of interest is that L1 could induce the innate immune response that is implicated in a range of cell processes including immunity, inflammation, aging, and cell death which is discussed in the subsequent paragraphs.

1.7 Beyond retrotransposition: L1 retrotransposition-independent implicated diseases

As complex organisms evolve, it is crucial to be able to differentiate between "self" and "non-self" compounds (Janeway, 1992) or alternatively, to detect "danger" signals that are associated with cellular damage (Matzinger, 2002) to ensure survivability. The best-studied

mechanism for these is nucleic acid sensing, where the host evolved several nucleic acid receptors that could induce the innate immune response once activated; generally called as pattern recognition receptors (PRRs), these receptors could recognize and function as DNA or RNA sensors. Once PRRs encounter “anomalous” nucleic acids, downstream signaling pathways will be activated to induce the expression of cytokines including type I IFN based on the respective pathways; innate immune response pathways have been extensively reviewed elsewhere (Motwani et al., 2019; Hopfner and Hornung, 2020; Rehwinkel and Gack, 2020; Onomoto et al., 2021). The secreted cytokines will activate the innate immune response by binding to their respective receptors (e.g., IFNAR in the case of type I IFN) that will induce the upregulation of interferon-stimulated genes (ISGs), collectively also known as IFN signatures. Although the type I IFN response is well known for their roles in host protection against pathogens, aberrant chronic and/or episodic activation of type I IFN is a common hallmark in many autoimmune diseases such as systemic lupus erythematosus (SLE), AGS, and Sjögren syndrome (SS) (Ivashkiv and Donlin, 2014; Tsokos et al., 2016; Crow et al., 2019; Ukadike and Mustelin, 2021).

Viruses are the main suspect behind chronic inflammation in autoimmune patients; however, recent evidence shifted the focus towards endogenous elements, including retrotransposons. Although most retrotransposons are repressed by the host or have accumulated mutations that impair their mobilization; however, retrotransposon intermediates (RNA and/or cDNA) including ERV, L1, and Alu still induce the type I IFN response in humans, most likely, independent of retrotransposition. Much study is needed to understand how these elements trigger the PRRs. In terms of ERV, the dsRNA is likely to be the trigger of the innate immune response, which has been extensively reviewed in (Grandi and Tramontano, 2018). Both of the L1 intermediates, i.e., RNA (including this study) (Mavragani et al., 2016; Zhao et al., 2018; Tunbak et al., 2020) and the cDNA (Brégnard et al., 2016; Cecco et al., 2019; Simon et al., 2019; Zhao et al., 2021) generated by ORF2p RT trigger the type I IFN response. RT inhibitors treatment such as nevirapine or tenofovir

ameliorates the inflammatory phenotypes, reinforcing the idea that L1 cDNA contributes to exacerbating inflammation (Thomas et al., 2017; Cecco et al., 2019; Simon et al., 2019). Furthermore, the cytoplasmic DNA sensor (cyclic GMP–AMP Synthase [cGAS]) knockdown negated the immune response (Simon et al., 2019; Zhao et al., 2021); however, as it is thought that L1 cDNA is only generated in the nucleus during the TPRT reaction, how L1 cDNA accumulates in the cytoplasm remains controversial. On the other hand, Alu RNA was reported to be reverse transcribed in the cytoplasm, suggested to be the cause of age-related macular degeneration through the IFN pathway, which is independent of retrotransposition (Fukuda et al., 2021). Until now, there is no conclusive evidence on L1 cytoplasmic reverse transcription; one hypothesis postulates that aborted L1 retrotransposition cDNA is exported into the cytoplasm (Brégnard et al., 2016).

Cytoplasmic RNA sensors, i.e., the RIG-I like receptors (RLRs) MDA5 and RIG-I were shown to detect L1 RNA and induce the type I IFN response (Zhao et al., 2018; Tunbak et al., 2020). MDA5 recognizes long dsRNA, while RIG-I specifically recognizes uncapped and unprocessed 5'-phosphate RNA structure as well as short dsRNA, reviewed in (Rehwinkel and Gack, 2020). It is unclear how L1 RNAs trigger the RNA sensors; however, it is likely that L1 RNA forms secondary structures such as hairpins or dsRNA (Chiappinelli et al., 2015; Cuellar et al., 2017; Tunbak et al., 2020), and not through uncapped and unprocessed 5'-phosphate RNA structure as L1 RNA is transcribed by RNA polymerase II (Swergold, 1990; Becker et al., 1993; Athanikar et al., 2004; Dmitriev et al., 2007). Intriguingly, L1 RNA can also trigger toll-like receptors (TLRs) 7/8 (Mavragani et al., 2016), which are membrane proteins that only sense RNA molecules in extracellular region and/or endosomes. How and if L1 RNA/RNPs are transported into endosomes remains unknown, but L1 RNA/RNP may be released into the extracellular space upon cell death or via secretion of exosomes or microvesicles (Balaj et al., 2011; Kawamura et al., 2019).

Non-LTR retrotransposon Alu RNA was also shown to strongly induce the innate immune response besides L1. Ro60, a common autoantigen in SLE patients, which is an

RNA-binding protein can directly bind Alu RNA to suppress the type I IFN response (Hung et al., 2015). Inverted-repeat Alu structure is suggested to be the cause of the immune response as shown through the Alu upregulation by DNA methyltransferase inhibitors, i.e., azacytidine and decitabine (Mehdipour et al., 2020). Unlike L1, Alu RNA is one of the major RNA molecules edited by ADAR1 and depletion of ADAR1 upregulates Alu expression, leading to a stronger immune response; ADAR1 editing is suggested to cause destabilization of Alu RNA to mitigate the immune response (Athanasiadis et al., 2004; Levanon et al., 2004; Chung et al., 2018; Mehdipour et al., 2020; Nichols et al., 2021). ADAR1 is also recently shown to detect and edit Z-RNA to avoid detection by Zbp1 RNA sensor (Hubbard et al., 2022; Zhang et al., 2022), suggesting an immunogenic Z-RNA conformation of Alu (Nichols et al., 2021).

Chronic inflammation in autoimmune patients with no history of persistent viral infection can be explained by retrotransposon-mediated innate immune responses. Indeed, type I IFN is induced by retrotransposons and is readily observed in some autoimmune patients, reviewed in (Crow et al., 2019). Intriguingly, all seven genes that are linked to AGS regulate L1 (Crow et al., 2015): (1) *TREX1* (Stetson et al., 2008; Li et al., 2017; Thomas et al., 2017); (2) *SAMHD1* (Zhao et al., 2013; Hu et al., 2015; White et al., 2016; Herrmann et al., 2018); (3-5) *RNASEH2A*, *RNASEH2B*, *RNASEH2C* (Benitez-Guijarro et al., 2018; Choi et al., 2018); (6) *IFIH1* (MDA5-encoding gene) (Zhao et al., 2018; Tunbak et al., 2020); and (7) *ADAR1* (Orecchini et al., 2017). In comparison to healthy individuals, a high level of the common hallmark of IFN signature and L1 intermediates (RNA and/or DNA) are observed in SLE and SS patients in parallel (Mavragani et al., 2016, 2018). Common antigens recognized by SLE autoantibodies such as Lupus La and Ro60 are also L1 RNP-interacting proteins (also found in this study) (Goodier et al., 2013; Moldovan and Moran, 2015). In addition, L1 ORF1p autoantibody was detected in SLE patients (Carter et al., 2020; Crow, 2020). As such, RT inhibitors are potential therapeutic approaches for autoimmune diseases to suppress the L1-mediated cDNA production (Volkman and Stetson, 2014). A caveat to

this approach is that L1 RNA can also trigger the innate immune response so that the RT inhibitors may not completely eliminate the retrotransposon-mediated immune responses. Hence, a more comprehensive understanding of the L1-mediated innate immune response is vital to target L1 for immune suppression.

Although a persistent activation of innate immune responses by the retrotransposons are deleterious due to the inflammatory exacerbation, an acute upregulation may be beneficial in tumor elimination. Indeed, epigenetic therapy for cancer is shown to occur through activation of HERV, Alu, and L1 to elicit a cytotoxic effect that leads to cell death (Chiappinelli et al., 2015; Roulois et al., 2015; Jones et al., 2019). A knockdown of *SETDB1* or *FBXO44* to upregulate L1 and Alu (Griffin et al., 2021; Shen et al., 2021) or treatment of DNA methyltransferase inhibitor (Mehdipour et al., 2020) induces the innate immune response and secretion of cytokines that "reactivate" exhausted T cells and enhance immune infiltration to eliminate cancer cells (Mehdipour et al., 2020; Griffin et al., 2021; Shen et al., 2021). A combination of epigenetic therapy with knockdown of L1 and/or Alu inhibitors, e.g., SETDB1 (Jones et al., 2019) or ADAR1 (Mehdipour et al., 2020; Zhang et al., 2022) has been proposed to increase the efficacy of cancer immunotherapy. Identification of new host factors that strongly inhibit retrotransposons including some of the factors reported in this thesis will provide more candidates that can be targeted in combination with epigenetic and/or immune-based cancer therapies. However, since relying on retrotransposons for cancer treatment might occasionally do more harm than good to the patients, more studies need to be done to weigh in the benefits of exploiting retrotransposons for cancer and/or autoimmune diseases therapy in the future.

1.8 Research focus

The initial aim of my study is to uncover host factors that regulate human L1 retrotransposon by immunoprecipitation coupled-mass spectrometry of epitope tagged

ORF1p-FLAG. Firstly, ORF1p mutational analyses were performed to find stably expressed ORF1p with defective RNA-binding ability. To search for L1 RNP-specific associated proteins, differential IP-MS analyses that compare the WT and mutant ORF1p-FLAG complexes revealed a group of antiviral proteins that associate with and potentially regulate L1 retrotransposition at different steps of L1 retrotransposition cycle. I studied helicase with zinc finger domain 2 (HELZ2) protein in detail and elucidated the mechanism of L1 retrotransposition inhibition by HELZ2. L1 RNA was also shown to induce an innate immune response and HELZ2 overexpression strikingly reduced the innate immune response, possibly through reduction of L1 and other immunogenic RNA levels. Notably, I found that ORF1p binding to L1 RNA potentially shields the L1 RNA from RNA sensors to reduce the innate immune response level.

Chapter 2

L1 ORF1p mutational analyses

2.1 Introduction

Approximately 30 years after the characterization of ORF2p reverse transcriptase activity (Dombroski et al., 1991; Mathias et al., 1991), the role of ORF1p in retrotransposition remains elusive. Consensus human ORF1p is made up of 338 amino acid residues, functions as a homotrimer, and contains a disordered region, a coiled-coil domain, an RNA recognition motif (RRM), and a C-terminal domain (CTD). ORF1p structure was elucidated through X-ray crystallography (Khazina et al., 2011; Khazina and Weichenrieder, 2018) except for the disordered region from the N-terminal end to the 51 amino acid residue that is difficult to crystallize. The disordered region was suggested to mediate ORF1p multimerization and higher-order structure formation, the coiled-coil domain (52-153 residue) is vital for ORF1p trimerization, while the RRM (157-252 residue) and CTD (254-323 residue) domains form a positively-charged cleft that bind to RNAs through ionic interaction (Khazina et al., 2011; Khazina and Weichenrieder, 2018) (**Fig. 2.1**). The coiled-coil domain is made up of heptad repeats and contains an insertion of three amino acids (known as stammer) that disturbs the periodicity of the coiled-coil and separate the conserved and non-conserved heptads (Khazina and Weichenrieder, 2018). Two conserved RhxxhE trimerization motifs where h denotes hydrophobic amino acid and x denotes any amino acid were found in ORF1p coiled-coil domain (Kammerer et al., 2005; Khazina et al., 2011).

ORF1p mostly binds to single-stranded nucleic acids and although it was suggested to lack sequence specificity, oligo(dA) DNA was found to be discriminated against (Khazina et al., 2011). Ectopically expressed ORF1p preferentially bind to its encoding RNA in *cis* (Wei et al., 2001; Khazina et al., 2011; Taylor et al., 2013); however, RNA Immunoprecipitation (RIP) of endogenous ORF1p from younger L1 families (L1PA1, L1PA2, L1PA3, L1PA4) in prostate cancer cell lines suggest that ORF1p bind to many non-L1 RNAs including mitochondrial RNA, and only a small fraction of endogenous ORF1p binds to L1 RNA at 0.1% ~ 0.15% of the reads (Briggs et al., 2021). This suggests that endogenous ORF1p may

bind to non-L1 RNA for pseudogene processing (Mandal et al., 2013) or to stabilize the bound RNA (Briggs et al., 2021).

ORF1p binding to L1 RNA is sufficient to form cytoplasmic L1 RNP (Kulpa and Moran, 2005), albeit it will not be retrotransposition-competent (RC) without ORF2p. Since ORF2p expression is extremely low, the bulk majority of cytoplasmic L1 RNP do not contain ORF2p. L1 RNP also localizes in foci-like structure that is observable through immunostaining (Goodier et al., 2007; Doucet et al., 2010). Ectopically overexpressed full-length L1 in L1 retrotransposition-permissive cell lines usually display a high percentage of cells (~80%) containing the foci-like structure (Doucet et al., 2010; Pereira et al., 2018), which is overlapping with stress granule (Goodier et al., 2007), a dynamic membraneless structure that forms in response to multiple stressors e.g., heat shock and UV irradiation (Anderson and Kedersha, 2002). However, it remains unclear if this structure appears due to L1 RNP localization in stress granule. The foci structure was also reported to co-localize in P-bodies, which is another membraneless structure that constitutively exist in the cytoplasm (Goodier et al., 2007; Briggs et al., 2021). Host factors that inhibit L1 retrotransposition such as the zinc-finger antiviral protein (ZAP) or MOV10 RNA helicase were found to co-localize with L1 cytoplasmic foci (Goodier et al., 2012, 2015; Moldovan and Moran, 2015).

In this chapter, a panel of ORF1p missense mutations were generated based on previous literature to further understand the role of ORF1p domains in L1 cytoplasmic foci formation and/or retrotransposition. I identified an ORF1p RNA-binding mutant that exhibits (1) stable expression in human cell lines, (2) markedly reduced ability to form L1 cytoplasmic foci, (3) impaired RNA-binding ability, and (4) loss of L1 retrotransposition ability. The mutated sites are in the RRM domain: R206A/R210A/R211A (a.k.a., M8/RBM); this mutant was used in the IP-MS analysis for the ORF1p complexes by comparison with the wild-type L1 to identify L1 regulators that will be described in chapter 3.

2.2 Results

2.2a Generation of ORF1p missense mutations

The alanine missense mutations were generated based on a full-length human RC-L1 expression construct. This construct (pJM101/L1.3FLAG) expresses ORF1p with FLAG epitope tag at its carboxyl-terminus (**Fig. 2.1**) (Sassaman et al., 1997; Moldovan and Moran, 2015).

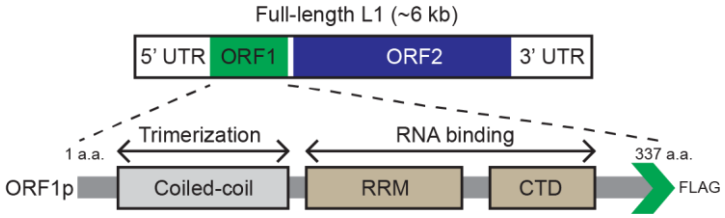


Fig. 2.1: *Full-length RC-L1 (L1.3: Genbank Accession #L19088) and ORF1p domains schematic.* Functional ORF1p domains are noted below the full-length RC-L1 including coiled-coil domain that is essential for trimerization, RNA recognition motif (RRM) and carboxyl-terminal domain (CTD) that are important for RNA-binding. The in-frame FLAG epitope tag is shown as green arrowhead. a.a. indicates the amino acid length of ORF1p.

Mutations in the ORF1p regions were made based on the putative roles suggested by respective literature as follows: (1) M1: a conserved pair of amino acids (N157A/R159A) essential for ORF1p cytoplasmic foci formation and L1 retrotransposition (Goodier et al., 2007); (2) M2: conserved amino acids that are essential for coiled-coil structure formation and predicted to play a role in ORF1p trimerization (Kammerer et al., 2005) (R117A/E122A); (3) M3 and M4: amino acids proposed to stabilize ORF1p homotrimer formation by mediating the coordination of chloride ions in the coiled-coil domain (Khazina et al., 2011) (N142A and R135A, respectively); (4) M5: an acidic patch that is proposed to act as a putative ORF1p protein-protein interaction surface that may interact with host factors (Khazina et al., 2011) (E116A/D123A); (5) M6-M9: amino acids required for ORF1p RNA-binding activity (Kulpa and Moran, 2005; Doucet et al., 2010; Khazina et al., 2011) (K137A/K140A, R235A, R206A/R210A/R211A, and R261A, respectively); and (6) M10: an

amino acid required to maintain nucleic acid chaperone activity (Doucet et al., 2010) (Y282A). In the ORF1p crystal structure (Khazina et al., 2011), the respective positions of each mutation with the putative functions are shown (**Table 1 and Fig. 2.2**).

Number	**M1**	**M2**	**M3**	**M4**	**M5**
Mutational Sites	N157A, R159A	R117A, E122A	N142A	R135A	E116A, D123A
Putative functions	Conserved site, abolished ORF1p foci	RhxxhE motif (trimerization)	Chloride ion mediator (trimerization)	Chloride ion mediator (trimerization)	Putative protein binding site, does not affect RNA-binding
References	Goodier, JL. *et al*., 2007	Kammerer, RA. *et al*., 2005	Khazina, E. *et al.*, 2011	Khazina, E. *et al.*, 2011	Khazina, E. *et al.*, 2011

Number	**M6**	**M7**	**M8**	**M9**	**M10**
Mutational Sites	K137A, K140A	R235A	R206A, R210A, R211A	R261A	Y282A
Putative functions	Decrease RNA-binding (coiled-coil)	Decrease RNA - inding (RRM)	Loss of RNA-binding ability (RRM)	Decrease RNA-binding and chaperone activity (CTD)	Decrease chaperone activity (CTD)
References	Khazina, E. *et al.*, 2011	Doucet, AJ. *et al.*, 2010	Khazina, E. *et al.*, 2011	Doucet, AJ. *et al.*, 2010	Doucet, AJ. *et al.*, 2010

Table 1: *A list of ORF1p mutants used in this study*. Ten alanine missense ORF1p-FLAG mutants (M1 to M10) were generated. Row 1, mutant number. Row 2, mutated site of the alanine mutations; commas indicate double (i.e., M1, M2, M5, M6) or triple (i.e., M8) mutants. Row 3, putative functions affected due to the alanine exchange. Row 4, previous studies references implicating the mutations effects on ORF1p.

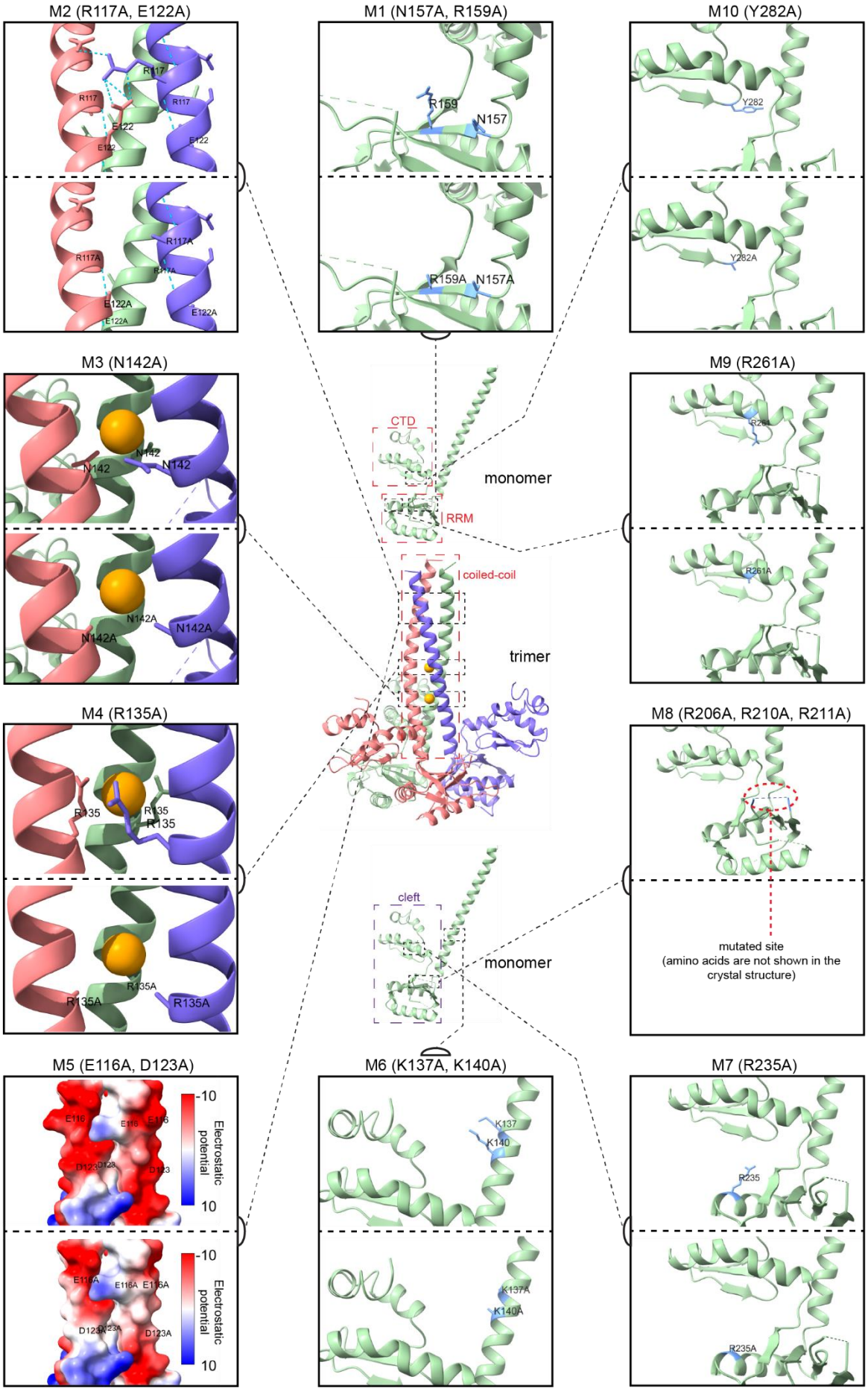

M2 (R117A, E122A)
M1 (N157A, R159A)
M10 (Y282A)
M3 (N142A)
M9 (R261A)
M4 (R135A)
M8 (R206A, R210A, R211A)
M5 (E116A, D123A)
M6 (K137A, K140A)
M7 (R235A)
CTD
RRM
monomer
coiled-coil
trimer
cleft
monomer
mutated site
(amino acids are not shown in the
crystal structure)
Electrostatic potential
-10
10

Fig. 2.2: *ORF1p mutants crystal structure*. Center: Crystal structure assembly of the ORF1p monomer (top and bottom) and trimer (middle) from amino acid residues 107 to 323 in the “lifted” conformation (Protein Data Bank ID: 2ykp); each monomer is annotated with distinct colors (red, green, and purple colors). Two orange spheres indicate the chloride ion residues in the predicted position inside the coiled-coil domain (red-dotted box, top of the trimer). Each monomer forms a flexible cleft (purple-dotted box as shown in the bottom monomer) made up of an RNA recognition motif (top monomer: RRM, bottom of the cleft, red-dotted box) and a C-terminal domain (top monomer: CTD, top of the cleft, red-dotted box) to bind RNA. The mutated amino acids relative positions are indicated in black-dotted boxes connected with black-dotted lines to the respective enlarged images of the mutated sites. Perimeter: ORF1p mutated sites used in this study. The mutants were arranged from the lowest number (i.e., M1 and M2) to highest (i.e., M10) in a counterclockwise direction beginning from the top (middle), where each of the mutants is enclosed in black boxes with the respective annotation noted at the top. The WT (upper) and the mutants (lower) are separated by black-dotted lines where the corresponding mutated amino acids are indicated within each box. M1, M6, M7, M9 and M10: mutated amino acids and side chains are indicated in blue. M2: blue-dotted lines indicate hydrogen bonds formed. Please note the hydrogen bond between R117 and E122 side chains (different monomers) that stabilize the trimer. M3 and M4: predicted side chains thought to stabilize the chloride ions. M5: the mutated site was suggested to be a potential recruitment site of host factors with a relative electrostatic potential map of the ORF1p trimers surface shown. Red indicates low positive electrostatic potential (high acidity) and blue indicates high positive electrostatic potential (high basicity). M8: red-dotted circle shows the relative position of the mutated site.

2.2b ORF1p cytoplasmic foci formation is dependent on ORF1p RNA-binding ability

The expression levels of the ORF1p mutants were investigated in human U-2 OS osteosarcoma, HeLa-JVM cervical cancer, and HEK293T embryonic kidney cell lines (**Fig. 2.3a**). Immunoblotting against ORF1p-FLAG epitope tag in western blot analyses revealed the stability of the ORF1p mutant constructs protein: a severe reduction in the steady state level of ORF1p in the M1 mutant or a change in the electrophoretic mobility of ORF1p in the M5 mutant was observed when compared to the WT ORF1p-FLAG control in each cell line (**Fig. 2.3a**). In comparison to the WT ORF1p-FLAG control, the steady state levels of ORF1p in the M9 and M10 mutants appeared to be reduced depending on cell line. In addition, the results generally agree with a previous alanine scanning mutational analysis of ORF1p (Adney et al., 2019). A cross examination using anti-ORF1p antibody showed a comparatively similar band pattern when compared to anti-FLAG antibody, suggesting that the signal reduction of ORF1p-FLAG was not due to a loss of FLAG epitope tag from ORF1p (**Fig. 2.3a**).

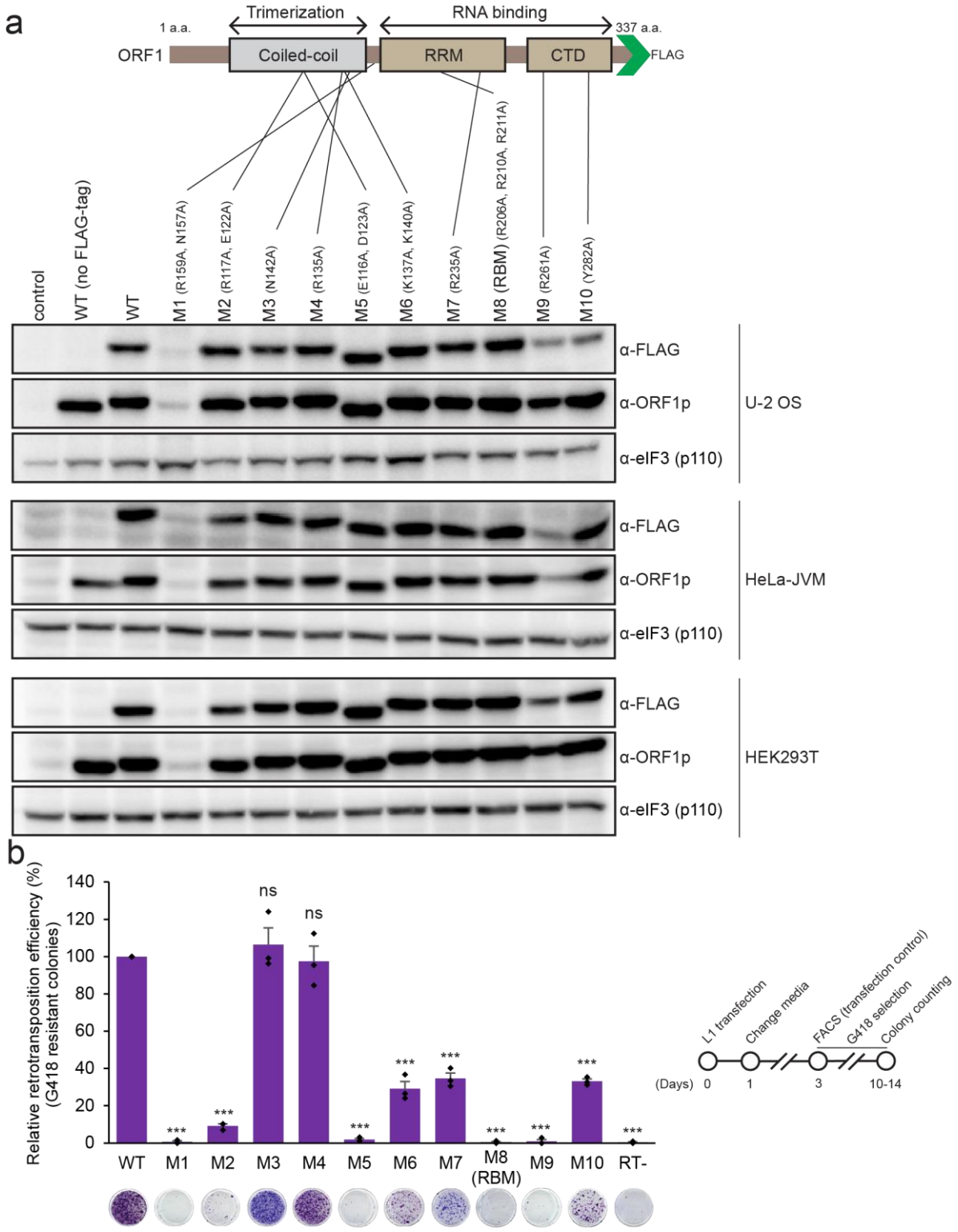


Fig. 2.3: *Steady state levels and the L1 retrotransposition efficiency in the ORF1p mutants.* (**a**) *ORF1p mutants proteins steady state levels.* Top, schematic representation of ORF1p functional domains with relative positions of the respective mutated amino acids. Bottom, western blot of ORF1p mutants' expression level in U-2 OS, HeLa-JVM, or HEK293T cells. The cells were transfected with: pJM101/L1.3FLAG (WT); pALAF001 (M1); pALAF002 (M2); pALAF003 (M3); pALAF004 (M4);

pALAF005 (M5); pALAF006 (M6); pALAF007 (M7); pALAF008 (M8); pALAF009 (M9); or pALAF010 (M10). U-2 OS, HeLa-JVM, or HEK293T cells were collected on day 5, day 9, or day 4 post-transfection, respectively, which were determined to be the optimal days to observe ORF1p steady state levels in the respective cell lines. An anti-FLAG antibody was used to detect ORF1p-FLAG. An anti-ORF1p antibody was used to cross-check that the ORF1p-FLAG signals correspond to the ORF1p levels. The eIF3 protein (p110) served as a loading control. (**b**) *L1 retrotransposition efficiencies in the ORF1p mutants.* HeLa-JVM cells were co-transfected with the plasmids used in panel (a) and a phrGFP-C plasmid to normalize for transfection efficiencies and subjected to *mneoI*-based retrotransposition assays (inset, timeline of the assay). X-axis, mutant name, and representative results from the assay; a missense mutation in the ORF2p RT domain (RT-) served as a negative control. Y-axis, the percentage of normalized G418-resistant foci compared to the WT (pJM101/L1.3FLAG) control. Pairwise comparisons relative to the WT control: p = 1.8 x 10^{-12}*** (M1); 7.6 x 10^{-12}*** (M2); 0.56^{ns} (M3); 0.67^{ns} (M4); 2.1 x 10^{-12}*** (M5); 5.7 x 10^{-10}*** (M6); 1.4 x 10^{-9}*** (M7); 2.1 x 10^{-12}*** (M8); 2.0 x 10^{-12}*** (M9); 1.3 x 10^{-9}*** (M10); 2.0 x 10^{-12}*** (RT-). Values represent the mean ± SEM of three independent biological replicates. The p-values were calculated using a one-way ANOVA followed by Bonferroni-Holm post-hoc tests: ns: not significant; *** p<0.001.

Next, the effects of ORF1p mutations on L1 retrotransposition efficiency were investigated. All ORF1p mutant constructs contain an *mneoI* retrotransposition indicator cassette within their 3'UTR. This retrotransposition indicator cassette will only be expressed upon the completion of a single round of retrotransposition to give rise to G418-resistant foci following G418 antibiotic treatment (Moran et al., 1996). The number of surviving G418-resistant foci, normalized to the transfection efficiency indicates L1 retrotransposition efficiency (Moran et al., 1996; Wei et al., 2000; Kopera et al., 2016) (**Fig. 2.4; see Methods**).

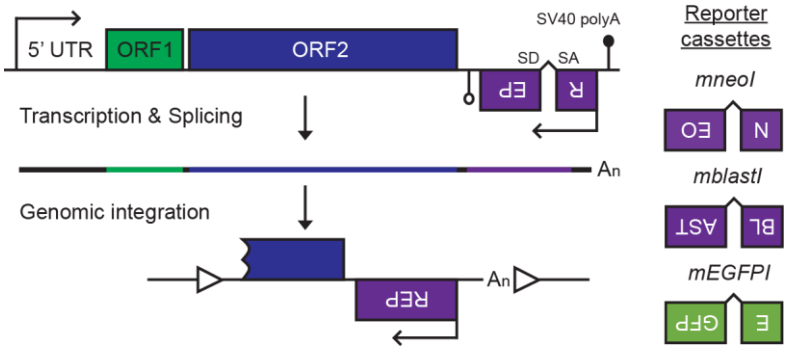


Fig. 2.4: *Retrotransposition indicator cassettes used in this study.* A retrotransposition indicator cassette (*REP*) was inserted into the 3'UTR of an L1 in the opposite orientation relative to sense strand L1 transcription. The *REP* gene contains an SV40 promoter (upside down arrow) and a polyadenylation signal (open lollipop). The *REP* gene is interrupted by an intron in the same orientation relative to sense strand L1 transcription. This arrangement ensures that *REP* will only be expressed if the sense strand L1 transcript is spliced and successfully integrated into genomic DNA by retrotransposition (bottom schematic, open triangles, target site duplications that typically are generated upon L1 retrotransposition). Three retrotransposition indicator cassettes are shown at the right of the figure: *mneoI*, which confers resistance to G418; *mblastI*, which confers resistance to blasticidin; and *mEGFPI*, which leads to enhanced green fluorescent protein (EGFP) expression.

L1 retrotransposition efficiency of the M1, M2, M5, M8, and M9 mutants are severely reduced in comparison to the positive control i.e., >90% the levels of pJM101/L1.3FLAG; the M6, M7, and M10 mutants exhibited an ~60 to 70% decrease in L1 retrotransposition efficiency, whereas the M3 and M4 mutants did not show a significant change on L1 retrotransposition efficiency, when compared to the pJM101/L1.3FLAG positive control (**Fig. 2.3b**). A reverse transcriptase deficient mutant construct (ORF2p D702A) is included as a negative control. The above data suggest that the RNA-binding, nucleic acid chaperone, putative trimerization, and ORF1p protein-binding domains are important for L1 retrotransposition (Moran *et al.*, 1996; Martin and Bushman, 2001; Wei *et al.*, 2001; Doucet *et al.*, 2010; Khazina *et al.*, 2011). M3 and M4 mutants contain single point mutations in the putative chloride-ion coordinating sites (R135A or N142A, respectively); however, the mutants did not show a reduction in L1 retrotransposition efficiency, suggesting that single

point mutations in the putative chloride-ion coordinating sites (R135A or N142A) are not sufficient to destabilize ORF1p trimerization when compared to the G132I/R135I/N142I triple mutant used in a previous study (Khazina *et al.*, 2011) or the M2 mutant.

I then focused the analyses on the ORF1p mutants that are stably expressed in HeLa-JVM cells due to the severe reduction in L1 retrotransposition efficiency with a loss of different putative functions, i.e., M2, M5, and M8 mutants. To determine whether the M2, M5, and M8 mutant ORF1p proteins localize to cytoplasmic foci and stress granule, G3BP1 was used as the stress granule marker. A U-2 OS cell line that expresses a doxycycline-inducible G3BP1 was established, where the G3BP1 is tagged at its amino terminus with an mCherry fluorescent protein (mCherry-G3BP1) (Tourrière *et al.*, 2003) (**Fig. 2.5a**); L1 cytoplasmic foci formation has been studied in U-2 OS cells previously (Doucet et al., 2010). Following transfection in the U-2 OS cells with either the WT (pJM101/L1.3FLAG), M2, M5, or M8 mutant ORF1p derivatives, ORF1p-FLAG was visualized ~48 hours post-transfection using an anti-FLAG primary antibody and Alexa Fluor 488-conjugated anti-mouse IgG secondary antibody to check for L1 cytoplasmic foci formation (**see Methods**). The cytoplasmic foci formation frequencies and intensities of the M2 and M5 mutants are relatively similar to the WT control (**Fig. 2.5b**). However, SG formation was not observed (mCherry-G3BP1 foci) without stress treatment; the cells were then treated with sodium arsenite for one hour prior to cell fixation to observe SG formation. As expected, ORF1p cytoplasmic foci was observed to co-localize with the stress granule marker mCherry-G3BP1 foci with an increase in size of the ORF1p cytoplasmic foci (**Fig. 2.5c**).

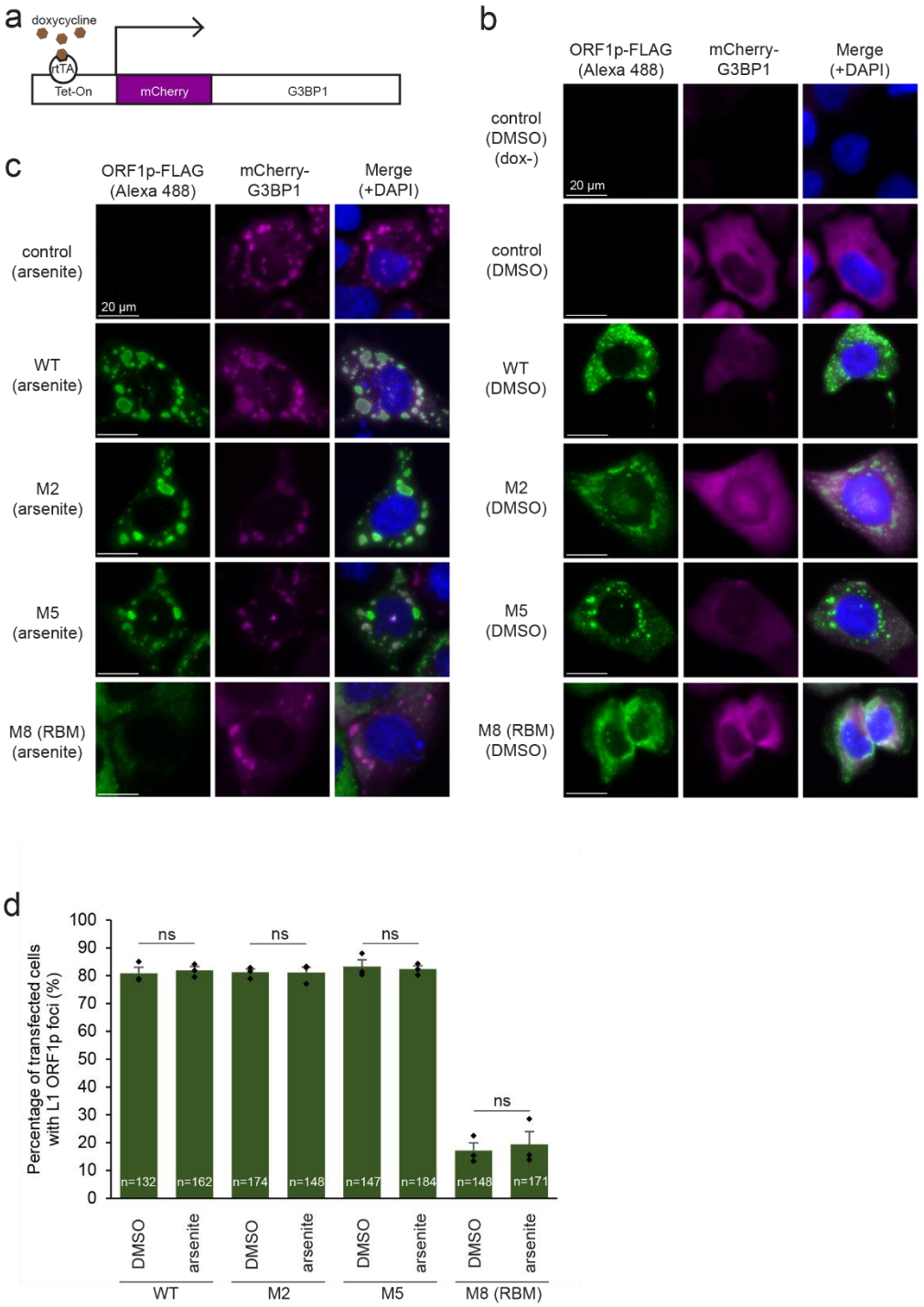


Fig. 2.5: *L1 cytoplasmic foci formation in the ORF1p mutants*. **(a)** *Doxycycline inducible mCherry-G3BP1 expression construct*. An mCherry-G3BP1 fusion protein will be expressed in U-2 OS cells upon doxycycline binding to the reverse tetracycline-controlled *trans*-activator protein (rtTA), which subsequently binds to the Tet-On promoter, activating the mCherry-G3BP1 transcription. **(b and c)** *Representative immunofluorescence images of WT, M2, M5, and M8 (RBM) ORF1p localization in the*

absence (right) or presence (left) of arsenite. The U-2 OS cells with inducible mCherry-G3BP1 expression cassette were transfected with pCEP4 (control), pJM101/L1.3FLAG (WT), pALAF002 (M2), pALAF005 (M5), or pALAF008 (M8). Two days post-transfection, the cells were treated with DMSO or 0.5 mM sodium arsenite for 1 hour prior to fixation. A mouse primary anti-FLAG antibody and secondary anti-mouse-Alexa Fluor 488 fluorescent dye-conjugated antibodies were used to visualize ORF1p. Cells not treated with doxycycline (dox-) were included as a control in the left panel. White bars, 20 µm. (**d**) *Quantification of ORF1p-FLAG cytoplasmic foci in U-2 OS cells transfected with WT, M2, M5, or M8 (RBM) ORF1p L1 expression constructs*. X-axis, construct name and whether the cells were treated with vehicle (DMSO) or arsenite. Y-axis, the percentage of transfected cells exhibiting ORF1p-FLAG cytoplasmic foci. The numbers (n) within the green rectangles indicate the number of cells analyzed in the experiment. Pairwise comparisons between DMSO and arsenite-treated cells: p = 1.00^{ns} (WT); 1.00^{ns} (M2); 1.00^{ns} (M5); 1.00^{ns} (M8 [RBM]). Values represent the mean ± SEM of three independent biological replicates. The p-values were calculated using a one-way ANOVA followed by Bonferroni-Holm post-hoc tests. ns: not significant.

A severe reduction in the percentage of cells containing ORF1p cytoplasmic foci (~15% of cells) was observed in the M8 ORF1p RNA-binding mutant when compared to U-2 OS cells expressing either the WT, M2, or M5 constructs (~80% of cells) (**Fig. 2.5d**). The reduction of M8 mutant RNA-binding ability when compared to WT ORF1p is confirmed by RNA-immunoprecipitation (RNA-IP) experiments showing markedly reduced enrichment of L1 RNA (**Fig. 2.6a)** and known ORF1p-bound RNA (HLTF and SMC2) (**Fig. 2.6b**) (Briggs et al., 2021), suggesting a general loss of ORF1p RNA-binding ability, which is consistent with the previous study (Khazina *et al.*, 2011).

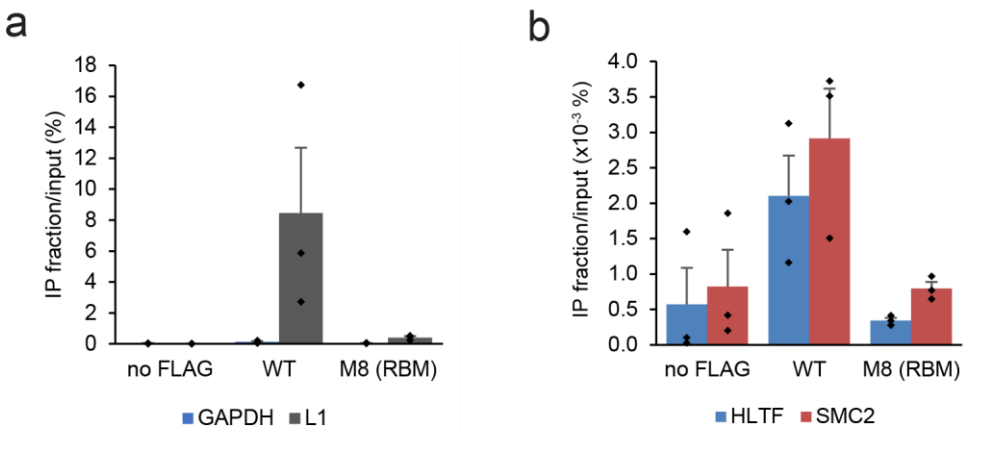


Fig. 2.6: **(a and b)** *RNA-immunoprecipitation (RNA-IP) reveals a general L1 RNA-binding defect in the ORF1p-FLAG R206A/R210A/R211A mutant (M8 [RBM]).* HeLa-JVM cells were transfected with either pJM101/L1.3 (no FLAG), WT ORF1p-FLAG (pJM101/L1.3FLAG), or the ORF1p-FLAG R206A/R210A/R211A mutant (pALAF008 [M8 (RBM)]). An anti-FLAG antibody was used to immunoprecipitate ORF1p-FLAG; for (a), reverse transcription-quantitative PCR (RT-qPCR) using a primer set (L1 [SV40]) that amplifies RNAs derived from the transfected L1 plasmid was used to quantify L1 RNA. Blue rectangles, relative levels of control GAPDH RNA (primer set: GAPDH). Gray rectangles, relative levels of L1 RNA. For (b), RT-qPCR using primer sets (HLTF and SMC2, respectively) that amplify RNAs derived from genes previously reported to be enriched in ORF1p RNA-IP experiments. Blue rectangles, relative levels of HLTF RNA. Red rectangles, relative levels SMC2 RNA. X-axes, constructs name. Y-axes, the enrichment of L1 RNA levels between the IP and input fractions. Values represent the mean ± SEM of three independent biological replicates.

Western blot experiments were also performed to confirm the reduced RNA-binding ability of the M8 ORF1p protein; HeLa-JVM cells were transfected with the pJM101/L1.3FLAG (ORF1p-FLAG) or pALAF008_L1.3FLAG_M8 (M8/RBM-FLAG) expression constructs and the resultant ORF1p complexes were immunoprecipitated using an anti-FLAG antibody. A similar level of WT and M8/RBM ORF1p-FLAG in whole cell extracts and immunoprecipitates from the HeLa-JVM cells was observed, but not in a negative control transfected with an L1 expression vector lacking the FLAG epitope tag (**Fig. 2.7a**). In addition, a ubiquitous mRNA-binding protein, the Poly(A) Binding Protein

Cytoplasmic 1 (PABPC1) was robustly detected in the IP fraction WT ORF1p-FLAG cell extracts but was severely reduced in the M8/RBM ORF1p-FLAG L1 IP fraction (**Fig. 2.7a**); the result is consistent with previous studies that found the interaction between PABPC1 and ORF1p is RNA-dependent (Dai *et al.*, 2012; Moldovan and Moran, 2015). In summary, the above data suggest that the M2, M5, and M8 mutants each produce similar steady state levels of ORF1p and reduce L1 retrotransposition efficiencies. However, ORF1p RNA-binding ability is essential for cytoplasmic foci formation as demonstrated by the M8/RBM ORF1p. Given these data, I focused subsequent studies on the WT ORF1p-FLAG and M8/RBM-FLAG proteins (herein called the RNA-Binding Mutant [RBM]) and performed label-free quantitative differential IP-MS between the WT ORF1p-FLAG and M8/RBM ORF1p-FLAG complexes in the next chapter to comprehensively identify L1 RNA/RNP-specific regulators.

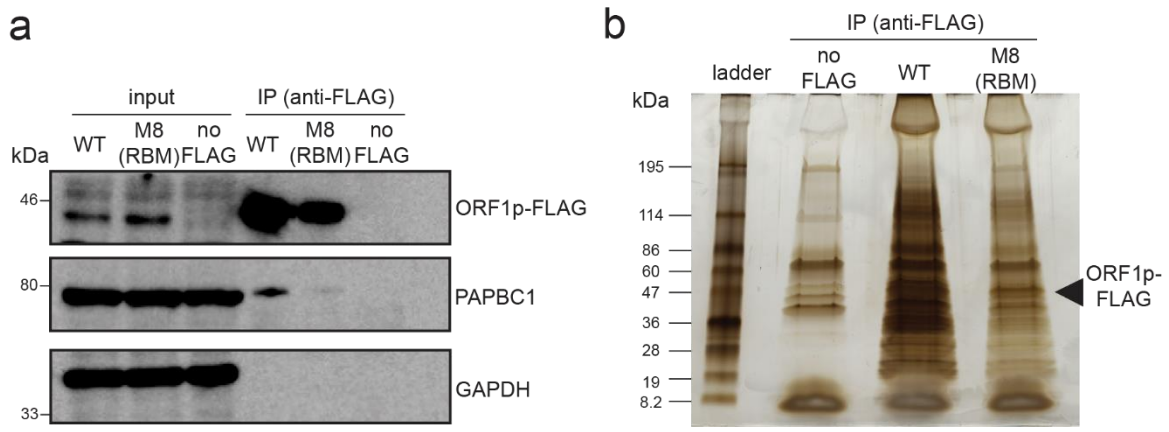


Fig. 2.7: *Immunoprecipitation of WT and M8 (RBM) mutant ORF1p*. **(a)** *The ORF1p (M8 [RBM]) mutant does not efficiently interact with Poly(A) Binding Protein Cytoplasmic 1 (PABPC1)*. HeLa-JVM cells were transfected with either pJM101/L1.3 (no FLAG), pJM101/L1.3FLAG (WT ORF1p-FLAG), or pALAF008 (ORF1p-FLAG [M8 [RBM]] mutant). An anti-FLAG antibody was used to immunoprecipitate ORF1p-FLAG. Western blots detected ORF1p (anti-FLAG), PABPC1 (anti-PABC1), and GAPDH (anti-GAPDH) in the input and IP fractions. GAPDH served as a loading control for the input fractions and a negative control in the IP experiments. **(b)** *Separation of proteins associated with the WT and mutant ORF1p-FLAG proteins.* The WT and M8 (RBM) mutant ORF1p-FLAG IP

complexes were separated by SDS-PAGE using a 4-15% gradient gel and silver staining visualized the proteins. Protein size standards (kDa) are shown at the left of the gel. Black arrowhead, the expected molecular weight of ORF1p-FLAG.

Chapter 3

A network of ISG proteins associates with L1 RNP

3.1 Introduction

An arms race between TEs and the host is proposed due to TE mutagenic effects; this includes L1 retrotransposition that can drive human genetic diseases on rare instances (Kazazian et al., 1988; Beck et al., 2011; Richardson et al., 2015; Hancks and Kazazian, 2016; Kazazian and Moran, 2017). More recently, L1 products generated in the L1 life cycle (i.e., L1 RNAs and single-stranded L1 cDNAs) are hypothesized to trigger a type I interferon (IFN) response that may contribute to inflammation and aging phenotypes (Mavragani et al., 2016; Li et al., 2017; Zhao et al., 2018; Cecco et al., 2019; Simon et al., 2019; Ardeljan et al., 2020; Tunbak et al., 2020). However, the mechanism behind L1-mediated type I IFN response and whether this process contributes directly to human immune-related diseases requires elucidation.

To prevent the potential harmful effects that L1 has on the cell, several host factors have been reported to inhibit L1 retrotransposition e.g., TREX1 (Stetson et al., 2008; Li et al., 2017; Thomas et al., 2017), ADAR1 (Orecchini et al., 2017), and ZAP (Goodier et al., 2015; Moldovan and Moran, 2015). To identify unknown host factors that specifically regulate L1 RNA/RNP, immunoprecipitation (IP) coupled with liquid chromatography-tandem mass spectrometry (LC-MS/MS), followed with differential analyses of the resultant immunoprecipitated protein in the WT ORF1p *vs.* the triple mutant ORF1p (M8/RBM) was performed. Gene Ontology (GO) (Sherman et al., 2022), Gene Set Enrichment Analysis (GSEA) (Subramanian et al., 2005), interferome database screening (Rusinova et al., 2013), and String database analyses (Szklarczyk et al., 2019) of the resultant IP-MS results revealed a network of ISGs that preferentially associates with WT ORF1p against M8/RBM ORF1p, including HELZ2, HERC5, OASL, DDX60L, and IFIT1. Biochemical and genetic analyses showed that HELZ2, HERC5, and OASL overexpression inhibits the retrotransposition of engineered L1s in cultured cells; each protein appears to act at different steps in the L1 retrotransposition cycle.

3.2 Results

3.2a Immune-related proteins associate with the WT ORF1p complex

Immunoprecipitation coupled with liquid chromatography-tandem mass spectrometry (IP/LC-MS/MS) was performed to screen for unknown cellular proteins that differentially interact with the WT ORF1p-FLAG and M8/RBM-FLAG protein complexes (**Figs. 2.7b and 3.1**), followed by label-free quantification (LFQ) analysis.

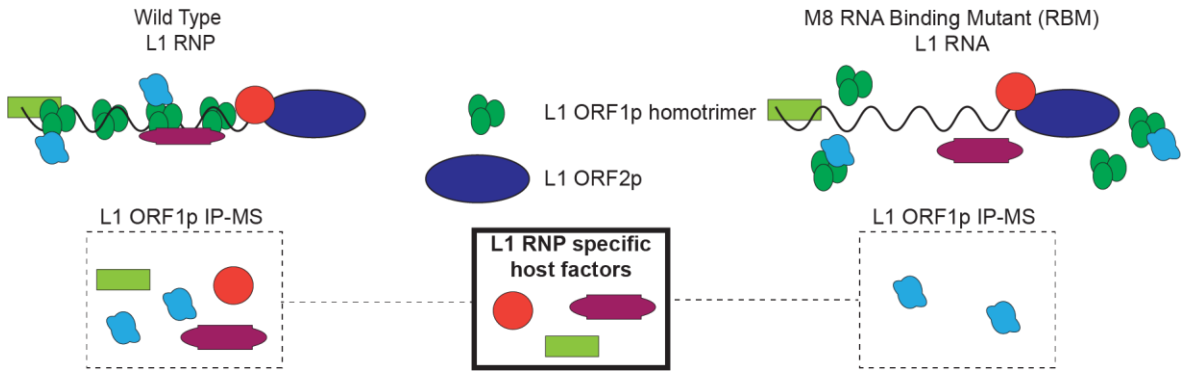


Fig. 3.1: *Rationale for a differential comparison between host factors enriched in WT ORF1p-FLAG vs. ORF1p-FLAG (M8 [RBM]) immunoprecipitation reactions.* Hypothetical diagrams of the proteins bound to WT and M8 (RBM) mutant RNP particles. Green circles, ORF1p-FLAG. Blue oval, ORF2p. Red circle, purple squared oval, and green rectangle, host factors that are differentially enriched in WT *vs*. M8 (RBM) immunoprecipitates and might associate with ORF1p-FLAG and/or L1 RNPs.

The resultant protein hits were analyzed using Database for Annotation, Visualization and Integrated Discovery (DAVID (Huang et al., 2009; Sherman et al., 2022)) gene ontology (GO) webtool that include proteins that have >0.5 $\log_2$ abundance ratio of WT *vs.* M8/RBM (**Source Data 1 and 2, see Methods**). An enrichment of viral-related GO terms was found, i.e., “host-virus interaction,” “innate immunity,” and “antiviral defense” (**Fig. 3.2a and Table 2**) associated with the WT ORF1p-FLAG protein complexes. Next, a preranked Gene Set Enrichment Analysis (GSEA) was conducted using the $\log_2$ abundance ratio of WT ORF1p-FLAG *vs*. M8/RBM-FLAG IP/LC-MS/MS protein hits to investigate if there was an enrichment of hallmark gene set signatures in the Molecular Signatures Database (MsigDB) (**see**

Methods). Two interferon-related gene sets were found—the interferon alpha and interferon gamma responses—among the top six most significantly enriched gene sets (**Fig. 3.2b and Table 3, see Methods**); both GO terms and GSEA analyses suggest a cohort of antiviral proteins preferentially associates with WT L1 RNPs.

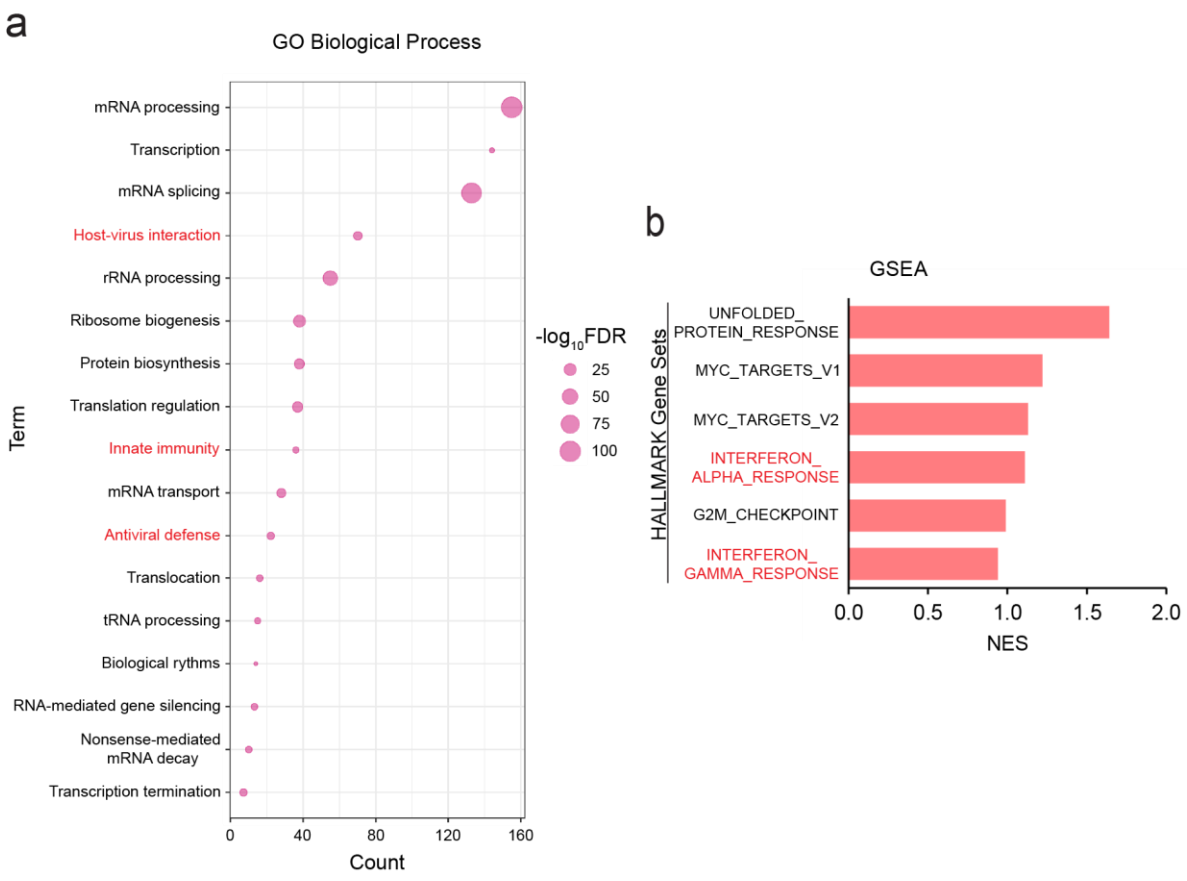


Fig. 3.2: *Differential analyses of proteins enriched in IP WT ORF1p-FLAG vs. the mutant ORF1p-FLAG complexes*. **(a)** *Gene Ontology (GO) analysis of proteins enriched in IP WT ORF1p-FLAG vs. the mutant ORF1p-FLAG complexes.* DAVID gene ontology analysis was performed using protein hits having a >0.5 $\log_2$ abundance ratio in the WT ORF1p *vs.* M8 [RBM]) complexes from the IP-MS. Listed are the "functional annotation of UniProt Keyword GO biological process" terms. X-axis, protein count, the number of proteins identified by mass spectrometry that are included in each respective GO term. Y-axis, GO term. Circle size, $-\log_{10}$FDR. Larger circles indicate higher confidence based on the FDR for each GO term. Red lettering, viral related GO terms. **(b)** *GSEApreranked analysis identifies interferon-related gene sets differentially enriched in WT ORF1p vs. M8 [RBM]*

immunoprecipitation. Gene Set Enrichment Analysis (GSEA) of log_2 abundance ratio of cellular proteins immunoprecipitated in the WT ORF1p-FLAG *vs.* (M8 [RBM])-FLAG IP complexes was performed using hallmark gene sets in the Molecular Signatures Database (MSigDB: https://www.gsea-msigdb.org/gsea/msigdb/), followed by Leading Edge Analysis to determine gene set enrichment scores. The top six hallmark gene sets with the highest normalized enrichment score (NES) are shown in descending order. X-axis, NES. Y-axis, hallmark gene sets.

Overexpression of engineered L1s previously was reported to modestly induce type I IFN response (Yu et al., 2015; Zhao et al., 2018; Ardeljan et al., 2020; Tunbak et al., 2020); to corroborate the finding, I tested whether there was an IFN-α induction in HEK293T cells transfected with full-length L1-expressing constructs. HEK293T cells was used due to the low amount of cyclic GMP–AMP synthase (cGAS, a DNA sensor), which can prevent a strong innate immune response by plasmid-based transfections and immunogenic DNAs (Tunbak et al., 2020).

HEK293T cells were transfected with either pJM101/L1.3FLAG (WT ORF1p-FLAG), pJM105/L1.3 (reverse transcriptase deficient [RT-]), or pALAF008_L1.3FLAG_M8 (M8/RBM-FLAG). IFN-α transcription in WT ORF1p-FLAG or RT-deficient mutant construct each showed a modest induction (~2.5-fold increase), while M8/RBM-FLAG expression induced a higher increase in IFN-α transcription (~4-fold increase), when compared to a mock control (**Fig. 3.3**). Using a primer set to amplify the *mneoI* retrotransposition reporter cassette to avoid amplification of endogenous L1 transcripts, the L1 RNA levels of the RT-deficient and M8/RBM-FLAG mutants were found to be similar to the WT (**Fig. 3.3**). Because the IFN-α induction between WT ORF1p-FLAG and RT-deficient mutant constructs are similar, these data suggest that L1 RNA, but not L1 RNP formation or ssDNA intermediates generated during L1 TPRT, *per se*, are primarily responsible for the modest induction of type I IFN expression in HEK293T cells (see **chapter 1.6**: "*Beyond retrotransposition: L1 retrotransposition-independent implicated diseases*" for more details regarding L1 cDNA and RNA sensing). To investigate the effect of the L1-mediated innate immune response on the

secretome, a Bio-Plex assay that allows a simultaneous assessment of 37 different cytokines and chemokines was conducted (**see Methods**). A positive control, i.e., polyinosinic:polycytidylic acid (poly[I:C])-treated cells was also included in this assay; poly(I:C) is a double-stranded RNA analog known to strongly induce the innate immune response as a comparison to the L1-mediated innate immune response (Tunbak et al., 2020). A modest but overall increase in secreted cytokines and chemokines was observed in L1-transfected cells, in particular, the M8/RBM-transfected cells when compared to a mock control. Several cytokines in the M8/RBM L1-transfected cells showed comparable upregulations to the poly(I:C)-treated cells (**Table 4**). The assay revealed that L1-transfected cells exhibited a modest increase in the secretion of several cytokines including IFN-β, IL-27, and MMP-3 when compared to respective controls; however, the levels of WT and RT-deficient L1s-transfected cells were generally lower than those in the M8/RBM L1-transfected cells, which is consistent with the IFN-α RT-qPCR result (**Fig. 3.3**) where M8/RBM L1-transfected cells induced a higher IFN-α expression compared to that of WT L1-transfected cells or RT-deficient L1-transfected cells (**Table 4**).

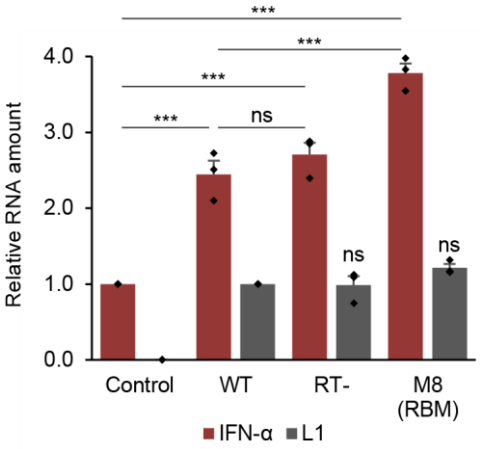


Fig. 3.3: *Engineered L1s expression modestly up-regulates IFN-α expression.* HEK293T were transfected with either pCEP4 (an empty vector control), pJM101/L1.3FLAG (WT), pJM105/L1.3 (RT-), or pALAF008 (M8 [RBM]). RT-qPCR was used to quantify IFN-α (primer set: IFN-α, which amplifies IFN-α1 and IFN-α13; this primer set act as a general indicator of IFN-α expression) and L1 expression (primer set: *mneoI* [Alu or L1]) ~96 hours post-transfection. IFN-α and L1 expression levels were normalized using β-actin (*ACTB*) as a control (primer set: Beta-actin). X-axis, constructs name. Control, pCEP4. Y-axis, relative RNA expression levels normalized to the pCEP4 empty vector control. Red bars, normalized IFN-α expression levels. Gray bars, normalized L1 expression levels. Values from three independent biological replicates ± SEM are depicted in the graph. The *p*-values were calculated using a one-way

ANOVA followed by Bonferroni-Holm post-hoc tests: pairwise comparisons of IFN-α relative to the pCEP4 control, p = 0.00028*** (WT); 0.00011*** (RT-); 3.14 x 10^{-6}*** (M8 [RBM]). Pairwise comparisons of IFN-α: WT *vs.* RT-, $p = 0.21^{ns}$; WT *vs.* M8 (RBM), p = 0.00036***. Pairwise comparisons of L1 relative to WT, $p = 0.87^{ns}$ (RT-), $p = 0.10^{ns}$ (M8 [RBM]); ns: not significant; *** p<0.001.

3.2b Proteins produced by Interferon-Stimulated Genes (ISGs) as potential L1 regulators

Several proteins expressed from interferon-stimulated genes (ISGs) have been reported to influence L1 retrotransposition e.g., MOV10, an RNA helicase (Arjan-Odedra et al., 2012; Goodier et al., 2012; Li et al., 2013); ADAR1, a double-stranded RNA-specific adenosine deaminase (Orecchini et al., 2017), TREX1, a three prime repair exonuclease 1 (Stetson et al., 2008; Li et al., 2017; Thomas et al., 2017), and ZAP, a zinc-finger antiviral protein (Goodier et al., 2015; Moldovan and Moran, 2015) (see **chapter 1.4**: "*L1 retrotransposition inhibitors*" for a comprehensive list of L1 retrotransposition-inhibiting ISGs). Due to the abundance of reported ISG proteins inhibiting L1 retrotransposition, I hypothesized that the ISG proteins associated with L1 RNPs may directly regulate L1 retrotransposition and/or L1-mediated IFN-α expression.

In line with the above hypothesis, identified proteins that have >0.5 $\log_2$ abundance ratio in WT ORF1p-FLAG *vs*. M8/RBM-FLAG in the IP/LC-MS/MS analyses was screened using the interferome database (www.interferome.org (Rusinova et al., 2013)) with a threshold of proteins that exhibit a >5-fold change in expression upon type I, II, and III IFNs induction (**Table 5**). To test for possible interactions among the ISG proteins that preferentially associated with WT ORF1p-FLAG, StringDB (https://string-db.org) was used (Szklarczyk et al., 2019).

Several interferon-inducible proteins that exhibited association as analyzed by StringDB and most (i.e., ADAR, ADARB1, APOBEC3B, DDX60, DHX58, EIF2AK2, HERC5, IFIT1, IFIT2, IFIT3, IFI16, OASL, TRIM25, TRIM56, and ZC3HAV1 [also known as ZAP]), with the exception of DDX60L, HELZ2, LGALS3BP, MOV10, and PARP12, were annotated as

antiviral defense (red) and/or innate immunity (blue) proteins in UniProt (https://www.uniprot.org/) (**Fig. 3.4a**, middle dotted box: ISG network that might regulate L1, red circles: FDR, $1.2x10^{-14}$, interaction strength, 1.51; blue circles: FDR, $5.2x10^{-12}$, interaction strength, 1.16; **see Methods**). Among the proteins mentioned, four proteins: ADAR, APOBEC3B, MOV10, and ZAP, were reported to inhibit L1 retrotransposition previously (Muckenfuss et al., 2006; Lovšin and Peterlin, 2009; Arjan-Odedra et al., 2012; Goodier et al., 2012, 2015; Li et al., 2013; Moldovan and Moran, 2015; Orecchini et al., 2017). Thus, it is possible that other ISG proteins in the identified ISG network might also be involved in the regulation of L1 retrotransposition (**Fig. 3.4a**). For further analyses, five of these unreported proteins were selected, i.e., HELZ2, IFIT1, DDX60L, OASL, and HERC5; these proteins are annotated on the volcano plot with their respective $\log_2$ abundance ratios (WT ORF1p-FLAG *vs*. M8/RBM-FLAG) and *p*-values as shown in **Fig. 3.4b**.

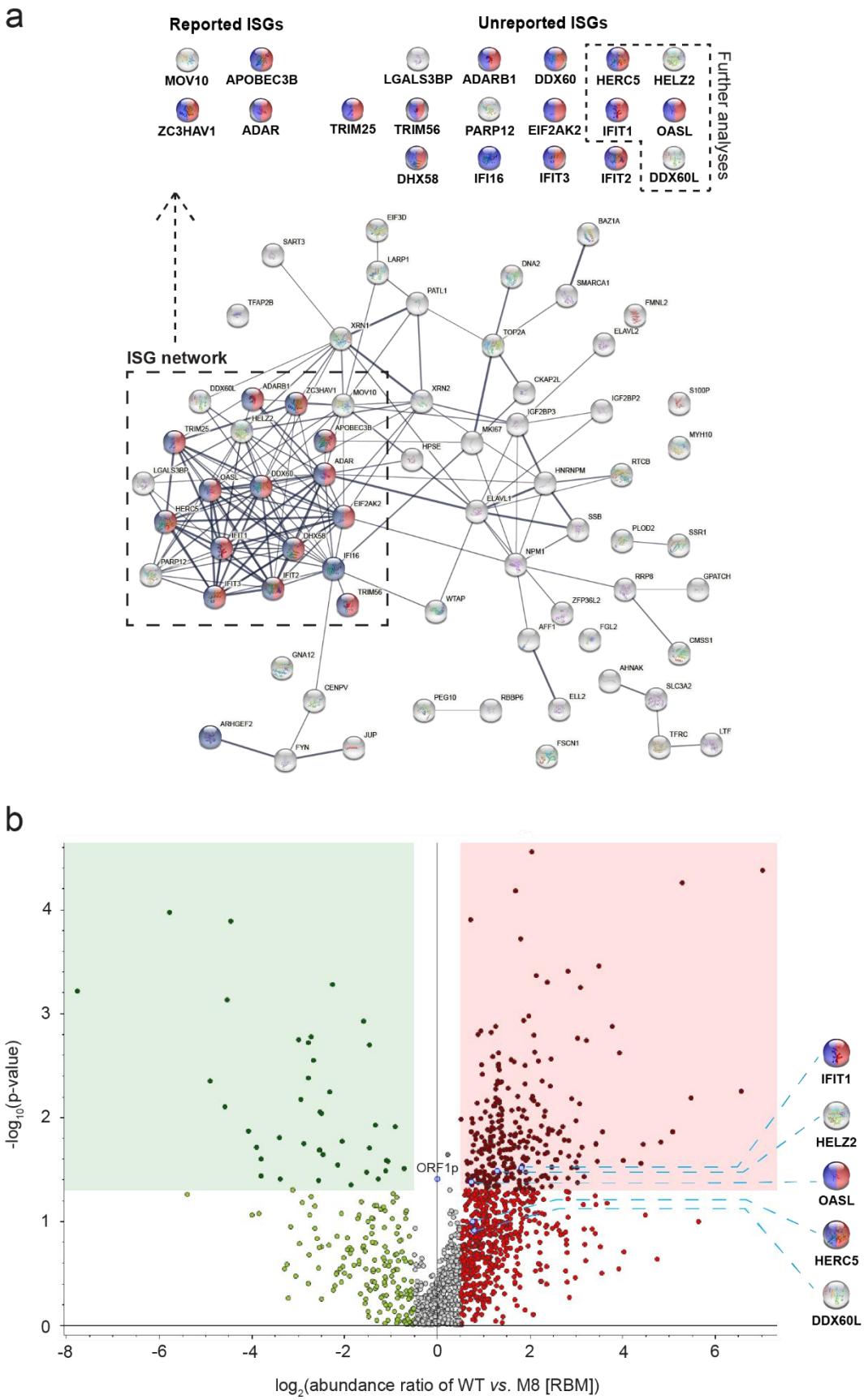

a
Reported ISGs
MOV10
APOBEC3B
ZC3HAV1
ADAR
Unreported ISGs
TRIM25
LGALS3BP
TRIM56
ADARB1
PARP12
DDX60
EIF2AK2
DHX58
IFI16
IFIT3
HERC5
IFIT1
IFIT2
HELZ2
OASL
DDX60L
Further analyses
ISG network
b
-log10(p-value)
log2(abundance ratio of WT vs. M8 [RBM])
ORF1p
IFIT1
HELZ2
OASL
HERC5
DDX60L

Fig. 3.4: *An ISG network associates with L1 RNP complexes*. **(a)** *String database analysis of ISG proteins enriched in IP WT ORF1p-FLAG vs. the mutant ORF1p-FLAG complexes*. Proteins with >0.5 log_2 abundance ratios in the WT ORF1p-FLAG *vs*. M8/RBM-FLAG complexes that exhibited a >5-fold increase in expression upon induction by type I, II, and III IFNs were subjected to String analysis. Red and blue spheres, proteins annotated in UniProt as antiviral defense and innate immunity proteins, respectively. Thickness of the inter-connecting lines indicates the strength of association based on the number of independent channels supporting the putative interactions. The black dotted box indicates a group of proteins that closely associate (i.e., a putative ISG network); the majority are annotated as antiviral defense proteins in UniProt. The proteins in the box are listed at the top of the figure (dotted arrow) based upon whether they have been reported to regulate L1 retrotransposition (left, Reported ISG), or not (right, Unreported ISG). Proteins used for further analyses are indicated in the black-dotted hexagon. **(b)** *Volcano plot of WT ORF1p-FLAG vs. M8 (RBM) ORF1p-FLAG label-free quantitative mass spectrometry analysis*. The ORF1p amounts obtained in the WT and M8 (RBM) were used to normalize protein abundance ratios as indicated in the middle of the plot (0 abundance ratio). X-axis, log_2 abundance ratios of WT ORF1p-FLAG *vs.* M8 (RBM) ORF1p-FLAG. Y-axis, $-log_{10}$ p-values of the abundance ratios. Cutoffs of >0.5 log_2 abundance ratio and <0.05 p-values are shown as references for the enrichment of proteins in the WT ORF1p-FLAG fraction (red rectangle) or M8 (RBM) ORF1p-FLAG fraction (green rectangle). Blue dotted lines, proteins enriched in WT ORF1p-FLAG complexes (i.e., HELZ2, HERC5, DDX60L, IFIT1, and OASL).

Additional co-IP experiments were performed to confirm that HELZ2, IFIT1, DDX60L, OASL, and HERC5 directly interacted with WT ORF1p-FLAG. Briefly, HEK293T cells were co-transfected with pJM101/L1.3FLAG (WT ORF1p-FLAG) and individual ISG protein expression vectors containing three copies of a MYC epitope tag at their respective carboxyl-termini (HELZ2-3xMYC, IFIT1-3xMYC, DDX60L-3xMYC, OASL-3xMYC, and HERC5-3xMYC). ORF1p-FLAG complexes from HEK293T whole cell extracts were then immunoprecipitated using an anti-FLAG antibody, and an anti-MYC antibody was used to confirm interactions between WT ORF1p-FLAG and the co-transfected ISG proteins; HERC5, OASL, IFIT1, DDX60L, and HELZ2 were found to be co-immunoprecipitated with WT

ORF1p-FLAG (**Fig. 3.5**). The results suggest that a network of antiviral ISG proteins may affect L1 RNA, L1 RNP, and/or L1 retrotransposition efficiencies.

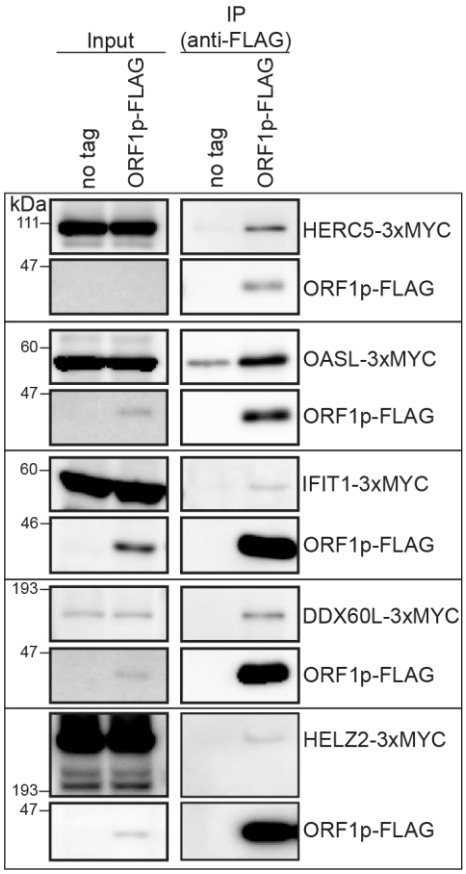


Fig. 3.5: *Co-immunoprecipitation of the ISG proteins from WT ORF1p-FLAG pull-down.* HEK293T cells were co-transfected with either pJM101/L1.3 (no tag) or pJM101/L1.3FLAG (ORF1p-FLAG) and the following individual carboxyl-terminal 3xMYC epitope-tagged ISG expression vectors: pALAF015 (HELZ2), pALAF016 (IFIT1), pALAF021 (DDX60L), pALAF022 (OASL), or pALAF023 (HERC5). The input and anti-FLAG IP reactions were analyzed by western blotting using an anti-FLAG (to detect ORF1p-FLAG) or an anti-MYC (to detect ISG proteins) antibody.

3.2c The ISG proteins, HELZ2, OASL, and HERC5 inhibit L1 retrotransposition

To investigate if ectopic overexpression of the ISG proteins identified above affect L1 retrotransposition, HeLa-JVM or HEK293T cells were co-transfected with a WT human L1 expression construct containing either a *mblastI* (pJJ101/L1.3) or *mEGFPI* (cepB-gfp-L1.3) retrotransposition indicator cassette and the carboxy-terminal 3xMYC epitope-tagged HELZ2, IFIT1, DDX60L, OASL, or HERC5 expression vectors. The resultant number of blasticidin-resistant foci or EGFP-positive cells counted indicate L1 retrotransposition efficiencies (**Figs. 2.4 and 3.6, see Methods**).

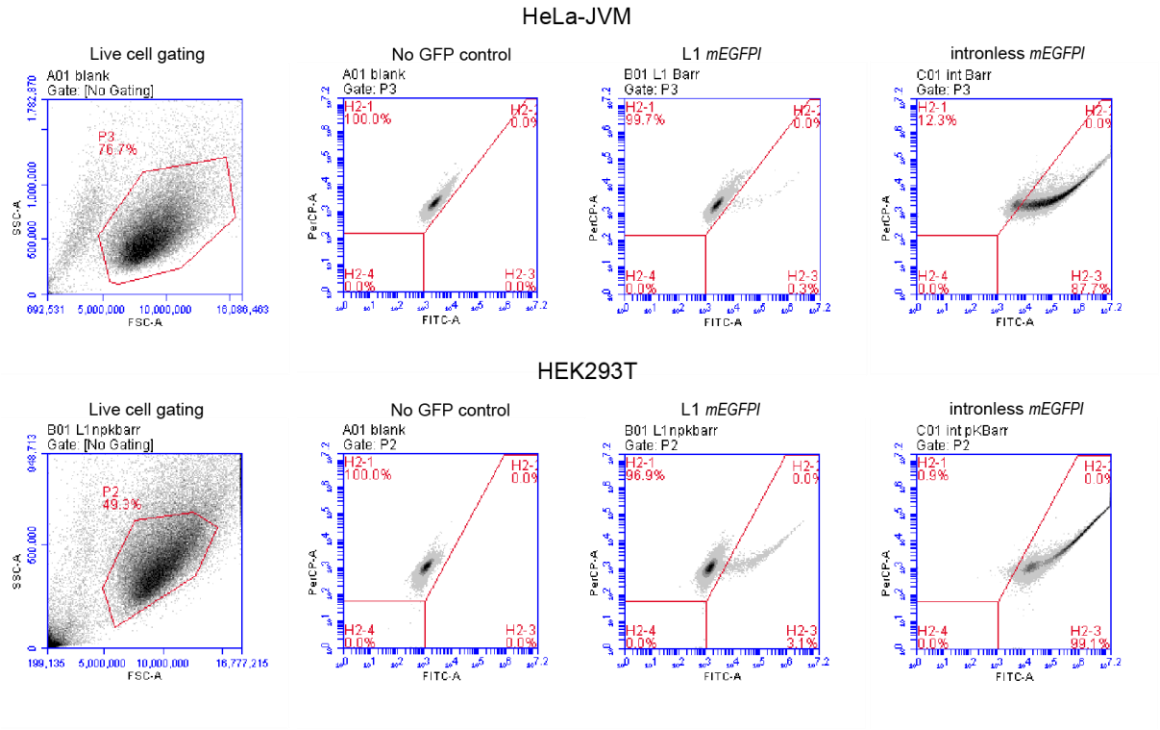
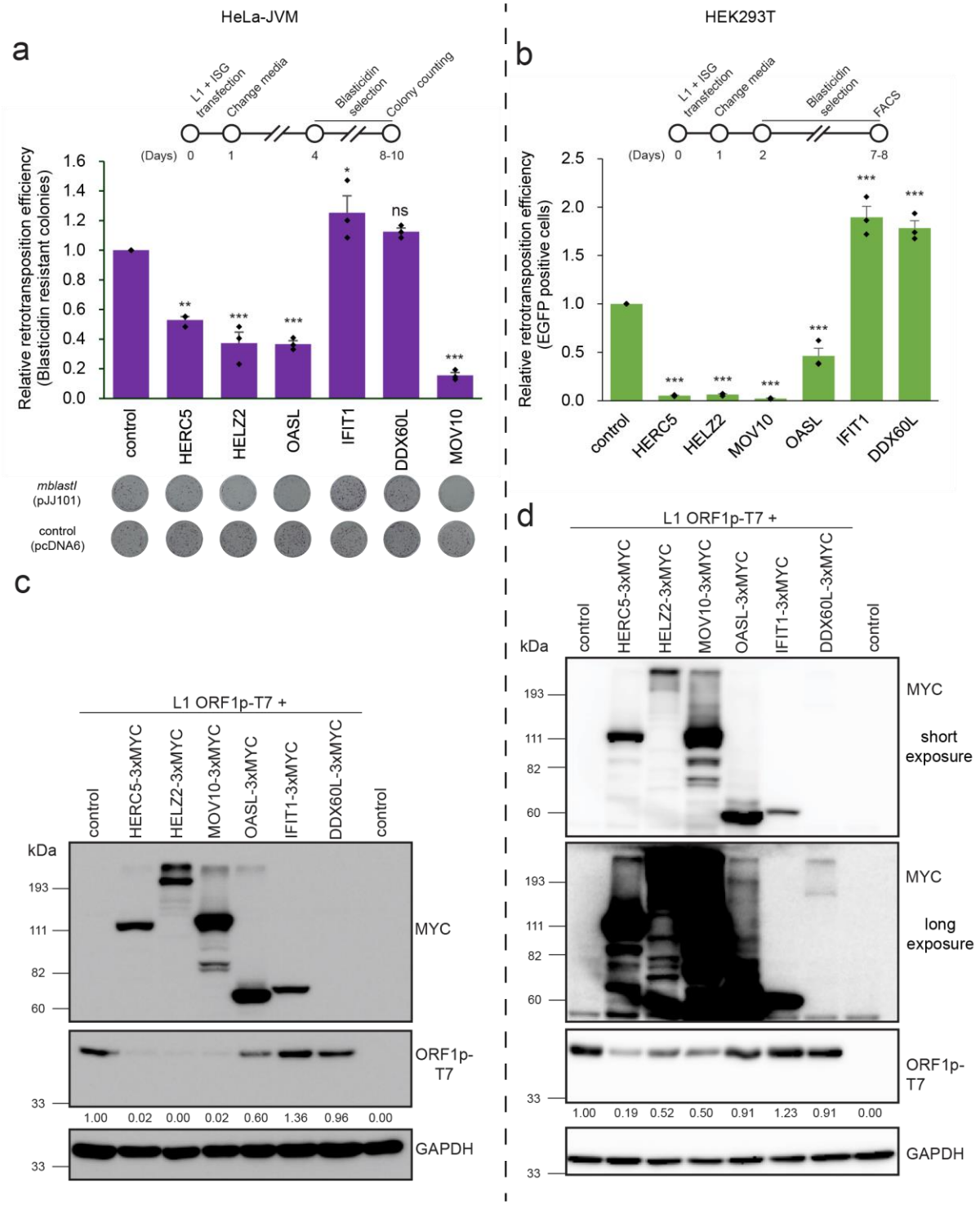


Fig. 3.6: *Representative flow cytometry plots of L1 mEGFPI retrotransposition assays in HeLa-JVM (top) and HEK293T (bottom)*. Thirty thousand live cells analyzed in both cell lines were gated in the hexagon-shaped box (i.e., P3 [HeLa-JVM] and P2 [HEK293T]). X-axis, forward scatter-area (FSC-A) channel. Y-axis, side scatter-area (SSC-A) channel (leftmost plot, live cell gating). EGFP+ cells were detected using Fluorescein isothiocyanate-area (FITC-A) channel (x-axis) with peridinin chlorophyll protein complex-area (PerCP-A) (y-axis) channel as a control. A threshold line was set accordingly based on the negative control (middle left plot, No GFP control). The H2-3 fraction indicates the percentage of EGFP+ cells. Middle right plots are representative of L1 containing and expressing the EGFP reporter cassette (*mEGFPI*) (middle right plot), while the rightmost plots show the intronless EGFP reporter cassette, which serves as a transfection control.

As a positive control, a potent L1 retrotransposition inhibitor, MOV10 expression vector, also containing a carboxyl-terminal 3xMYC epitope tag was included in the analyses. No significant decrease was observed with co-overexpression of DDX60L and IFIT1 on L1 retrotransposition in HeLa-JVM or HEK293T cells (**Figs. 3.7a and 3.7b)**; however, the expression of DDX60L was barely detected by western blot in either cell line (**Figs. 3.7c and 3.7d**). On the other hand, overexpression of HERC5, HELZ2, and OASL reduced

retrotransposition by at least 2-fold in the *mblastI*-based L1 retrotransposition assay conducted in HeLa-JVM cells (**Fig. 3.7a**) and by ~90% in the *mEGFPI*-based L1 retrotransposition assay conducted in HEK293T cells (**Fig. 3.7b**).



Legend is shown in the next page…

Fig. 3.7: *ISG proteins effects on L1 retrotransposition and ORF1p-FLAG steady state levels in HeLa-JVM (left) and HEK293T (right) cells.* **(a)** *Overexpression of HERC5, HELZ2, or OASL inhibit L1 retrotransposition in HeLa-JVM cells*. HeLa-JVM cells were co-transfected with pJJ101/L1.3, which contains the *mblastI* retrotransposition indicator cassette, and either pCMV-3Tag-8-Barr or one of the following carboxyl-terminal 3xMYC epitope-tagged ISG protein expression plasmids: pALAF015 (HELZ2), pALAF016 (IFIT1), pALAF021 (DDX60L), pALAF022 (OASL), pALAF023 (HERC5), or pALAF024 (MOV10) according to the timeline shown at the top of the respective figure. A blasticidin expression vector (pcDNA6) was co-transfected into cells with either pCMV-3Tag-8-Barr or an individual ISG protein expression plasmid (see plates labeled control [pcDNA6]) to assess cell viability. The retrotransposition efficiencies then were normalized to the respective toxicity control. X-axis, name of the control (pCMV-3Tag-8-Barr) or ISG protein expression plasmid. Y-axis, relative retrotransposition efficiency normalized to the pJJ101/L1.3 + pCMV-3Tag-8-Barr co-transfected control. Representative results of the retrotransposition (see plates labeled *mblastI* [pJJ101]) and toxicity (see plates labeled *control* [pcDNA6]) assays are shown below the graph. Pairwise comparisons relative to the pJJ101/L1.3 + pCMV-3Tag-8-Barr control: p = 8.0 x 10^{-5}** (HERC5); 4.4 x 10^{-6}*** (HELZ2); 4.9 x 10^{-6}*** (OASL); 0.011* (IFIT1); 0.12^{ns} (DDX60L); and 1.7 x 10^{-7}*** (MOV10). MOV10 served as a positive control in the assay. **(b)** *Overexpression of HERC5, HELZ2, and OASL inhibit L1 retrotransposition in HEK293T.* Top: the timeline of the assay. HEK293T cells we co-transfected with cep99-gfp-L1.3 (which has the *mEGFPI* retrotransposition indicator cassette) and either pCEP4 (control) or the following individual ISG protein expression plasmids as in panel (a). EGFP-positive cells transfected with cep99-gfp-L1.3 were counted using flow cytometry and normalized to the number of EGFP-positive cells in the transfection control (i.e., cells independently transfected with the cep99-gfp-L1.3RT(-) intronless plasmid and each of the individual ISG protein expression plasmids as described in panel (a)). X-axis, name of constructs co-transfected with cep99-gfp-L1.3. Y-axis, relative percentage of EGFP-positive cells relative to the cep99-gfp-L1.3 + pCEP4 control. Pairwise comparisons relative to the control: p = 4.8 x 10^{-7}*** (HERC5); 4.6 x 10^{-7}*** (HELZ2); 6.1 x 10^{-7}*** (MOV10); 3.9 x 10^{-5}*** (OASL); 6.2 x 10^{-7}*** (IFIT1); 1.5 x 10^{-6}*** (DDX60L). Values represent the mean ± SEM from three independent biological replicates. The p-values were calculated using a one-way ANOVA followed by Bonferroni-Holm post-hoc tests (*** p<0.001). **(c and d)** *Expression of the ISG proteins in HeLa-JVM (c) or HEK293T (d) cells.* HeLa-JVM (c) or HEK293T (d)

cells were co-transfected with pTMF3, which expresses a version of ORF1p containing a T7 gene 10 carboxyl epitope tag (ORF1p-T7), and either a pCMV-3Tag-8-Barr (control) or the individual ISG-expressing plasmids used in panel (a). Whole cell extracts were subjected to western blot analysis 48 hours post-transfection. ISG proteins were detected using an anti-MYC antibody. ORF1p was detected using an anti-T7 antibody. GAPDH served as a loading control. The relative band intensities of ORF1p-T7 are indicated under the ORF1p-T7 blot; they were calculated using ImageJ software and normalized to the respective GAPDH bands.

3.2d Some ISG proteins affect L1 RNA, ORF1p, and cytoplasmic foci formation

To further understand how ISG proteins might inhibit L1 retrotransposition, HeLa-JVM or HEK293T cells were co-transfected with a full-length RC-L1 (pTMF3) and either the HELZ2, IFIT1, DDX60L, OASL, HERC5, or MOV10 expressing vectors, and investigated whether the ISG proteins affected ORF1p and/or L1 RNA levels. A similar data trend in HeLa-JVM and HEK293T cells was observed by western blot experiments: ORF1p steady state levels were severely decreased by co-expression of HERC5, HELZ2, and MOV10, were modestly reduced by the co-expression of OASL, but were not significantly changed by the co-expression of IFIT1 or DDX60L (**Figs. 3.7c and 3.7d**). RT-qPCR analyses revealed that HELZ2 co-expression significantly reduced L1 RNA levels in HeLa-JVM cells; a probe set that specifically recognizes the SV40 poly(A) signal of the plasmid-expressed L1 RNA was used (**Fig. 3.8**, ~90% reduction of the WT L1 control). Consistent with previous reports, MOV10 co-expression resulted in a ~70% reduction in L1 RNA when compared to the WT L1 control (Li et al., 2013; Warkocki et al., 2018).

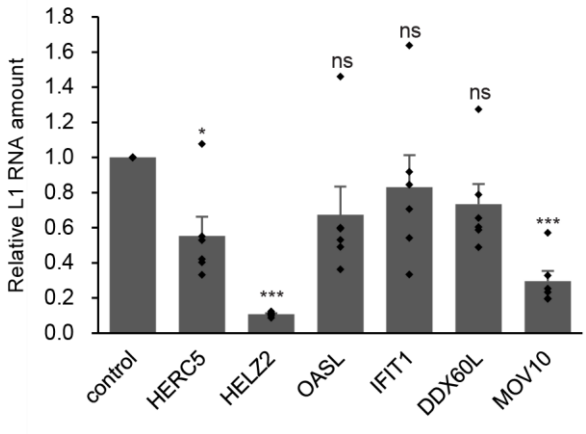


Fig. 3.8: *L1 RNA steady state level is severely reduced by HELZ2 expression.* HeLa-JVM cells were transfected pTMF3, and either pCMV-3Tag-8-Barr or one of the following carboxyl-terminal 3xMYC epitope-tagged ISG protein expression plasmids: pALAF015 (HELZ2), pALAF016 (IFIT1), pALAF021 (DDX60L), pALAF022 (OASL), pALAF023 (HERC5), or pALAF024 (MOV10). L1 RNA levels were determined by performing RT-qPCR using a primer set specific to RNAs derived from the transfected L1 (primer set: L1 [SV40]) and then were normalized to *ACTB* RNA levels (primer set: Beta-actin). X-axis, name of the constructs. Y-axis, relative level of L1 RNA normalized to the ORF1-T7 + pCMV-3Tag-8-Barr control. Pairwise comparisons relative to the control: $p = 0.032^{*}$ (HERC5); $1.7 \times 10^{-5***}$ (HELZ2): 0.14^{ns} (OASL); 0.29^{ns} (IFIT1); 0.20^{ns} (DDX60L); and $4.4 \times 10^{-4***}$ (MOV10). Values represent the mean ± SEM from three independent biological replicates. The *p*-values were calculated using a one-way ANOVA followed by a Bonferroni-Holm post-hoc tests. ns: not significant; * $p<0.05$; *** $p<0.001$.

Co-transfection of pJM101/L1.3FLAG (WT ORF1p-FLAG) with the individual ISG protein expression vectors (i.e., HELZ2, HERC5, OASL, and MOV10) effects on ORF1p-FLAG cytoplasmic foci formation in HeLa-JVM cells were also examined (**Fig. 3.9a**). Consistent with previous results (Doucet et al., 2010), more than 70% of transfected cells expressing WT ORF1p-FLAG were observed to form ORF1p cytoplasmic foci (**Fig. 3.9b**). HERC5 co-expression did not show dramatic effect on ORF1p cytoplasmic foci formation in HeLa-JVM cells (**Fig. 3.9b**, ~55% of cells contained ORF1p cytoplasmic foci, some of the ORF1p foci was observed to associate with HERC5 foci). By comparison, HELZ2, OASL, and MOV10 co-expression resulted in a decrease in ORF1p-FLAG cytoplasmic foci (**Fig. 3.9b**, ~30%, ~15%, and ~5% of cells, respectively) and these foci was not observed to associate with the relevant ISG protein (**Fig. 3.9a**). In sum, these data suggest: (1) HERC5 destabilizes ORF1p, but does not affect its cellular localization; (2) OASL mainly impairs ORF1p cytoplasmic foci

formation; and (3) HELZ2 reduces the levels of L1 RNA, ORF1p, and ORF1p cytoplasmic foci formation. Thus, different ISGs appear to affect different steps of the L1 retrotransposition cycle. As HELZ2 negatively affects L1 in all the experiments performed, i.e., decrease of L1 RNA and ORF1p steady state levels, ORF1p cytoplasmic foci formation, and L1 retrotransposition efficiency, I decided to study HELZ2 mechanism of action in more detail.

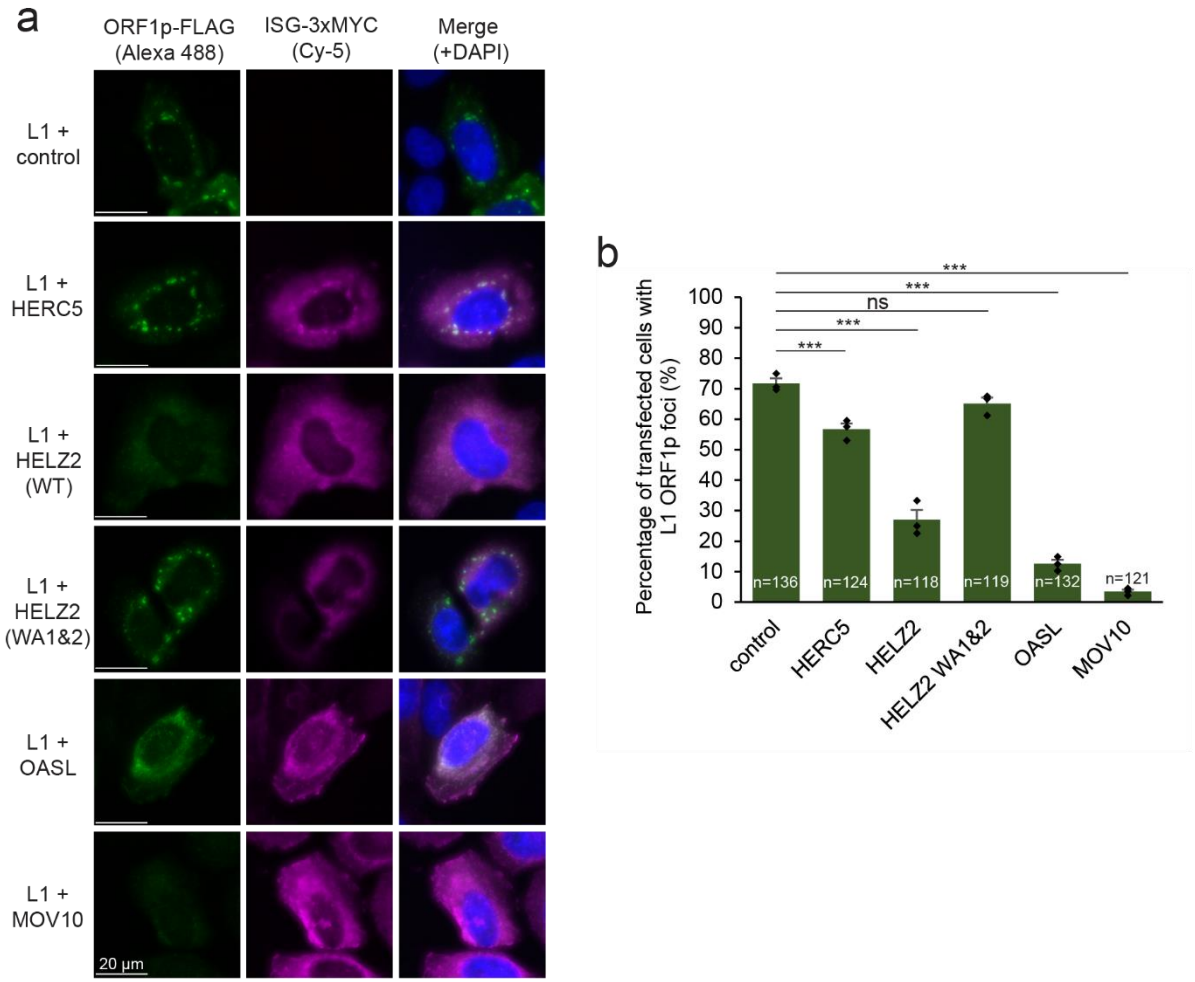


Fig. 3.9: *ISG proteins effects on ORF1p-FLAG cytoplasmic foci formation*. **(a)** *Representative images of ISG co-expression effects on ORF1p-FLAG cytoplasmic formation*. HeLa-JVM cells were co-transfected with pJM101/L1.3FLAG (WT ORF1p-FLAG) and either a pCEP4 empty vector (control) or one of the following carboxyl-terminal 3xMYC epitope-tagged ISG protein expression plasmids: pALAF015 (HELZ2); pALAF027 (HELZ2 WA1&2); pALAF022 (OASL); pALAF023 (HERC5); or

pALAF024 (MOV10) to visualize WT ORF1p-FLAG cytoplasmic foci and co-localization between WT ORF1p-FLAG and the candidate ISG protein. Representative fluorescent microscopy images are shown. White scale bars, 20 µm. **(b)** *L1 cytoplasmic foci formation quantification.* X-axis, name of the constructs co-transfected with pJM101/L1.3FLAG (WT ORF1p-FLAG); control, pCEP4. Y-axis, percentage of transfected cells with visible ORF1p signal exhibiting ORF1p-FLAG cytoplasmic foci. The numbers (n) within the green rectangles indicate the number of analyzed cells in each experiment. Pairwise comparisons relative to the pJM101/L1.3FLAG (WT ORF1p-FLAG) + pCEP4 control: p = 8.6 x 10^{-4}*** (HERC5); 1.2 x 10^{-7}*** (HELZ2); 0.098^{ns} (HELZ2 WA1&2); 1.0 x 10^{-10}*** (OASL); 2.7 x 10^{-9}*** (MOV10). Values represent the mean ± SEM from independent biological replicates. The p-values were calculated using one-way ANOVA followed by Bonferroni-Holm post-hoc tests; ns: not significant; *** p<0.001.

Protein FDR Confidence: Combined	High	High	High	Medium	Low
Accession	Q08211	P48634	Q86SQ0-3	Q96FV9	E9PE96
Description	ATP-dependent RNA helicase A OS=Homo sapiens OX=9606 GN=DHX9 PE=1 SV=4	Protein PRRC2A OS=Homo sapiens OX=9606 GN=PRRC2A PE=1 SV=3	Isoform 3 of Pleckstrin homology-like domain family B member 2 OS=Homo sapiens OX=9606 GN=PHLDB2	THO complex subunit 1 OS=Homo sapiens OX=9606 GN=THOC1 PE=1 SV=1	Protein piccolo OS=Homo sapiens OX=9606 GN=PCLO PE=1 SV=1
Coverage [%]	59	53	58	1	1
# Peptides	68	91	80	1	1
# PSMs	1863	574	996	1	6
# Unique Peptides	68	87	2	1	1
# AAs	1270	2157	1237	657	1217
MW [kDa]	140.9	228.7	139.9	75.6	136.3
calc. pI	6.84	9.45	7.12	4.98	7.9
# Razor Peptides	0	0	0	0	0
Abundance Ratio: (WT) / (M8)	3.007	1.382	0.043	2.168	0.698
Abundance Ratio P-Value: (WT) / (M8)	0.106005	0.134432	0.000738	0.192712	0.990043
Abundance Ratio Adj. P-Value: (WT) / (M8)	0.286181	0.3193	0.066003	0.388785	0.991904
Abundances (Grouped): WT	1.76E+10	8.81E+08	2305458	2246661	10564377
Abundances (Grouped): M8	5.85E+09	6.38E+08	53012380	1036502	15141478
Abundances (Grouped) CV [%]: WT	22.78	17.37	25.11	59.28	57.57
Abundances (Grouped) CV [%]: M8	64.2	37.52	42.45	28.89	34.49
Found in Sample: [S2] F2: Sample, WT	High	High	Peak Found	Peak Found	Peak Found
Found in Sample: [S5] F5: Sample, WT	High	High	Peak Found	High	Peak Found
Found in Sample: [S8] F8: Sample, WT	High	High	Peak Found	Peak Found	High
Found in Sample: [S3] F3: Sample, M8	High	High	High	Not Found	Peak Found
Found in Sample: [S6] F6: Sample, M8	High	High	High	Peak Found	Peak Found
Found in Sample: [S9] F9: Sample, M8	High	High	High	Peak Found	High

Legend is shown in the next page…

Source Data 1: *Representative label-free quantitative analysis data from IP-MS peptide hits processed by Proteome Discoverer 2.3*. Leftmost column, parameters for each protein group; other columns, representative protein groups showing high, medium, or low combined protein FDR confidence**. Legends**: Protein FDR Confidence: Combined, level of confidence of identified proteins that are determined by the Protein FDR Validator node in Proteome Discoverer 2.3 (High < 0.01, 0.01 ≤ Medium < 0.05, 0.05≤ Low); Accession, accession number of each protein hits assigned by UniProt KB; Description, description of the protein based on the accession number; Coverage [%], percentage of the protein sequences covered by identified peptides; # Peptides, total number of distinct peptide sequences (both unique and razor peptides); # PSMs, total number of all identified peptide spectrum matches (including redundant peptides); # Unique Peptides, number of distinct peptide sequences unique to the protein group; # AAs, amino acid length of the protein sequences; MW [kDa], molecular weight of the protein in kDa; calc. pI, theoretically calculated isoelectric point of the proteins; # Razor Peptides, peptides that have been assigned to the protein group with the largest number of total peptide combined with the unique peptides; Abundance Ratio: (WT) / (M8), abundance ratios of the values of Abundances (Grouped): WT *vs.* Abundances (Grouped): M8 (RBM); Abundance Ratio P-Value: (WT) / (M8), p-values of abundance ratio of WT *vs.* M8 (RBM) calculated by running the Tukey HSD test (post-hoc) after an analysis of variance (ANOVA) test; Abundance Ratio Adj. P-Value: (WT) / (M8), p-values adjusted by using the Benjamini-Hochberg correction for the false discovery rate; Abundances (Grouped): WT, abundance values of WT samples combined; Abundances (Grouped): M8, abundance values of M8 (RBM) samples combined; Abundances (Grouped) CV [%]: WT, coefficient of variation (%) of WT samples combined; Abundances (Grouped) CV [%]: M8, coefficient of variation (%) of M8 (RBM) samples combined; Abundances (Normalized): WT, normalized abundance values of WT samples with the ORF1p peptides; Abundances (Normalized): M8, normalized abundance values of M8 (RBM) samples with the ORF1p peptides; Abundances : WT, abundance values of WT samples; Abundances : M8, abundance values of M8 (RBM) samples; Found in Sample, peaks found in each of the samples (3 WT, and 3 M8 (RBM) samples, respectively). (High = peak with high confidence, Peak Found = peak with medium or low confidence, Not Found = peak was not found)

Gene	log_2 of abundance ratio	Gene	log_2 of abundance ratio	Gene	log_2 of abundance ratio	Gene	log_2 of abundance ratio
NOC4L	7.034326	PKP4	3.125155	KIF5B	2.463361	LGALS3BP	2.173447
CCNT1	6.571753	RNPC3	3.115366	ALYREF	2.461529	HNRNPC	2.168
HEXIM1	5.635987	TRMT2A	3.10249	HNRNPDL	2.440421	THOC2	2.164786
CDK9	5.479651	RRP12	3.084745	PACSIN3	2.435095	HNRNPA3	2.163177
PNO1	5.288137	EDC4	3.070389	HNRNPL	2.416029	KBTBD3	2.148934
RIOK2	5.078567	PWP2	3.0358	DDX5	2.411426	GLE1	2.144699
DDX10	5.004546	AFF1	3.008451	HNRNPA2B1	2.401903	MSI2	2.140779
CCNT2	4.854145	SNRPC	3.007375	ZCCHC8	2.394788	TBL3	2.121015
LTV1	4.826599	DDX6	2.998196	U2AF1	2.385431	EMG1	2.116697
SCAF11	4.748193	LUC7L	2.904773	IFRD2	2.379066	G3BP2	2.111365
IPO7	4.493455	WDR3	2.901881	PAIP2	2.371001	CCDC12	2.105678
FYN	4.445462	CSE1L	2.856587	CSTF2	2.369048	SNRPE	2.096262
ZNF277	4.401426	NLE1	2.842979	MTDH	2.365972	HNRNPAB	2.095587
DHX58	4.218316	RBMX2	2.839355	WDR36	2.365693	SRRT	2.094236
PABPN1	4.162855	SRSF3	2.838548	DLG5	2.356707	LSM6	2.086444
SPATA5L1	4.031042	AQR	2.834104	AKAP17A	2.348516	TIMM50	2.08168
LARP7	4.017566	RO60	2.829647	PRPF40A	2.34568	LUC7L3	2.079975
MEPCE	3.942515	HNRNPA1	2.826599	FIP1L1	2.338282	RRP7A	2.076901
SON	3.916763	MRPL55	2.825175	ZCRB1	2.335998	IFI16	2.073135
BYSL	3.78419	U2AF2	2.822934	EFTUD2	2.329124	SRSF10	2.067639
KLC1	3.749749	ELL2	2.814961	SSB	2.316146	TMA16	2.067639
TRMT61A	3.74685	EAF1	2.810237	CLK3	2.314986	NCBP3	2.065917
LSM2	3.711164	ATXN2L	2.799502	KIF1C	2.304803	RALYL	2.049631
NXF1	3.676267	SRSF9	2.799087	MRPS6	2.290424	EXOSC3	2.03984
CDK12	3.57022	SPTBN1	2.793688	DDX46	2.285106	EXOSC6	2.025029
LGALS1	3.550285	JUP	2.78136	EIF4E	2.268434	ZBTB11	2.014712
LSM4	3.520799	WTAP	2.78031	NCL	2.260327	KIF1B	2.004322
HNRNPH3	3.490698	LSM10	2.756596	SF3A1	2.260327	RPF1	2.000361
SLX9	3.457332	HNRNPD	2.738984	LARP1	2.25882	NSUN5	1.995304
SRPRA	3.435095	ZC3H13	2.709511	AATF	2.251568	THAP11	1.989502
SREK1	3.428812	ARGLU1	2.67333	LSM14B	2.245191	SCAF8	1.986593
DDX47	3.425862	METTL17	2.665393	MTERF1	2.24245	LTF	1.985865
TNRC6A	3.413052	TRMT6	2.643625	LRPPRC	2.236033	NOL6	1.976364
NOB1	3.406673	POLR1F	2.636451	PNPLA6	2.226817	EIF4G3	1.973428
GNG12	3.404631	NUP42	2.624569	PDCD7	2.226509	SAP18	1.970117
CLK2	3.363732	APOBEC3C	2.602409	MBD4	2.217231	RBM7	1.966061
EXOSC2	3.351063	ZC3H3	2.593115	SPATA5	2.21692	CRNKL1	1.949722
MRPS18B	3.218781	AIMP1	2.567059	PLOD3	2.214747	HNRNPA0	1.944109
KLC2	3.211012	TSR1	2.559492	DHX37	2.197865	VIRMA	1.943359
CPSF6	3.189508	SLIRP	2.551885	MRPS21	2.190931	WDR74	1.936968
MLLT1	3.180466	NFX1	2.496207	SNRNP200	2.187768	NOMO3	1.936591
AFF4	3.167519	CFAP20	2.486714	SRSF5	2.18301	CHD3	1.933195
MPHOSPH6	3.157529	AP2B1	2.469626	LSM7	2.182057	RTRAF	1.932061

Gene	log_2 of abundance ratio	Gene	log_2 of abundance ratio	Gene	log_2 of abundance ratio	Gene	log_2 of abundance ratio
RRS1	1.931305	G3BP1	1.773152	EIF2S1	1.617298	SENP3-EIF4A1	1.486457
YTHDC2	1.92448	TSNAX	1.761285	DDX1	1.601221	PRPF4B	1.483364
MOV10	1.913033	CAPRIN1	1.760434	SNRPB2	1.597889	PUM2	1.481299
LSM1	1.913033	RBMS1	1.748461	AIMP2	1.592636	TAF1A	1.476122
CPSF2	1.910349	SUGP2	1.742437	KRR1	1.591679	NOP10	1.469886
PTBP1	1.906891	HPSE	1.741143	ACIN1	1.589763	ZNF326	1.468323
MATR3	1.904966	PHF6	1.738552	ADAR	1.589284	SNU13	1.358959
S100A10	1.90458	MRPS10	1.733354	FXR1	1.589284	IGF2BP3	1.356707
DDX41	1.898402	GADD45GIP1	1.731618	DHX9	1.588325	RBM19	1.354452
GEMIN2	1.894139	RBM42	1.728138	YWHAG	1.586885	XRN1	1.345396
LSM12	1.878921	SNCA	1.726395	MKRN1	1.584963	HNRNPUL2	1.343124
RBMS2	1.877744	BLM	1.723777	MKRN2	1.58111	KHDRBS1	1.342555
RRP9	1.873026	EIF5B	1.719841	PEG10	1.57918	MRPL24	1.341417
RPF2	1.868292	HNRNPU	1.715015	EXOSC7	1.567059	AGO1	1.339708
RBM15	1.865127	ATXN10	1.707525	RBM14	1.566572	YBX3	1.337996
SAFB2	1.861161	ZC3H18	1.702214	TUBA4A	1.563646	AGO3	1.333996
MRPS28	1.849199	ELAVL1	1.699107	PPIH	1.539035	SRSF7	1.332851
EXOSC8	1.841571	LSM14A	1.694657	C1orf167	1.530071	KIAA1522	1.332851
RTCB	1.835924	CLTC	1.69332	MRPL48	1.529071	EXOSC5	1.332278
MRPS23	1.835116	HNRNPCL4	1.685267	DKC1	1.527571	HNRNPH2	1.329411
SF3A3	1.834307	LSM3	1.676719	DDX3X	1.525568	MYO1E	1.329411
SNRPG	1.831067	POLR2F	1.674913	POLR1B	1.522558	PCBP2	1.329411
IFIT1	1.825379	HNRNPR	1.671746	POP4	1.518031	MRPS22	1.328836
MYBBP1A	1.823749	RBM47	1.668119	TCAF1	1.516519	DDX17	1.327112
PTBP3	1.82171	ZFP36L2	1.665393	CHTOP	1.51248	DNA2	1.327112
EXOSC1	1.819668	CPSF7	1.664938	SNRNP70	1.511468	POP1	1.325386
PUF60	1.809003	FGL2	1.664938	EPRS1	1.510456	ZNF346	1.325386
XPA	1.808591	MRM1	1.662661	SNRNP35	1.509949	RBM45	1.324235
PIP5K1A	1.807355	PTBP2	1.662661	RBBP6	1.508936	TCOF1	1.318461
GNA12	1.80653	DROSHA	1.662205	FBXW4	1.508429	HNRNPF	1.315566
MYH6	1.798258	CCDC86	1.662205	ZNF638	1.506907	SNRPD2	1.311503
PATL1	1.798258	GRSF1	1.660381	SEC61G	1.505891	TRA2B	1.309758
TTF1	1.797013	FAM98A	1.65764	MTREX	1.505383	MAGOH	1.307429
EIF4G1	1.792022	ICE2	1.655352	SART3	1.503858	MRPS27	1.304511
THAP12	1.78785	SNRPD1	1.654894	SYNCRIP	1.500292	HELZ2	1.299245
VANGL1	1.78785	TARDBP	1.644779	MRPS24	1.499782	SRBD1	1.296311
SRSF1	1.787014	PRR3	1.643856	RBM39	1.499272	PRKDC	1.290424
EIF2AK2	1.786596	SRSF8	1.641546	NCOA5	1.498761	PES1	1.288654
AHNAK	1.784504	MRPS16	1.641084	POLR1A	1.498251	FAM120C	1.286881
IFIT3	1.78157	CCDC47	1.641084	TFAP2B	1.497229	POLR2E	1.28629
LUC7L2	1.78031	NGDN	1.637378	IARS1	1.496207	ZNF66	1.280956
SRSF2	1.777788	TUT7	1.634129	PELP1	1.49057	PDCD11	1.276199
IPO5	1.775683	NVL	1.620117	NHP2	1.486457	PHRF1	1.276199

Gene	log_2 of abundance ratio	Gene	log_2 of abundance ratio	Gene	log_2 of abundance ratio	Gene	log_2 of abundance ratio
GRWD1	1.273217	FXR2	1.212258	SFSWAP	1.132248	PABPC4L	1.041243
TFB1M	1.273217	TTN	1.212258	NIFK	1.130272	KDELR2	1.037031
SRP14	1.27262	PPAN-P2RY11	1.206643	TIMM13	1.130272	ZC3H15	1.037031
GPATCH8	1.270828	RBM34	1.203514	ZNF574	1.123666	DHX30	1.03492
TSN	1.269033	EPPK1	1.195348	YBX2	1.123666	MRPS31	1.033511
GNL2	1.266637	RRBP1	1.194717	DIS3L2	1.122341	SHFL	1.032101
SRP9	1.266037	FGF2	1.189034	U2SURP	1.120352	ELAVL2	1.028569
PPIG	1.264837	PNN	1.183963	EME1	1.119688	NUDT21	1.027862
SEC61A2	1.263635	CPSF3	1.183328	IGF2BP1	1.115033	DGCR8	1.026446
KCNMA1	1.26123	RBMXL1	1.182692	IFIT2	1.115033	SART1	1.02432
RRP15	1.260628	RBFOX1	1.181421	SPATS2	1.113034	SLC25A11	1.0229
EIF3C	1.260026	CHD6	1.180784	EXOSC9	1.111699	AGO2	1.020769
RPS15	1.260026	EIF5	1.179511	MRPS35	1.111031	PRPF31	1.019346
MRPS5	1.259423	RBM8A	1.178874	EEF2	1.107688	XRN2	1.013641
PRPF8	1.25882	TNRC6B	1.178874	NIP7	1.105678	HAX1	1.012926
DDX51	1.25882	MOGS	1.176323	CSNK1A1	1.105008	ZNF16	1.010064
EIF4A3	1.257614	ZFC3H1	1.175684	URB1	1.102322	PATZ1	1.005759
NCLN	1.256407	APOBEC3B	1.169925	MRPS17	1.098958	GEMIN5	1.005041
PCBP1	1.253384	SF3B5	1.162855	WDR18	1.097611	PLCD3	1.002883
MRPS26	1.250962	TRIM56	1.16092	RSL1D1	1.096936	BTF3L4	1.002162
WDR12	1.249142	HNRNPH1	1.160275	UTP18	1.090853	GEMIN4	1.001442
PCID2	1.24732	TOP3B	1.160275	CDKN2AIP	1.08134	ZNF512	1.001442
KTN1	1.245496	PAIP1	1.160275	DHX36	1.079975	UTP20	0.999278
IFRD1	1.243669	ZC3H8	1.160275	MYO1C	1.079975	KNOP1	0.999278
PGAM5	1.24062	GPATCH11	1.155749	DDB1	1.079975	MTCL1	0.997112
FUS	1.24062	RPL7L1	1.155102	TAF1B	1.07861	SNRPB	0.994942
RRP1	1.238175	DDX50	1.153157	NPM1	1.077243	CPS1	0.994942
PHF5A	1.236952	CPEB4	1.14991	SF3B2	1.07519	MRPL9	0.994218
FAM120A	1.23634	ZC3HAV1	1.149259	YTHDC1	1.07519	TRIM25	0.992043
SRPK1	1.233888	CCDC124	1.149259	ERH	1.073135	RPL18	0.989139
PRRC2C	1.233275	SEMA3B	1.147307	MRPS2	1.072449	DDX54	0.988412
ZCCHC3	1.233275	MRPS7	1.145351	SPOUT1	1.067639	MRPL37	0.986957
MRPS30	1.233275	UPF3B	1.144699	RPS19BP1	1.06695	SMARCA1	0.986957
FASTK	1.232661	CHD2	1.144046	POLRMT	1.060047	YBX1	0.9855
TFAM	1.229588	UTP23	1.14274	EIF3K	1.05797	NMT1	0.9855
PUM1	1.227741	FMR1	1.142087	YWHAZ	1.05589	PUM3	0.983313
RALY	1.223423	NUFIP1	1.141433	EIF3D	1.050328	SFPQ	0.977463
DAP3	1.222805	EIF4G2	1.140779	MRPS15	1.04684	WDR33	0.97673
SRFBP1	1.220949	MRPS34	1.140124	MRPL47	1.046142	XAB2	0.97673
MRPS11	1.215989	BTF3	1.138159	IGHMBP2	1.045443	LRRC59	0.97673
SURF6	1.214747	LARP1B	1.136191	CMAS	1.044744	C8orf33	0.975997
RPP25L	1.214125	MRPS33	1.136191	SPTY2D1	1.043345	PLOD2	0.97159
FBL	1.213503	WDR46	1.132248	RPP25	1.041944	CCDC137	0.968644

Gene	log_2 of abundance ratio	Gene	log_2 of abundance ratio	Gene	log_2 of abundance ratio	Gene	log_2 of abundance ratio
EIF3B	0.967169	GNB1	0.887525	SF3B4	0.827006	DDX56	0.768078
ARHGEF2	0.967169	LSG1	0.885184	SRSF11	0.827006	SRRM1	0.767231
PABPC1	0.96643	MBNL3	0.884403	PRPF4	0.823749	PABPC1L	0.759582
USP36	0.96643	TBC1D10B	0.883621	TIMM8B	0.820485	PHAX	0.759582
ZCCHC17	0.965692	DPM1	0.882056	DYNLL2	0.818851	SRP72	0.757023
RING1	0.961253	LYAR	0.88049	ASPH	0.818032	PLRG1	0.756169
MRPL18	0.95977	EIF3L	0.879706	CNTNAP4	0.817214	FBXW11	0.75446
STAU2	0.956057	NOL9	0.877352	MRPL53	0.816395	ELAVL3	0.752749
RFC1	0.955313	MYO3B	0.876566	SEC61B	0.814755	RUVBL2	0.751035
NOP2	0.954569	RBBP7	0.87578	TRIM26	0.813935	LAS1L	0.74932
DDX24	0.953079	SRP68	0.874207	KCMF1	0.813935	PLEC	0.748461
SF3B6	0.952334	EXOSC10	0.868687	SMARCC1	0.813114	DDX55	0.747602
WBP11	0.948601	PURA	0.867106	EIF3G	0.812293	RPS27L	0.747602
PLOD1	0.944858	MRPL40	0.865523	PTDSS1	0.810649	SRSF4	0.745022
SF3B1	0.944109	NEPRO	0.863938	BAZ1A	0.809826	OGT	0.745022
RBMX	0.943359	ZBTB24	0.863146	DDX60L	0.808179	POLR1C	0.744161
HNRNPM	0.941858	MAP7D1	0.859174	BCAS2	0.805705	PWP1	0.744161
SRP54	0.935083	EIF3F	0.857583	RAN	0.804054	OASL	0.7433
CHD1	0.933573	ZCCHC4	0.85599	TOP2B	0.801573	SND1	0.741575
NSUN2	0.929791	SKP1	0.853597	STRBP	0.800745	SPTAN1	0.730314
COPA	0.929791	ZCCHC10	0.853597	GIGYF2	0.799916	BAZ1B	0.727703
ZFR	0.929033	DDX52	0.852798	MMTAG2	0.791606	GDI2	0.727703
Nucleolar protein 12 (Fragment)	0.925999	STRAP	0.852798	MRPL34	0.791606	WDR76	0.725959
RBM25	0.92524	PHF8	0.852798	REV1	0.790772	EIF3I	0.725087
NOP53	0.92524	CPSF1	0.851999	FSCN1	0.789938	CCDC59	0.725087
CCAR2	0.919149	PARP12	0.851999	SPATS2L	0.788268	EIF2S2	0.725087
AP2M1	0.916094	SNRPA	0.851199	MKI67	0.78576	NAT10	0.724214
SRRM2	0.913033	CSDE1	0.849599	SEC61A1	0.784923	DDX20	0.721591
SSR1	0.913033	RPP38	0.849599	DHX15	0.784085	DYNLL1	0.720716
TENT4B	0.9115	MARS1	0.849599	OLA1	0.784085	BANF1	0.719841
SRSF6	0.905351	MRPL27	0.848798	DNTTIP1	0.782409	EIF3H	0.71721
HNRNPUL1	0.90381	DDX18	0.847195	FUBP3	0.782409	PPHLN1	0.713696
LMNA	0.901494	CASC3	0.847195	NONO	0.78073	SDAD1	0.711054
SRP19	0.901494	PRPF19	0.842375	GTF3C1	0.778209	TOE1	0.711054
BMS1	0.899949	FMNL2	0.840765	TBPL2	0.775683	TMEM33	0.710173
ILF2	0.897628	CHERP	0.838347	SNRPD3	0.77484	CKAP2L	0.707525
TDRD3	0.896853	MRPL46	0.836732	NSD2	0.77484	EED	0.703987
TOP2A	0.896078	NUFIP2	0.835116	RPP30	0.771463	RPL10	0.703101
CRTAP	0.893751	EIF3E	0.832688	TMCO1	0.770618	CCDC77	0.703101
LARP4	0.892974	ZNF787	0.832688	TAF6L	0.770618	GPATCH4	0.70044
MRPL54	0.891419	MAGEB1	0.831067	DDX21	0.768925	EIF3A	0.699552
DHX33	0.888305	SRPK2	0.831067	HERC5	0.768078	MYO5A	0.694212

Gene	log_2 of abundance ratio	Gene	log_2 of abundance ratio	Gene	log_2 of abundance ratio
REXO4	0.694212	BAG1	0.60502	RPS27	0.528071
MRPS14	0.694212	ABT1	0.601221	POLR1G	0.528071
TEX10	0.690641	S100P	0.60027	UBAP2L	0.528071
MRPL16	0.687061	TFRC	0.60027	RBM17	0.525067
BCLAF1	0.68437	COIL	0.593593	ZNF692	0.52306
EIF4E2	0.678072	LARS1	0.590722	ILF3	0.522056
CAAP1	0.678072	MRPL38	0.590722	BRIX1	0.521051
TSPYL2	0.675364	IQGAP1	0.589763	C7orf50	0.518031
PRKRA	0.673556	ELOC	0.587845	POLR1D	0.515006
PURB	0.67084	RPS29	0.584	CELF1	0.513996
MYH10	0.669934	BOP1	0.584	CENPV	0.513996
RPL38	0.669027	YWHAE	0.583038	ADARB1	0.513996
CUL4B	0.668119	DDX60	0.582074	UTP3	0.509949
RPS28	0.666302	TBL2	0.582074	ATAD3B	0.507921
EWSR1	0.663572	RPP14	0.58111	RECQL	0.504875
RBM28	0.662661	TSPYL1	0.58111		
POLR2L	0.662661	TRAP1	0.57918		
RUVBL1	0.66175	CDC5L	0.571434		
DNAJC9	0.659925	CEBPZ	0.571434		
PDCD6	0.657183	GNAS	0.571434		
P4HA1	0.652601	MRPL12	0.569491		
CMSS1	0.650765	GTF3C3	0.565597		
MRPL58	0.649845	PPIE	0.565597		
DUSP11	0.648005	TRMT112	0.558757		
PRPF3	0.646163	SQOR	0.557778		
DIMT1	0.643394	RAB1A	0.555816		
YTHDF1	0.642471	ANXA2	0.553852		
SLC25A12	0.634129	STAU1	0.550901		
ZNF7	0.633199	NEDD8	0.550901		
CYFIP1	0.632268	RPL11	0.549916		
TRIM28	0.627607	EBNA1BP2	0.54893		
CAVIN1	0.626673	YTHDF3	0.546956		
MAP7	0.624803	QARS1	0.545968		
C18orf21	0.623867	SLC16A3	0.545968		
DDX31	0.618239	WRAP53	0.54498		
POLR2J	0.615416	KPNA2	0.541019		
ENY2	0.614474	ATP5MC1	0.539035		
POP7	0.612588	SLC3A2	0.538041		
MTPAP	0.611645	ZNF142	0.535058		
YTHDF2	0.609755	MRPL28	0.531069		
TRAM1	0.608809	NOC2L	0.530071		
CPSF4	0.607863	GLYR1	0.529071		
SF3B3	0.605968	PRMT5	0.529071		

Source Data 2: *Genes with > 0.5 log_2(abundance ratio of WT vs. M8 [RBM] ORF1p complexes)*. UniProt accession numbers (see **Source Data 1**) were converted to the respective gene symbols and sorted in descending order (top to bottom, followed by left to right) of the log_2(abundance ratio of WT vs. M8 [RBM] ORF1p complexes).

GO Term	Count	P-Value	FDR	-log$_{10}$FDR
mRNA processing	155	8.20E-107	5.20E-105	104.2839967
Transcription	144	6.40E-03	2.50E-02	1.602059991
mRNA splicing	133	2.60E-97	8.10E-96	95.09151498
Host-virus interaction	70	3.80E-11	3.00E-10	9.522878745
rRNA processing	55	3.70E-45	7.80E-44	43.1079054
Ribosome biogenesis	38	2.20E-26	3.40E-25	24.46852108
Protein biosynthesis	38	1.40E-16	1.50E-15	14.82390874
Translation regulation	37	5.50E-18	6.90E-17	16.16115091
Innate immunity	36	2.70E-04	1.20E-03	2.920818754
mRNA transport	28	7.20E-12	6.50E-11	10.18708664
Antiviral defense	22	1.60E-06	9.90E-06	5.004364805
Translocation	16	1.70E-05	9.50E-05	4.022276395
tRNA processing	15	3.50E-04	1.50E-03	2.823908741
Biological rhythms	14	2.90E-02	1.10E-01	0.958607315
RNA-mediated gene silencing	13	3.00E-05	1.40E-04	3.853871964
Nonsense-mediated mRNA decay	10	1.90E-05	9.90E-05	4.004364805
Transcription termination	7	3.30E-07	2.30E-06	5.638272164

Table 2: *DAVID Gene Ontology analysis (Functional annotation set UP_KW (UniProt_KeyWord) biological process) of a >0.5 cutoff of log$_2$(abundance ratio WT vs. M8[RBM]).* The GO terms are arranged in descending order of the count. FDR, False Discovery Rate. Red lettering, viral-related GO terms.

Gene Set	Size	Enrichment Score (ES)	Normalized Enrichment Score (NES)	Nominal (NOM) p-value	False Discovery Rate (FDR) q-value
HALLMARK_UNFOLDED_PROTEIN_RESPONSE	46	0.49	1.63	0.001	0.075
HALLMARK_MYC_TARGETS_V1	121	0.32	1.21	0.138	0.928
HALLMARK_MYC_TARGETS_V2	31	0.37	1.14	0.291	0.853
HALLMARK_INTERFERON_ALPHA_RESPONSE	17	0.41	1.12	0.315	0.685
HALLMARK_G2M_CHECKPOINT	48	0.29	0.99	0.495	0.875
HALLMARK_INTERFERON_GAMMA_RESPONSE	21	0.33	0.95	0.542	0.836
HALLMARK_P53_PATHWAY	19	0.27	0.76	0.79	1
HALLMARK_HEME_METABOLISM	15	0.28	0.72	0.832	0.97
HALLMARK_UV_RESPONSE_UP	20	0.24	0.7	0.87	0.89
HALLMARK_OXIDATIVE_PHOSPHORYLATION	34	-0.17	-0.67	0.946	0.932
HALLMARK_E2F_TARGETS	15	-0.23	-0.74	0.806	0.971
HALLMARK_DNA_REPAIR	36	-0.2	-0.82	0.794	0.98
HALLMARK_EPITHELIAL_MESENCHYMAL_ TRANSITION	50	-0.21	-0.89	0.719	0.953
HALLMARK_MITOTIC_SPINDLE	35	-0.24	-0.98	0.533	0.899
HALLMARK_MTORC1_SIGNALING	42	-0.35	-1.43	0.026	0.13
HALLMARK_ALLOGRAFT_REJECTION	19	-0.47	-1.62	0.023	0.05
HALLMARK_GLYCOLYSIS	25	-0.55	-1.96	0	0.002
HALLMARK_HYPOXIA	19	-0.68	-2.35	0.005	0.002

Table 3: *Hallmark list from preranked Gene Set Enrichment Analysis of log_2(abundance ratio WT vs. M8[RBM])*. Hallmark gene sets from MSigDB were used in the analysis. Red lettering, interferon gene sets.

Set	Number	1	2	3	4	5	6	7	8
	Cytokines	**APRIL (42)**	**BAFF (37)**	**CD30 (53)**	**CD163 (46)**	**Chitinase-3 (72)**	**gp130 (14)**	**IFN-a2 (20)**	**IFN-b (44)**
	Description	**Obs Conc (pg/mL)**	**Obs Conc (pg/mL)**	**Obs Conc (pg/mL)**	**Obs Conc (pg/mL)**	**Obs Conc (pg/mL)**	**Obs Conc (pg/mL)**	**Obs Conc (pg/mL)**	**Obs Conc (pg/mL)**
(a)	**control**	2401.43	405.14	*0.80	435.52	36.59	*9.72	OOR <	OOR <
	poly(I:C)	2401.43	399.94	*1.88	800.92	OOR <	*4.69	15.51	41.27
(b)	**pCEP4 24h**	1783.78	*139.66	*2.62	435.52	28.21	*4.69	OOR <	OOR <
	L1 WT 24h	1279.82	168.68	*3.36	OOR <	32.59	*6.33	OOR <	2.35
	L1 M8 24h	3117.89	*13.62	*4.48	237.79	43.76	*8.01	OOR <	1.45
	L1 RT- 24h	*250.66	324.74	*2.25	115.81	28.21	*7.17	11.26	0.24
(c)	**pCEP4 48h**	1452.93	192.99	*1.88	78.93	36.59	*10.59	OOR <	OOR <
	L1 WT 48h	3526.61	214.33	OOR <	34.84	50.18	*11.03	OOR <	1.77
	L1 M8 48h	7137.98	367.2	11.34	945.69	66.71	*17.73	32.33	9.07
	L1 RT- 48h	1783.78	297.65	OOR <	OOR <	38.47	*8.01	OOR <	8.38

Set	Number	9	10	11	12	13	14	15	16
	Cytokines	**IFN-g (21)**	**IL-2 (38)**	**IL-6Ra (19)**	**IL-8 (54)**	**IL-10 (56)**	**IL-11 (39)**	**IL-12 (p40) (28)**	**IL-12 (p70) (75)**
	Description	**Obs Conc (pg/mL)**	**Obs Conc (pg/mL)**	**Obs Conc (pg/mL)**	**Obs Conc (pg/mL)**	**Obs Conc (pg/mL)**	**Obs Conc (pg/mL)**	**Obs Conc (pg/mL)**	**Obs Conc (pg/mL)**
(a)	**control**	OOR <	OOR <	19.59	OOR <	OOR <	0.39	OOR <	OOR <
	poly(I:C)	OOR <	OOR <	10.48	OOR <	2.06	0.61	77.37	OOR <
(b)	**pCEP4 24h**	3.37	57.76	4.93	19.06	OOR <	3.38	OOR <	OOR <
	L1 WT 24h	OOR <	29.48	OOR <	OOR <	5.09	3.77	OOR <	OOR <
	L1 M8 24h	OOR <	OOR <	30.91	OOR <	2.78	4.3	OOR <	OOR <
	L1 RT- 24h	OOR <	OOR <	19.59	19.06	OOR <	3.23	12.71	OOR <
(c)	**pCEP4 48h**	OOR <	OOR <	24.62	30.26	OOR <	6.72	77.37	OOR <
	L1 WT 48h	OOR <	3.59	21.83	9.44	0.96	6.16	66.42	OOR <
	L1 M8 48h	22.97	95.82	41.2	62.29	5.48	4.35	170.32	10.52
	L1 RT- 48h	OOR <	52.17	19.59	OOR <	OOR <	4.97	OOR <	4.21

Table continues in the next page…

Set	Number	17	18	19	20	21	22	23	24
	Cytokines	IL-19 (29)	IL-20 (30)	IL-22 (18)	IL-26 (22)	IL-27 (p28) (13)	IL-28A (66)	IL-29 (33)	IL-32 (35)
	Description	Obs Conc (pg/mL)	Obs Conc (pg/mL)	Obs Conc (pg/mL)	Obs Conc (pg/mL)	Obs Conc (pg/mL)	Obs Conc (pg/mL)	Obs Conc (pg/mL)	Obs Conc (pg/mL)
(a)	control	149.29	*1.03	11.03	OOR <	5.06	12.61	OOR <	OOR <
	poly(I:C)	*20.14	*0.70	*1.75	OOR <	15.17	102.4	OOR <	7.79
(b)	pCEP4 24h	46.72	OOR <	OOR <	OOR <	OOR <	*7.56	OOR <	OOR <
	L1 WT 24h	OOR <	*0.05	OOR <	OOR <	5.06	14.23	OOR <	OOR <
	L1 M8 24h	*7.28	OOR <	OOR <	OOR <	OOR <	*4.83	OOR <	OOR <
	L1 RT- 24h	30.07	OOR <	*3.88	597.8	OOR <	*5.76	OOR <	OOR <
(c)	pCEP4 48h	OOR <	OOR <	4.42	206.42	0.14	OOR <	OOR <	OOR <
	L1 WT 48h	38.77	*0.05	OOR <	498.17	41.96	*1.83	OOR <	OOR <
	L1 M8 48h	149.29	*6.17	14.94	1518.49	51.72	34.18	163.87	41.98
	L1 RT- 48h	OOR <	*0.70	OOR <	333.68	28.08	15.81	OOR <	7.79

Set	Number	25	26	27	28	29	30	31	32
	Cytokines	IL-34 (15)	IL-35 (34)	LIGHT (51)	MMP-1 (43)	MMP-2 (26)	MMP-3 (45)	Osteocalcin (65)	Osteopontin (77)
	Description	Obs Conc (pg/mL)	Obs Conc (pg/mL)	Obs Conc (pg/mL)	Obs Conc (pg/mL)	Obs Conc (pg/mL)	Obs Conc (pg/mL)	Obs Conc (pg/mL)	Obs Conc (pg/mL)
(a)	control	*2.61	26.68	OOR <	198.94	OOR <	OOR <	OOR <	190.37
	poly(I:C)	OOR <	32.83	OOR <	255.09	1391.92	1846.18	55.21	146.31
(b)	pCEP4 24h	OOR <	22.7	OOR <	168	OOR <	OOR <	8.79	OOR <
	L1 WT 24h	OOR <	24.67	OOR <	134.17	OOR <	2477.99	36.84	OOR <
	L1 M8 24h	OOR <	26.68	OOR <	227.81	OOR <	OOR <	OOR <	184.64
	L1 RT- 24h	OOR <	24.67	OOR <	95.8	OOR <	OOR <	OOR <	OOR <
(c)	pCEP4 48h	OOR <	34.93	OOR <	168	OOR <	OOR <	OOR <	OOR <
	L1 WT 48h	OOR <	11.55	OOR <	168	OOR <	2288.92	OOR <	OOR <
	L1 M8 48h	80.22	39.2	3.91	557.7	8935.8	3111.52	223.87	226.89
	L1 RT- 48h	OOR <	13.31	OOR <	134.17	OOR <	1846.18	OOR <	OOR <

Table continues in the next page…

Set	Number	33	34	35	36	37
	Cytokines	Pentraxin-3 (48)	TNF-R1 (73)	TNF-R2 (67)	TSLP (52)	TWEAK (62)
	Description	Obs Conc (pg/mL)	Obs Conc (pg/mL)	Obs Conc (pg/mL)	Obs Conc (pg/mL)	Obs Conc (pg/mL)
(a)	control	2620.15	*6.33	16.18	OOR <	OOR <
	poly(I:C)	1949.55	19.84	18.21	1.17	2.74
(b)	pCEP4 24h	12996.9	26.11	*8.71	OOR <	12.43
	L1 WT 24h	13865.8	30.59	16.48	OOR <	10.18
	L1 M8 24h	14285.9	34.92	*12.48	OOR <	9.43
	L1 RT- 24h	11133.8	43.91	*7.96	OOR <	7.18
(c)	pCEP4 48h	12321.1	100.09	*11.49	OOR <	7.55
	L1 WT 48h	12972.7	86.5	13.44	OOR <	10.18
	L1 M8 48h	12166.7	113.28	22.32	13.98	19.25
	L1 RT- 48h	11836.1	65.09	13.44	6.87	10.18

Table 4: *Bio-Plex analysis of cytokines and chemokines secreted in poly(I:C) or L1-transfected HEK293T cells*. A dsRNA synthetic analog poly(I:C) was used as a positive control to compare the upregulation of cytokines or chemokines secreted in L1-transfected cells. For control and poly(I:C), the media was collected 24h post-transfection. For the rest, the extracellular media was changed with fresh DMEM at Day 3 post-transfection; the extracellular media was collected at 24 and 48 hours post-media change, respectively. Upregulation of all three sets, i.e., (a), (b), and (c) are highlighted in blue, upregulation of two of them are highlighted in green; control, pCEP4 24h, and pCEP4 48h are highlighted in gray; values with asterisk (*, see legends below) are not considered in the analysis due to low confidence. **Legends**: control, transfection with only the reagents control. poly(I:C), dsRNA immunostimulant poly(I:C)-transfected cells; pCEP4, pCEP4-transfected cells as a mock control; L1 WT, pJM101/L1.3FLAG-transfected cells; L1 M8, pALAF008_M8-transfected cells; L1 RT-, pJM105/L1.3-transfected cells; 24h, 24 hours post Day-3 post-transfection media change; 48h, 48 hours post Day-3 post-transfection media change; *, value is still in the detection range, but below than the last point on the standard curve; OOR<, value is out of range of the standard curve; Obs Conc, Observed concentration.

Gene Symbol	Description
Adar	adenosine deaminase, RNA-specific [Source:MGI Symbol;Acc:MGI:1889575]
ADAR	adenosine deaminase, RNA-specific [Source:HGNC Symbol;Acc:225]
Adarb1	adenosine deaminase, RNA-specific, B1 [Source:MGI Symbol;Acc:MGI:891999]
AFF1	AF4/FMR2 family, member 1 [Source:HGNC Symbol;Acc:7135]
AHNAK	AHNAK nucleoprotein [Source:HGNC Symbol;Acc:347]
APOBEC3B	apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3B [Source:HGNC Symbol;Acc:17352]
ARHGEF2	Rho/Rac guanine nucleotide exchange factor (GEF) 2 [Source:HGNC Symbol;Acc:682]
BAZ1A	bromodomain adjacent to zinc finger domain, 1A [Source:HGNC Symbol;Acc:960]
CENPV	centromere protein V [Source:HGNC Symbol;Acc:29920]
CKAP2L	cytoskeleton associated protein 2-like [Source:HGNC Symbol;Acc:26877]
CMSS1	cms1 ribosomal small subunit homolog (yeast) [Source:HGNC Symbol;Acc:28666]
Ddx60	DEAD (Asp-Glu-Ala-Asp) box polypeptide 60 [Source:MGI Symbol;Acc:MGI:2384570]
DDX60	DEAD (Asp-Glu-Ala-Asp) box polypeptide 60 [Source:HGNC Symbol;Acc:25942]
DDX60L	DEAD (Asp-Glu-Ala-Asp) box polypeptide 60-like [Source:HGNC Symbol;Acc:26429]
Dhx58	DEXH (Asp-Glu-X-His) box polypeptide 58 [Source:MGI Symbol;Acc:MGI:1931560]
DHX58	DEXH (Asp-Glu-X-His) box polypeptide 58 [Source:HGNC Symbol;Acc:29517]
DNA2	DNA replication helicase/nuclease 2 [Source:HGNC Symbol;Acc:2939]
Eif2ak2	eukaryotic translation initiation factor 2-alpha kinase 2 [Source:MGI Symbol;Acc:MGI:1353449]
EIF2AK2	eukaryotic translation initiation factor 2-alpha kinase 2 [Source:HGNC Symbol;Acc:9437]
EIF3D	eukaryotic translation initiation factor 3, subunit D [Source:HGNC Symbol;Acc:3278]
Elavl1	ELAV (embryonic lethal, abnormal vision)-like 1 (Hu antigen R) [Source:MGI Symbol;Acc:MGI:1100851]
Elavl2	ELAV (embryonic lethal, abnormal vision, Drosophila)-like 2 (Hu antigen B) [Source:MGI Symbol;Acc:MGI:1100887]
ELL2	elongation factor, RNA polymerase II, 2 [Source:HGNC Symbol;Acc:17064]
Fgl2	fibrinogen-like protein 2 [Source:MGI Symbol;Acc:MGI:103266]
FGL2	fibrinogen-like 2 [Source:HGNC Symbol;Acc:3696]
Fmnl2	formin-like 2 [Source:MGI Symbol;Acc:MGI:1918659]
FSCN1	fascin homolog 1, actin-bundling protein (Strongylocentrotus purpuratus) [Source:HGNC Symbol;Acc:11148]
FYN	FYN oncogene related to SRC, FGR, YES [Source:HGNC Symbol;Acc:4037]
Gna12	guanine nucleotide binding protein, alpha 12 [Source:MGI Symbol;Acc:MGI:95767]
Gpatch4	G patch domain containing 4 [Source:MGI Symbol;Acc:MGI:1913864]
Helz2	helicase with zinc finger 2, transcriptional coactivator [Source:MGI Symbol;Acc:MGI:2385169]
HELZ2	helicase with zinc finger 2, transcriptional coactivator [Source:HGNC Symbol;Acc:30021]
HERC5	HECT and RLD domain containing E3 ubiquitin protein ligase 5 [Source:HGNC Symbol;Acc:24368]
HNRNPM	heterogeneous nuclear ribonucleoprotein M [Source:HGNC Symbol;Acc:5046]

Gene Symbol	Description
HPSE	heparanase [Source:HGNC Symbol;Acc:5164]
IFI16	interferon, gamma-inducible protein 16 [Source:HGNC Symbol;Acc:5395]
Ifit1	interferon-induced protein with tetratricopeptide repeats 1 [Source:MGI Symbol;Acc:MGI:99450]
IFIT1	interferon-induced protein with tetratricopeptide repeats 1 [Source:HGNC Symbol;Acc:5407]
Ifit2	interferon-induced protein with tetratricopeptide repeats 2 [Source:MGI Symbol;Acc:MGI:99449]
IFIT2	interferon-induced protein with tetratricopeptide repeats 2 [Source:HGNC Symbol;Acc:5409]
Ifit3	interferon-induced protein with tetratricopeptide repeats 3 [Source:MGI Symbol;Acc:MGI:1101055]
IFIT3	interferon-induced protein with tetratricopeptide repeats 3 [Source:HGNC Symbol;Acc:5411]
Igf2bp2	insulin-like growth factor 2 mRNA binding protein 2 [Source:MGI Symbol;Acc:MGI:1890358]
IGF2BP3	insulin-like growth factor 2 mRNA binding protein 3 [Source:HGNC Symbol;Acc:28868]
JUP	junction plakoglobin [Source:HGNC Symbol;Acc:6207]
Larp1	La ribonucleoprotein domain family, member 1 [Source:MGI Symbol;Acc:MGI:1890165]
Lgals3bp	lectin, galactoside-binding, soluble, 3 binding protein [Source:MGI Symbol;Acc:MGI:99554]
LGALS3BP	lectin, galactoside-binding, soluble, 3 binding protein [Source:HGNC Symbol;Acc:6564]
Ltf	lactotransferrin [Source:MGI Symbol;Acc:MGI:96837]
MKI67	marker of proliferation Ki-67 [Source:HGNC Symbol;Acc:7107]
Mki67	antigen identified by monoclonal antibody Ki 67 [Source:MGI Symbol;Acc:MGI:106035]
Mov10	Moloney leukemia virus 10 [Source:MGI Symbol;Acc:MGI:97054]
MOV10	Mov10, Moloney leukemia virus 10, homolog (mouse) [Source:HGNC Symbol;Acc:7200]
Myh10	myosin, heavy polypeptide 10, non-muscle [Source:MGI Symbol;Acc:MGI:1930780]
NPM1	nucleophosmin (nucleolar phosphoprotein B23, numatrin) [Source:HGNC Symbol;Acc:7910]
OASL	2'-5'-oligoadenylate synthetase-like [Source:HGNC Symbol;Acc:8090]
Parp12	poly (ADP-ribose) polymerase family, member 12 [Source:MGI Symbol;Acc:MGI:2143990]
PARP12	poly (ADP-ribose) polymerase family, member 12 [Source:HGNC Symbol;Acc:21919]
PATL1	protein associated with topoisomerase II homolog 1 (yeast) [Source:HGNC Symbol;Acc:26721]
PEG10	paternally expressed 10 [Source:HGNC Symbol;Acc:14005]
PLOD2	procollagen-lysine, 2-oxoglutarate 5-dioxygenase 2 [Source:HGNC Symbol;Acc:9082]
RBBP6	retinoblastoma binding protein 6 [Source:HGNC Symbol;Acc:9889]
RP11-395B7.7	(does not map to a known gene)
RP3-437C15.1	(does not map to a known gene)
RRP8	ribosomal RNA processing 8, methyltransferase, homolog (yeast) [Source:HGNC Symbol;Acc:29030]
RTCB	RNA 2',3'-cyclic phosphate and 5'-OH ligase [Source:HGNC Symbol;Acc:26935]
S100P	S100 calcium binding protein P [Source:HGNC Symbol;Acc:10504]

Gene Symbol	Description
SART3	squamous cell carcinoma antigen recognized by T cells 3 [Source:HGNC Symbol;Acc:16860]
SLC3A2	solute carrier family 3 (amino acid transporter heavy chain), member 2 [Source:HGNC Symbol;Acc:11026]
SMARCA1	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 1 [Source:HGNC Symbol;Acc:11097]
SSB	Sjogren syndrome antigen B (autoantigen La) [Source:HGNC Symbol;Acc:11316]
Ssr1	signal sequence receptor, alpha [Source:MGI Symbol;Acc:MGI:105082]
TFAP2B	transcription factor AP-2 beta (activating enhancer binding protein 2 beta) [Source:HGNC Symbol;Acc:11743]
Tfrc	transferrin receptor [Source:MGI Symbol;Acc:MGI:98822]
TFRC	transferrin receptor [Source:HGNC Symbol;Acc:11763]
TOP2A	topoisomerase (DNA) II alpha 170kDa [Source:HGNC Symbol;Acc:11989]
Trim25	tripartite motif-containing 25 [Source:MGI Symbol;Acc:MGI:102749]
TRIM25	tripartite motif containing 25 [Source:HGNC Symbol;Acc:12932]
TRIM56	tripartite motif containing 56 [Source:HGNC Symbol;Acc:19028]
WTAP	Wilms tumor 1 associated protein [Source:HGNC Symbol;Acc:16846]
XRN1	5'-3' exoribonuclease 1 [Source:HGNC Symbol;Acc:30654]
Xrn2	5'-3' exoribonuclease 2 [Source:MGI Symbol;Acc:MGI:894687]
Zc3hav1	zinc finger CCCH type, antiviral 1 [Source:MGI Symbol;Acc:MGI:1926031]
ZC3HAV1	zinc finger CCCH-type, antiviral 1 [Source:HGNC Symbol;Acc:23721]
Zfp36l2	zinc finger protein 36, C3H type-like 2 [Source:MGI Symbol;Acc:MGI:107945]
ZFP36L2	ZFP36 ring finger protein-like 2 [Source:HGNC Symbol;Acc:1108]

Table 5: *ISG proteins from the interferome database that preferentially associate with WT ORF1p complexes*. Protein hits with >0.5 of $\log_2$(abundance ratio WT *vs.* M8 [RBM]) were submitted to the interferome database to search for ISGs with a cutoff of 5-fold increase/decrease by type I, II, and III IFN. Left column, gene symbols. Right column, gene descriptions. HGNC, Human Genome Nomenclature Committee. MGI, Mouse Genome Informatics. Acc, accession number.

Chapter 4

ISG protein HELZ2 potently inhibits L1 retrotransposition through L1 RNA 5′UTR recognition

4.1 Introduction

HELZ2 was discovered as a transcriptional co-activator of transcriptional factors including peroxisome proliferator-activated receptor (PPAR)-α (Surapureddi et al., 2002) and PPAR-γ (Tomaru et al., 2006), and was shown to augment both PPAR-α- and PPAR-γ-mediated gene expressions, possibly through recruitment of other coactivators (Tomaru et al., 2006; Katano-Toki et al., 2013). PPAR-α and PPAR-γ are essential during adipocyte differentiation. In this context, HELZ2 was suggested to increase PPAR-γ-mediated gene expressions through stable association with another transcriptional factor, TR-associated protein 3 (Katano-Toki et al., 2013). However, HELZ2 domains and their mechanism of action in detail is yet to be elucidated.

Based on the conserved sequence alignment, HELZ2 was predicted to have two helicase domains that flank an exoribonuclease RNase II/R (RNB) domain (Tomaru et al., 2006; Katano-Toki et al., 2013); however, the role of these domains in inhibiting L1 retrotransposition is unclear. In this chapter, I studied the helicase and exoribonuclease activities of HELZ2 in more detail and described a distinct function of HELZ2 to destabilize L1 RNA.

4.2 Results

4.2a Endogenous HELZ2 knockdown enhances L1 retrotransposition

Firstly, small interfering RNAs (siRNAs) were used to examine whether endogenous HELZ2 also inhibit L1 retrotransposition in HeLa-JVM cells; siRNAs against HELZ2 or MOV10 RNAs revealed an ~70% or ~80% knockdown of the target RNA levels, respectively, when compared to a non-targeting siRNA control (**Fig. 4.1a**). Using *mEGFPI*-based assays, ~1.5-fold or ~3-fold increase in L1 retrotransposition efficiency was observed in the siHELZ2 or siMOV10 treated cells, respectively (**Fig. 4.1b**). The results suggest that endogenous HELZ2 may also inhibit L1 retrotransposition.

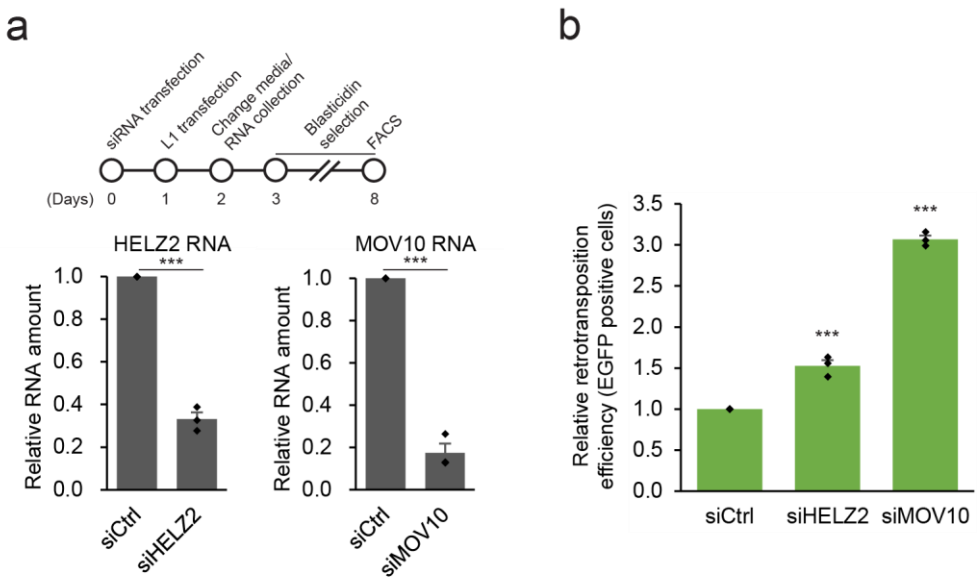


Fig. 4.1: *Small-interfering RNA (siRNA)-mediated knockdown of HELZ2 increased L1 retrotransposition efficiency* **(a)** *The efficiency of siRNA-mediated knockdown of HELZ2 and MOV10 RNAs*. Top, timeline of the assay conducted in HeLa-JVM cells. HeLa-JVM cells were transfected with a non-targeting siRNA control (siCtrl), siRNA targeting HELZ2 (siHELZ2), or siRNA targeting MOV10 (siMOV10). Left, HELZ2 RNA levels in siRNA treated cells. Right, MOV10 RNA levels in siRNA treated cells. X-axes, name of the siRNA. HELZ2 and MOV10 RNA levels were determined using RT-qPCR (primer sets: HELZ2 and MOV10, respectively) and then were normalized to *ACTB* RNA levels (primer set: Beta-actin). Y-axes, relative HELZ2 or MOV10 RNA levels normalized to the siCtrl. A two-tailed, unpaired Student's t-test was used to calculate the *p*-values relative to the siRNA control: p = 3.1 x 10^{-5}*** (siHELZ2); and 5.2 x 10^{-5}*** (siMOV10). **(b)** *HELZ2 and MOV10 siRNA-mediated knockdown increases L1 retrotransposition efficiency.* HeLa-JVM cells were transfected with either siCtrl, siHELZ2, or siMOV10, followed by transfection with either cepB-gfp-L1.3 or cepB-gfp-L1.3RT(-) intronless, which was used to normalize transfection efficiencies. X-axis, name of the siRNA. Y-axis, relative retrotransposition efficiency. Pairwise comparisons relative to the non-targeting siRNA control: p = 2.9 x 10^{-4}*** (siHELZ2); and 2.0 x 10^{-7}*** (siMOV10). All the reported values represent the mean ± SEM from three independent biological replicates. The *p*-values, except for the RT-qPCR experiment shown in panel (a), were calculated using a one-way ANOVA followed by a Bonferroni-Holm post-hoc tests. ns: not significant; *** p<0.001.

4.2b The HELZ2 helicase activity is essential for L1 retrotransposition inhibition

HELZ2 contains a putative exoribonuclease RNase II/R (RNB) domain that is flanked by two putative helicase domains (helicase 1 and helicase 2) (**Fig. 4.2**). A RNB domain often possesses a 3′ to 5′ single-strand exoribonuclease activity (Amblar et al., 2006; Frazão et al., 2006).

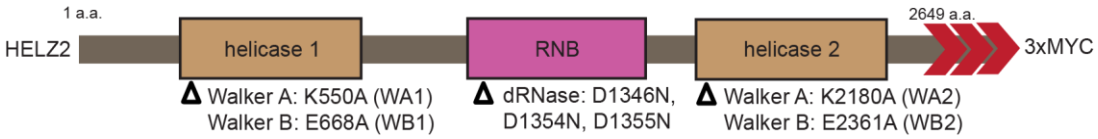


Fig. 4.2: *Schematic of the HELZ2 protein domains.* HELZ2 contains a central putative RNB exonuclease domain with two flanking putative helicase domains (helicase 1 and helicase 2). Open triangles, relative positions of missense mutations made in conserved amino acids within the domains: Walker A (WA) boxes in the helicase 1 and helicase 2 domains (K550A [WA1] and K2180A [WA2], Walker B (WB) boxes in the helicase 1 and helicase 2 domains (E668A [WB1] and E2361A [WB2], respectively). The RNB mutant contains three missense mutations, D1346N/D1354N/D1355N (dRNase). Red arrowheads, relative positions of the 3xMYC carboxyl-terminal epitope tag in the HELZ2 expression constructs. a.a. indicates the amino acid length of HELZ2.

To identify evolutionarily conserved amino acid residues in the RNB domain, protein sequences of RNB-containing proteins from human, yeast, and *E. coli* were aligned. Three evolutionarily conserved aspartic acid residues, which when mutated, are predicted to impair the exoribonuclease activity were found (Amblar et al., 2006; Frazão et al., 2006; Barbas et al., 2008) (**Fig. 4.3a**), which led to the creation of a HELZ2 triple mutant (D1346N/D1354N/D1355N, a.k.a. dRNase mutant).

WT HELZ2-3xFLAG and dRNase HELZ2-3xFLAG mutant proteins from HEK293T cells were purified for an *in vitro* ribonuclease assay to examine the exoribonuclease activity (**Fig. 4.3b**). The ribonuclease assay was performed using a poly$(A)_{30}$ RNA oligonucleotide labeled with IRDye800 at its 5′ end as an RNA substrate. The WT HELZ2-3xFLAG protein, but not the dRNase HELZ2-3xFLAG mutant protein, degraded the single-strand RNA substrate in a

3′ to 5′ direction, suggesting that the HELZ2 RNB domain contains a functional ribonuclease activity and that the dRNase mutant loses this activity (**Fig. 4.3c**). However, the dRNase mutant showed minimal change (i.e., less than 2-fold) on L1 retrotransposition efficiency in HeLa-JVM and HEK293T cells when compared to the WT HELZ2 control (**Figs. 4.3d and 4.3e, see discussion in chapter 5**).

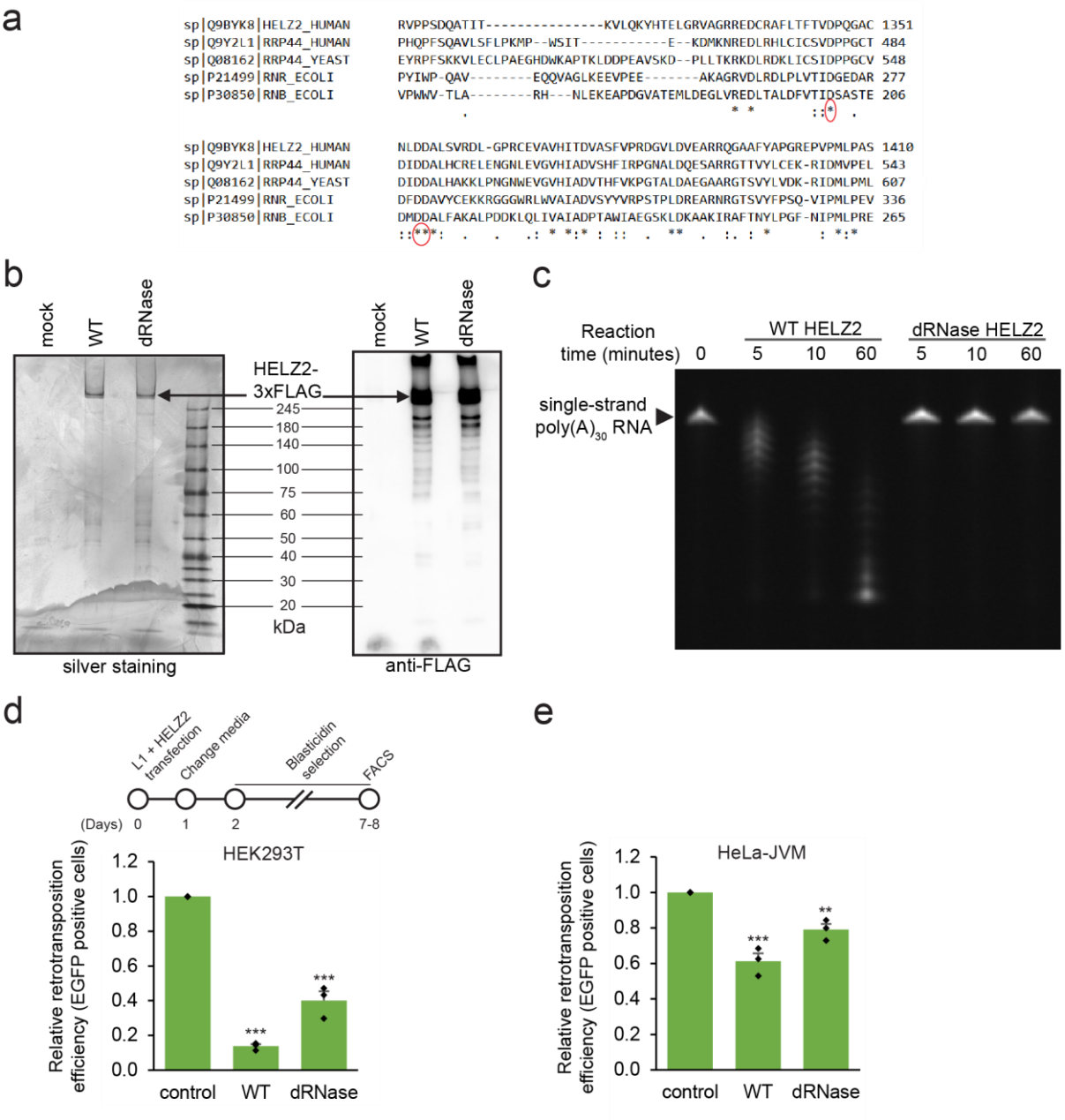


Fig. 4.3: *Functional analyses of the HELZ2 RNB domain using a dRNase mutant*. **(a)** *Identification of conserved amino acids in the RNB domain.* RNB-containing proteins were aligned: *Homo sapiens*

exosome complex exonuclease Rrp44 (RRP44_HUMAN) and HELZ2 (HELZ2_HUMAN); *Saccharomyces cerevisiae* exosome complex exonuclease Rrp44 (RRP44_YEAST); and *Escherichia coli* RNase R (RNR_ECOLI) and Exoribonuclease 2 (RNB_ECOLI). Red circles, amino acids mutated in the D1346N/D1354N/D1355N (dRNase) triple mutant. **(b)** *Protein purification of recombinant WT and mutant dRNase HELZ2*. The hHELZ2_3xFLAG (pALAF071) or hHELZ2_3xFLAG_dRNase (pALAF073) plasmid construct was independently transfected in HEK293T cells and the corresponding proteins were purified using an anti-FLAG antibody, subjected to SDS-PAGE, and visualized by silver staining (left) and western blot using an anti-FLAG antibody (right). The pCEP4-transfected cells served as a negative control (mock). The arrow indicates the position of the HELZ2-3xFLAG protein. **(c)** *HELZ2 has a 3' to 5' exoribonuclease activity*. The WT HELZ2-3xFLAG and dRNase HELZ2-3xFLAG proteins in panel (b) were incubated with a single-strand poly$(A)_{30}$ RNA oligonucleotide labeled with IRDye800 at its 5' end (poly[rA_{30}]) for 0, 5, 10, and 60 minutes at 37˚C. Reactions containing only single-strand poly$(A)_{30}$ RNA without the recombinant protein served as a negative control (leftmost lane). The single-stranded (ss) RNAs were separated on a polyacrylamide/urea gel with 1x TBE buffer. **(d)** *HELZ2 dRNase mutant effect on L1 retrotransposition efficiency in HEK293T cells*. Top: the timeline for the retrotransposition assays shown in panels (d) and (e). HEK293T cells were co-transfected with cepB-gfp-L1.3 (*mEGFPI*) and either pCMV-3Tag-8-Barr (control), pALAF015 (WT), or pALAF030 (dRNase). The retrotransposition efficiency was normalized to the transfection efficiency control (i.e., cells co-transfected with cepB-gfp-L1.3RT(-) intronless and either pCMV-3Tag-8-Barr (control), pALAF015 (WT), or pALAF030 (dRNase)). X-axis, name of the plasmid co-transfected with cepB-gfp-L1.3 (*mEGFPI*). Y-axis, relative retrotransposition efficiency relative to the cepB-gfp-L1.3 (*mEGFPI*) + pCMV-3Tag-8-Barr control. Pairwise comparisons relative to the cepB-gfp-L1.3 (*mEGFPI*) + pCMV-3Tag-8-Barr control: $p = 9.4 \times 10^{-10}$*** (WT HELZ2), 4.1×10^{-8}*** (dRNase). **(e)** *HELZ2 dRNase mutant effect on L1 retrotransposition efficiency in HeLa-JVM cells*. Experiments were conducted as summarized in panel (d). Pairwise comparisons relative to the control: $p = 9.5 \times 10^{-5}$*** (WT); 0.0073** (dRNase). Values represent the mean ± SEM from three independent biological replicates. The *p*-values were calculated using a one-way ANOVA followed by a Bonferroni-Holm post-hoc tests. ** $p<0.01$; *** $p<0.001$.

As the mutation in the RNB domain showed minimal effect, the putative HELZ2 helicase

domains effect on L1 retrotransposition was examined. Conserved amino acids in the Walker A and Walker B boxes, which will be indispensable for ATP binding (WA1 [K550A] in the helicase 1 domain and WA2 [K2180A] in the helicase 2 domain) (**Fig. 4.2**) or ATP hydrolysis (WB1 [E668A] in the helicase 1 domain and WB2 [E2361A] in the helicase 2 domain), were mutated, respectively (Walker et al., 1982; Finn et al., 2014; Miller and Enemark, 2016) (**Fig. 4.2**). The WA1 mutant still inhibits L1 retrotransposition almost as effectively as WT HELZ2 in HEK293T (**Fig. 4.4a**), but not in HeLa-JVM (**Fig. 4.4b**) cells. The WA2 mutants were significantly impaired in their ability to inhibit L1 retrotransposition in both HEK293T (**Fig. 4.4a**) and HeLa-JVM cells (**Fig. 4.4b**). A similar data trend was observed for the Walker B box mutations in HEK293T and HeLa-JVM cells (**Figs. 4.4c and 4.4d**). The WA1&2 double helicase mutant showed a complete loss in the ability to inhibit L1 retrotransposition (**Figs. 4.4a and 4.4b**). In sum, mutations in the helicase 2 (WA2 and WB2 mutants) domains generally showed a more severe impairment in the ability to inhibit L1 retrotransposition in comparison to mutations in the helicase 1 (WA1 and WB1 mutants), indicating the importance of the helicase 2 domain in L1 retrotransposition inhibition.

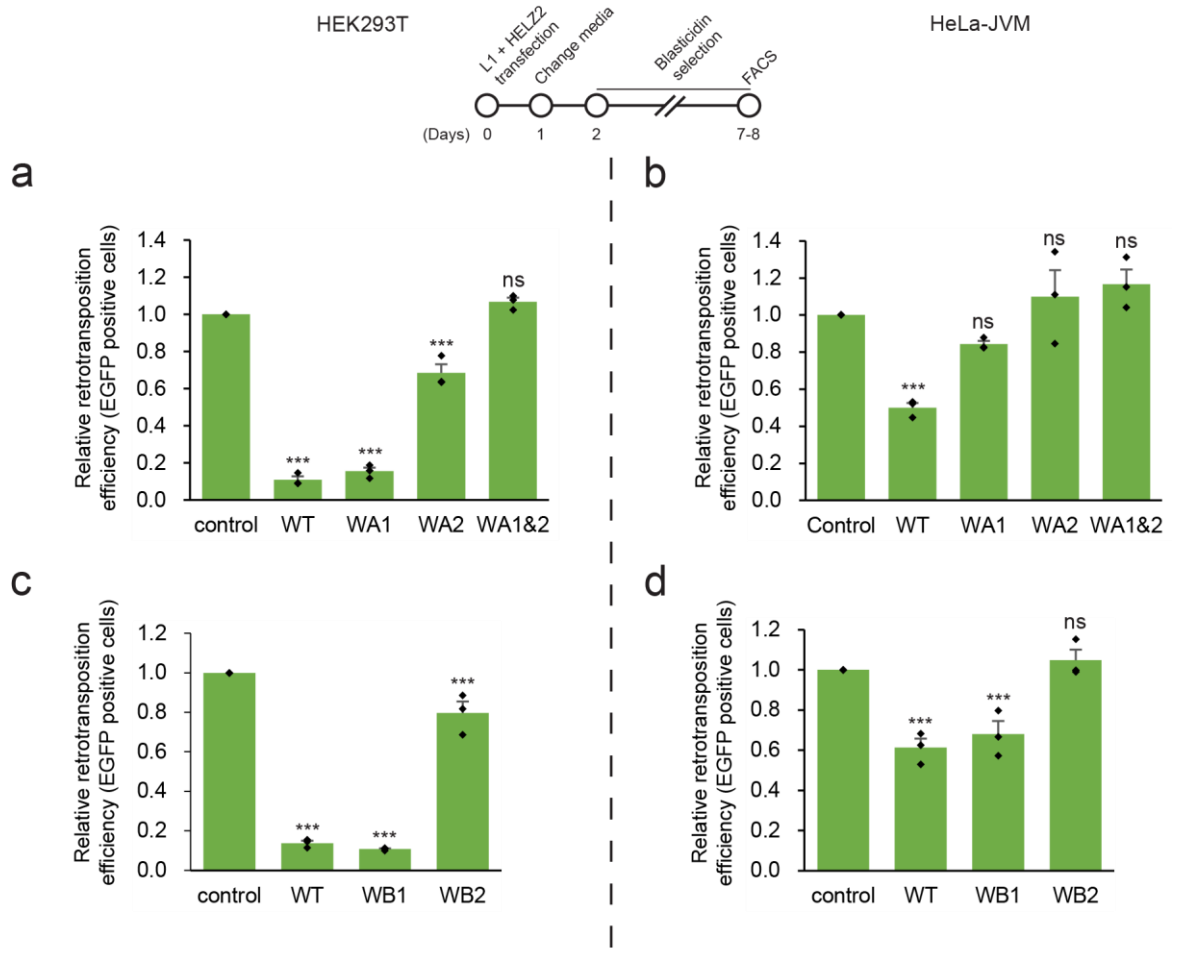


Fig. 4.4: *The effect of conserved sites mutation in HELZ2 helicase domains on L1 retrotransposition efficiency in HEK293T (left) or HeLa-JVM (right)*. Top, timeline of the assay for experiments shown in panels (a, b, c, and d). **(a and b)** *The effect of mutations in the Walker A box on L1 retrotransposition in HEK293T (a) or HeLa-JVM (b) cells.* The cells were co-transfected with cepB-gfp-L1.3, which contains an *mEGFPI* retrotransposition indicator cassette, and either pCMV-3Tag-8-Barr (control), pALAF015 (WT HELZ2), or one of the following HELZ2 expression plasmids that contain a mutation(s) in the Walker A box: pALAF025 (WA1); pALAF026 (WA2); or pALAF027 (WA1&2). Cells co-transfected with cepB-gfp-L1.3RT(-) intronless and either pCMV-3Tag-8-Barr, pALAF015 (WT HELZ2), or a mutant HELZ2 plasmid served as transfection normalization and toxicity controls. X-axes, name of HELZ2 expression constructs co-transfected into cells with cepB-gfp-L1.3; control, pCMV-3Tag-8-Barr. Y-axes, relative retrotransposition efficiency normalized to the cepB-gfp-L1.3 + pCMV-3Tag-8-Barr control. Pairwise comparisons in (a) relative to the cepB-gfp-L1.3 (*mEGFPI*) + pCMV-3Tag-8-Barr control: p = 2.5 x 10^{-11}*** (WT HELZ2); 3.5 x 10^{-11}*** (WA1); 1.7 x 10^{-6}*** (WA2); and 0.070^{ns} (WA1&2). Pairwise comparisons in (b) relative to the control: p = 0.00087*** (WT HELZ2);

0.26^{ns} (WA1); 0.32^{ns} (WA2); and 0.32^{ns} (WA1&2). **(c and d)** *The effects of mutations in the Walker B box on L1 retrotransposition in HEK293T (c) or HeLa-JVM (d) cells*. The cells were co-transfected with cepB-gfp-L1.3, which contains an *mEGFPI* retrotransposition indicator cassette, and either pCMV-3Tag-8-Barr (control), pALAF015 (WT HELZ2), or one of the following HELZ2 expression plasmids that contain a mutation(s) in the Walker B box (i.e., pALAF028 [WB1] or pALAF029 [WB2]). Cells co-transfected with cepB-gfp-L1.3RT(-) intronless and either pCMV-3Tag-8-Barr, pALAF015 (WT HELZ2), or a mutant HELZ2 plasmid served as transfection, normalization, and toxicity controls. Retrotransposition efficiencies were calculated as described in panel (a). Pairwise comparisons in (c) relative to the cepB-gfp-L1.3 (mEGFPI) + pCMV-3Tag-8-Barr control: $p = 9.4 \times 10^{-10}$*** (WT); 8.4×10^{-10}*** (WB1); and 8.7×10^{-4}*** (WB2). Pairwise comparisons in (d) relative to the L1.3 + pCMV-3Tag-8-Barr control: $p = 9.5 \times 10^{-5}$*** (WT); 0.0004*** (WB1); and 0.43^{ns} (WB2). Values represent the mean ± SEM of three independent biological replicates. The *p*-values were calculated using a one-way ANOVA followed by Bonferroni-Holm post-hoc tests. ns: not significant; *** $p<0.001$.

Western blot and RT-qPCR showed that WA1 mutant reduced both ORF1p-T7 and L1 RNA levels almost as effectively as WT HELZ2 in HeLa-JVM cells, respectively (**Fig. 4.5**). The WA2 and WA1&2 mutants showed a small effect on the steady state levels of the ORF1p-T7 protein, but still partially reduced L1 RNA levels when compared to the WT control (**Fig. 4.5**). The change in HELZ2 mutants′ ability to inhibit L1 retrotransposition is unlikely due to the protein instability as I did not observe a noticeable reduction in the steady state levels of the HELZ2 mutant proteins (**Fig. 4.5**). Finally, the ability of ORF1p-FLAG to localize in the cytoplasmic foci when co-transfected with WA1&2 double helicase mutant was not significantly affected, as opposed to when co-transfected with WT HELZ2 (**Fig. 3.9**). In sum, the above data suggest that the HELZ2 RNase activity showed lesser effects on L1 retrotransposition when compared to the HELZ2 helicase activity, and that mutations in the HELZ2 helicase domains adversely affect L1 RNA stability, ORF1p levels, and ORF1p cytoplasmic localization to different extents.

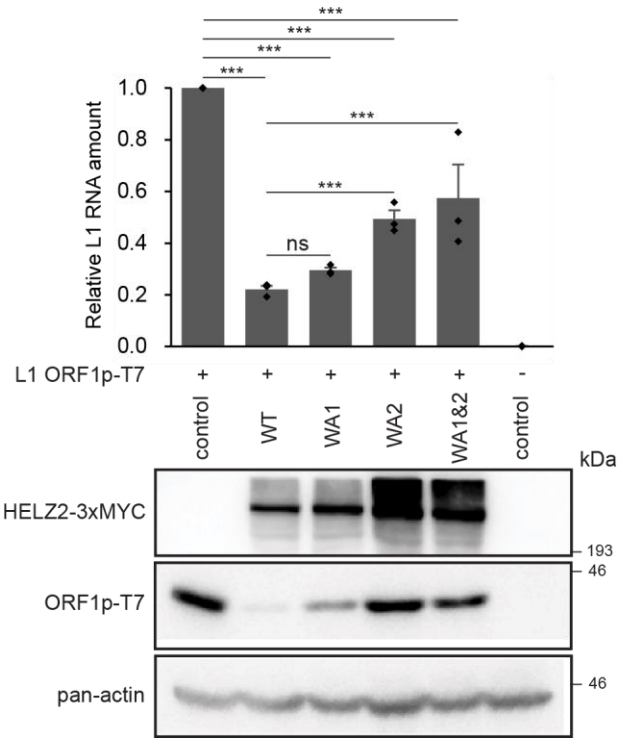


Fig. 4.5: *Mutations in the HELZ2 helicase domains reduce the ability to decrease L1 ORF1p and RNA.* HeLa-JVM cells were transfected with pTMF3 (L1 ORF1p-T7), denoted by + symbol, and either pCMV-3Tag-8-Barr (control), pALAF015 (WT HELZ2), or an individual HELZ2 expression plasmid containing a mutation(s) in the Walker A box: pALAF025 (WA1), pALAF026 (WA2), or pALAF027 (WA1&2). Top: L1 RNA levels were determined by RT-qPCR using primers directed against sequences in the transfected L1 RNA (primer set: L1 [SV40]) and then were normalized to *ACTB* RNA levels (primer set: Beta-actin). Pairwise comparisons relative to the pTMF3 (L1 ORF1p-T7) + pCMV-3Tag-8-Barr control: p = 9.5 x 10^{-9}*** (WT); 1.9 x 10^{-8}*** (WA1); 7.3 x 10^{-7}*** (WA2); and 1.5 x 10^{-6}*** (WA1&2). Pairwise comparisons relative to the pTMF3 (L1 ORF1p-T7) + WT HELZ2: p = 0.56^{ns} (WA1); 5.9 x 10^{-4}*** (WA2); 1.9 x 10^{-4}*** (WA1&2). Bottom: western blot image displaying ORF1p-T7 bands. HELZ2 expression was detected using an anti-MYC antibody. ORF1p was detected using an anti-T7 antibody. Pan-actin served as a loading control.

4.2c HELZ2 and HERC5 recognize L1 RNA independent of L1 RNP formation

The mechanism of association between ORF1p-FLAG and HELZ2 was further examined. An RNase A treatment of the WT ORF1p RNP complex abolished the ORF1p-FLAG-HELZ2 interaction, suggesting that HELZ2 association with ORF1p is RNA-dependent, similarly with the ORF1p-PABPC1 interaction (Dai et al., 2012; Moldovan and Moran, 2015) (**Fig. 4.6**).

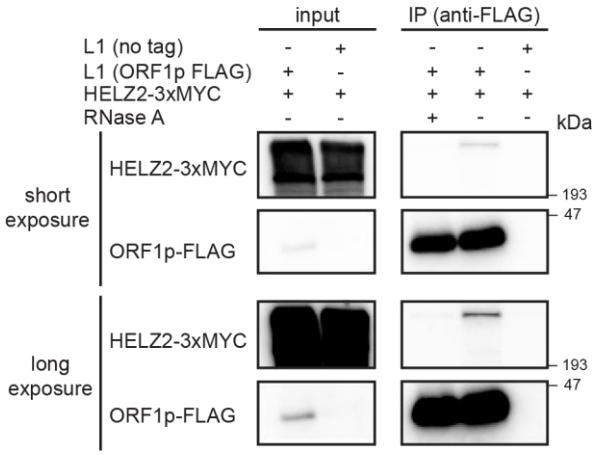


Fig. 4.6: *ORF1p and HELZ2 association is RNA-dependent*. HEK293T cells were co-transfected with pALAF015 (HELZ2-3xMYC) and either pJM101/L1.3FLAG (WT ORF1p-FLAG) or pJM101/L1.3 (no tag). The input and anti-FLAG IP fractions were analyzed by western blot using an anti-FLAG antibody to detect ORF1p-FLAG or an anti-MYC antibody to detect HELZ2-3xMYC. Shown are short (top blots) and longer (bottom blots) chemiluminescence western blot exposures.

To examine whether L1 RNP formation is necessary for the association between WT ORF1p-FLAG and HELZ2, the effects of HELZ2 overexpression on L1 RNA and ORF1p protein abundance in HeLa-JVM cells transfected with either pJM101/L1.3FLAG (ORF1p-FLAG) or pALAF008_L1.3FLAG_M8 (M8/RBM-FLAG) were compared. Both WT ORF1p-FLAG and M8/RBM ORF1p-FLAG transfected cells showed a marked reduction in L1 RNA (~80% reduction) and ORF1p levels upon HELZ2 overexpression when compared to controls (**Fig. 4.7a**); RT-qPCR was performed using a probe set that specifically recognizes the SV40 poly(A) signal of a plasmid expressing L1 RNA, while ORF1p-FLAG were detected with an anti-FLAG antibody. A similar reduction in L1 WT ORF1p-FLAG and M8/RBM ORF1p-FLAG protein levels was observed upon the co-expression of HERC5 in HeLa-JVM cells (**Fig. 4.7b**). Thus, both HELZ2 and HERC5 overexpression appear to reduce L1 RNA and ORF1p, respectively, independent of L1 RNP formation.

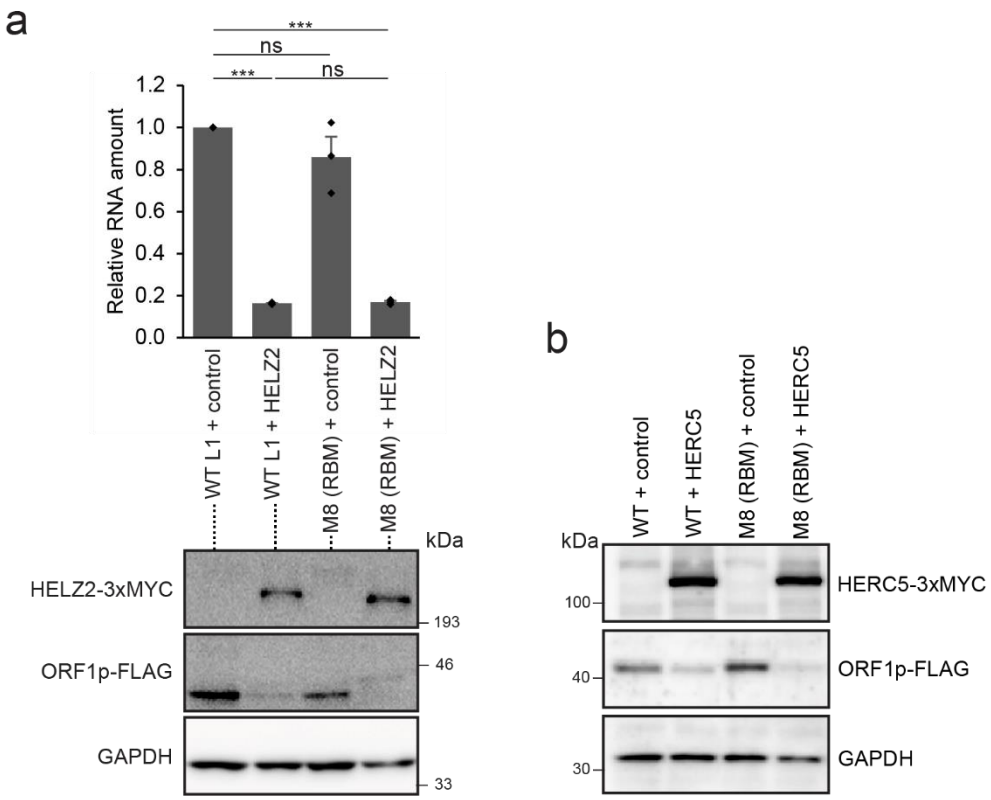


Fig. 4.7: *HERC5 and HELZ2 reduced L1 ORF1p and RNA levels independent of ORF1p RNA-binding ability*. **(a)** *Reduction of steady state levels of L1 RNA and ORF1p by HELZ2 expression is independent of ORF1p RNA-binding*. HeLa-JVM cells were co-transfected with pJM101/L1.3FLAG (WT ORF1p-FLAG) or the pALAF008 ORF1p-FLAG (M8 [RBM]) mutant expression plasmid and either pCEP4 (control) or pALAF015 (HELZ2). Top: L1 RNA amounts were determined by RT-qPCR (primer set: L1 [SV40]) and were normalized to *ACTB* RNA levels (primer set: Beta-actin). The L1 RNA values were normalized to the WT L1 or ORF1p-FLAG (M8 [RBM]) + pCEP4 control transfections. Pairwise comparisons (in parentheses) relative to the (WT L1 + control) are shown: p = 7.1×10^{-7}*** (WT L1 + HELZ2); 0.090^{ns} (M8 [RBM] + control); 6.7×10^{-7}*** (M8 [RBM] + HELZ2). Pairwise comparisons of (WT L1 + HELZ2) *vs*. (M8 [RBM] + HELZ2), $p = 0.92^{ns}$. Values represent the mean ± SEM from three independent biological replicates. The *p*-values were calculated using a one-way ANOVA followed by a Bonferroni-Holm post-hoc tests. ns: not significant; *** $p<0.001$. Bottom: ORF1p-FLAG and HELZ2 protein levels were detected by western blot using anti-MYC and anti-FLAG antibodies, respectively. GAPDH served as a loading control. **(b)** *Reduction of L1 ORF1p steady state levels by HERC5 is independent of ORF1p RNA-binding*. HeLa-JVM cells were co-transfected with pJM101/L1.3FLAG (WT ORF1p-FLAG) or the pALAF008 (M8 [RBM] ORF1p-FLAG) mutant expression plasmid and either pCMV-3Tag-8-Barr (control) or pALAF023 (HERC5). ORF1p-FLAG

and HERC5 protein levels were detected by western blot using anti-MYC and anti-FLAG (Sigma Aldrich, F7425) antibodies, respectively. GAPDH served as a loading control.

4.2d HELZ2 modestly suppresses Alu retrotransposition

HeLa-HA cells (Hulme et al., 2007) were transfected with a plasmid that expresses both a monocistronic L1 ORF2p-3xFLAG expression cassette (Miyoshi et al., 2019) and an engineered Alu-element containing a *neo*-based retrotransposition indicator cassette (*neo*^{Tet}) (Dewannieux et al., 2003) to test whether HELZ2 overexpression affects Alu retrotransposition. Alu retrotransposition was reduced by ~2-fold when compared to the respective controls upon HELZ2 overexpression (**Fig. 4.8a**). HELZ2 overexpression was observed to reduce L1 ORF2p and Alu RNA levels by ~80% and ~35%, respectively; consistently, L1 ORF2p protein levels were reduced in parallel with L1 RNA reduction (**Fig. 4.8b**). Notably, the reductions in the levels of full-length and monocistronic L1 RNAs upon HELZ2 overexpression were quite similar (i.e., **Fig. 4.8b *vs.* Fig. 4.7a**), suggesting that the decrease in Alu retrotransposition may happen mainly due to the L1 RNA destabilization by HELZ2. That being stated, co-transfection in HeLa-HA cells with HELZ2 and an Alu only expression plasmid (*Alu_neo*^{Tet}) still showed a ~40% reduction in Alu RNA levels, although this reduction was not as significant as observed with L1 RNA (**Fig. 4.8b *vs.* Fig. 4.8c**).

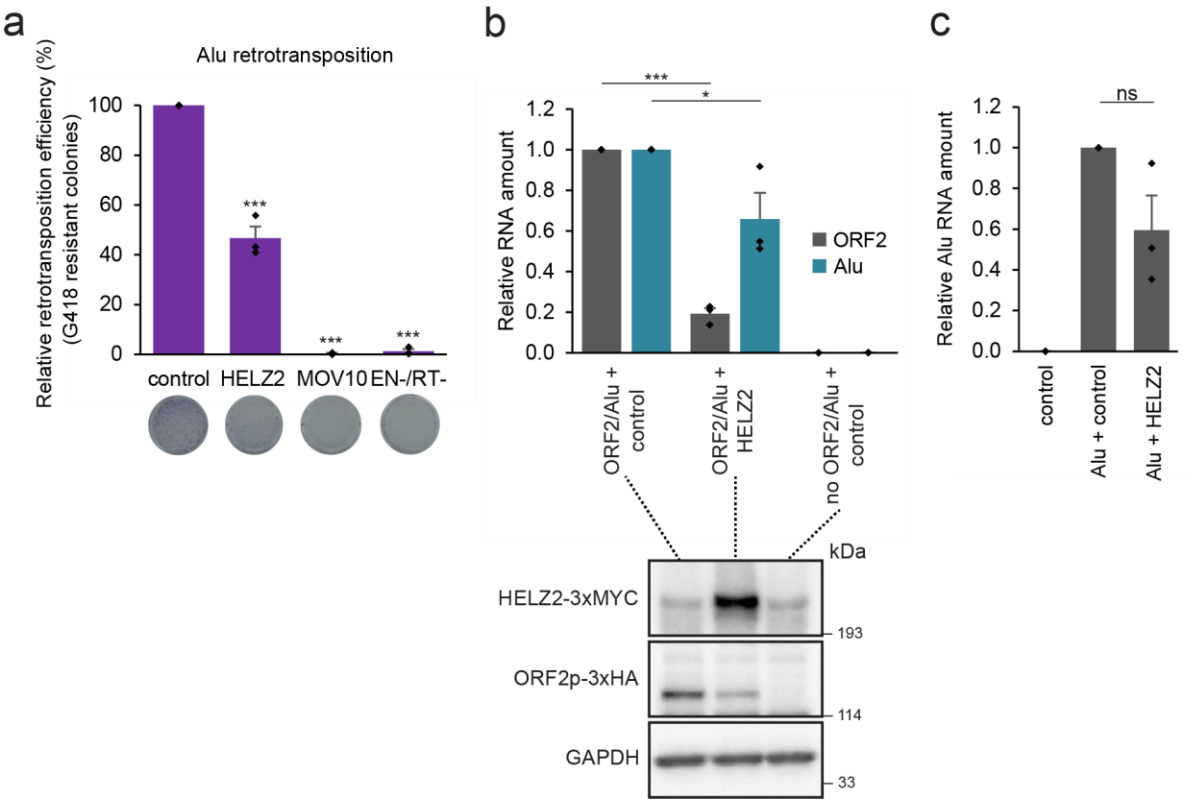


Fig. 4.8: *HELZ2 expression effects on Alu RNA steady state level and Alu retrotransposition efficiency* **(a)** *HELZ2 expression inhibits Alu retrotransposition*. HeLa-HA cells were co-transfected with pTMO2F3_Alu (which expresses an Alu element marked with *neo*-based retrotransposition indicator cassette and monocistronic version of L1 ORF2p [see Methods]), pTMO2F3D145AD702A_Alu (which expresses an Alu element marked with *neo*-based retrotransposition indicator cassette and an EN-/RT- mutant version of L1 ORF2 [see Methods]), or phrGFP-C (a transfection normalization control) and either pCMV-3Tag-8-Barr (control), pALAF015 (WT HELZ2), or pALAF024 (WT MOV10). X-axis, name of constructs. Y-axis, the percentage of G418-resistant foci, indicative of Alu retrotransposition, relative to the pTMO2F3_Alu + pCMV-3Tag-8-Barr control (see Methods for more detail). Representative images of G418-resistant foci are shown below the graph. Pairwise comparisons relative to the pTMO2F3_Alu + pCMV-3Tag-8-Barr control: p = 7.8 x 10^{-5}*** (HELZ2); 1.8 x 10^{-7}*** (MOV10); and 1.6 x 10^{-7}*** (EN-/RT-). **(b)** *HELZ2 expression leads to a reduction in monocistronic ORF2 L1 RNA and ORF2p levels.* HeLa-HA cells were co-transfected with pTMO2H3_Alu (ORF2p-3xHA and Alu) and either pCMV-3Tag-8-Barr (control) or pALAF015 (HELZ2). Top: ORF2 (gray bars) and Alu RNA (blue bars) levels were determined using RT-qPCR (primer sets: L1 [SV40] and *mneoI* [Alu or L1], respectively) and normalized to *ACTB* RNA levels (primer set: Beta-actin). X-axis, co-transfected constructs name. Y-axis, relative RNA level normalized to the pTMO2H3_Alu (ORF2p-

3xHA and Alu) + pCMV-3Tag-8-Barr control. L1 ORF2 RNA pairwise comparison (ORF2/Alu + control *vs.* ORF2/Alu + HELZ2), p = 7.2 x 10^{-8}***. Alu RNA pairwise comparison (ORF2/Alu + control *vs.* ORF2/Alu + HELZ2), p = 0.018*. Bottom: western blotting using an anti-HA antibody was used to detect ORF2p. GAPDH served as a loading control. **(c)** *HELZ2 modestly reduces Alu RNA levels.* HeLa-HA cells were transfected with only pCMV-3Tag-8-barr (control), or co-transfected with Alu-*neo*^{Tet} (Alu) and either pCMV-3Tag-8-barr (control) or pALAF015 (HELZ2). X-axis, constructs name. Y-axis, the relative amount of Alu RNA (primer set: *mneoI* [Alu or L1]). Pairwise comparison: p = 0.64^{ns}. Values represent the mean ± SEM from three independent biological replicates. The *p*-values were calculated using a one-way ANOVA followed by a Bonferroni-Holm post-hoc tests. ns: not significant; * p<0.05; *** p<0.001.

4.2e The L1 RNA 5′UTR is recognized by HELZ2 to reduce both L1 RNA levels and IFN-α induction

HELZ2 overexpression inhibits both L1 and Alu retrotransposition; intriguingly, L1 5′UTR and L1 3′UTR sequences are shared between the monocistronic ORF2p and full-length L1 expression constructs used in these assays. However, the monocistronic ORF2p expression cassette that drives Alu retrotransposition contains a deletion of a conserved polypurine tract (Δppt) in the L1 3′UTR, although this deletion does not dramatically affect L1 retrotransposition (Moran et al., 1996); the full-length bicistronic L1 does not contain this deletion. Thus, I hypothesized that HELZ2 may recognize either the RNA structures or sequences in the L1 5′UTR and/or 3′UTR to destabilize L1 RNA.

To test this hypothesis, the L1 5′UTR sequence from a WT L1 expression construct (pTMF3) that also contains the 3′UTRΔppt sequence was deleted and L1 expression was driven solely from the cytomegalovirus immediate-early (CMV) promoter (**Fig. 4.9a**, L1 [Δ5′UTR]; a.k.a. pTMF3_Δ5UTR). We also replaced the L1.3 coding sequences (*ORF1* and *ORF2*) with a firefly luciferase gene, creating a construct that has the L1 5′UTR and L1 3′UTRΔppt sequences surrounding the luciferase gene as a control (**Fig. 4.9a**, Fluc; a.k.a pL1_[5&3UTRs]_Fluc). HeLa-JVM cells were then co-transfected with either pTMF3, pTMF3_Δ5UTR, or Fluc and HELZ2 followed by RT-qPCR (i.e., using probe sets that

specifically recognize the SV40 poly(A) signal of the plasmid, pTMF3, pTMF3_Δ5UTR, or Fluc RNAs; **see Methods**). Irrespective of their downstream sequences, HELZ2 overexpression was found to significantly reduce the RNA levels derived from only the L1 5′UTR-containing constructs (pTMF3 and Fluc, but not L1 [Δ5′UTR] construct) (**Fig. 4.9a**); in congruent, no significant change was observed for L1 retrotransposition in the L1 [Δ5'UTR] construct when HELZ2 was expressed (**Fig. 4.9b**).

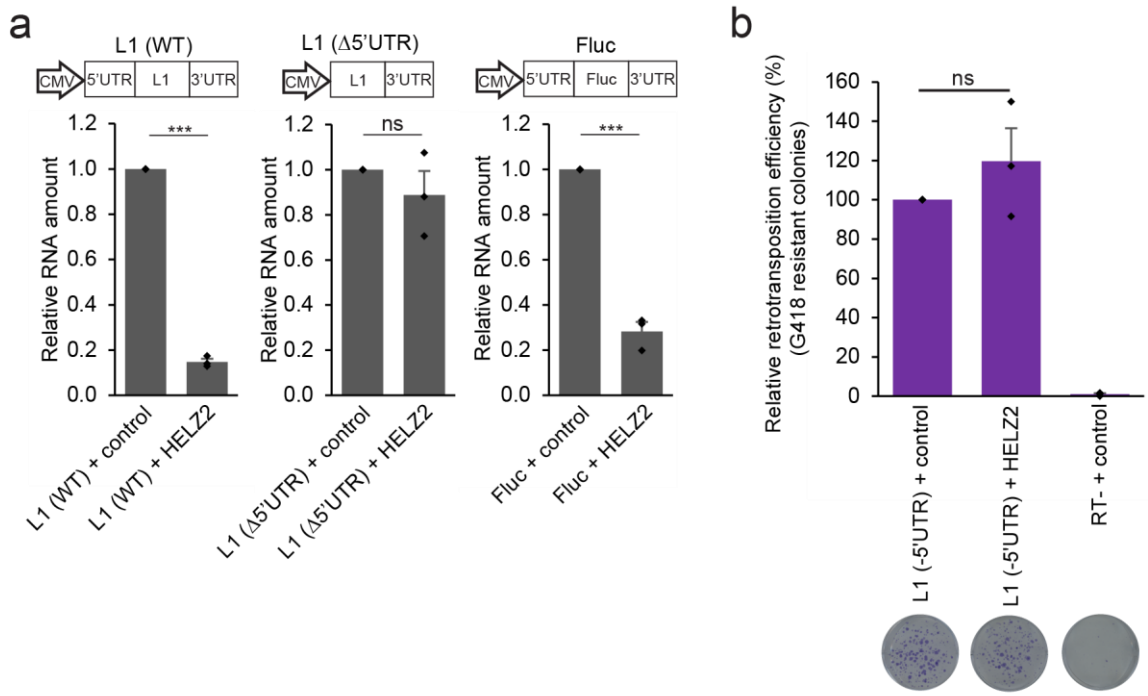


Fig. 4.9: *HELZ2 recognizes L1 5'UTR for L1 RNA destabilization.* **(a)** *The L1 5'UTR is required for HELZ2-mediated reduction of L1 RNA levels.* HeLa-JVM cells were co-transfected with L1 (WT), L1 (Δ5'UTR), or Fluc (a firefly luciferase gene flanked by the L1 5' and 3'UTRs) and either pCMV-3Tag-8-barr (control) or pALAF015 (HELZ2). Schematics of the constructs are above the respective bar charts. RNA levels were determined by RT-qPCR using the following primer sets: L1 (SV40) (for L1 WT and L1[Δ5'UTR]) or Luciferase (for Fluc) and then were normalized to *GAPDH* RNA levels (primer set: GAPDH). X-axis, name of respective constructs co-transfected with pCMV-3Tag-8-Barr (control) or pALAF015 (HELZ2); Y-axis, the relative amount of L1 or Fluc-based RNA relative to the relevant pairwise control (e.g., the L1 expression plasmid + pCMV-3Tag-8-Barr or the Fluc-based plasmid +

pCMV-3Tag-8-Barr). Two-tailed, unpaired Student's t-tests: p = 3.9 x 10^{-7}*** (left plot); 0.35^{ns} (middle plot); 7.1 x 10^{-5}*** (right plot). **(b)** *The L1 5'UTR is required for HELZ2-mediated reduction of L1 retrotransposition.* HeLa-JVM cells were co-transfected with L1 (Δ5'UTR) and either pCMV-3Tag-8-barr (control) or pALAF015 (HELZ2). Another set of HeLa-JVM cells were co-transfected with pCMV-3Tag-8-barr (control) or pALAF015 (HELZ2) and a phrGFP-C plasmid to normalize for transfection efficiencies and subjected to *mneoI*-based retrotransposition assays. X-axis, constructs name and representative results from the assay; a missense mutation in the ORF2p RT domain (RT-) served as a negative control. Y-axis, the percentage of normalized G418-resistant foci compared to the WT (pJM101/L1.3FLAG) control. Pairwise comparison: p = 0.2^{ns}. Values represent the mean ± SEM of three independent biological replicates. The p-values were calculated using a one-way ANOVA followed by Bonferroni-Holm post-hoc tests; ns: not significant.

HELZ2 overexpression in HeLa-JVM cells also did not affect steady state RNA or protein levels produced from an inducible Tet-On firefly luciferase or human L1 ORFeus construct that lack the L1 5′UTR (Ardeljan et al., 2020) (**Figs. 4.10**). Overall, these data suggest that HELZ2 destabilizes L1 RNA by recognizing RNA sequences and/or RNA structure(s) within the L1 5′UTR.

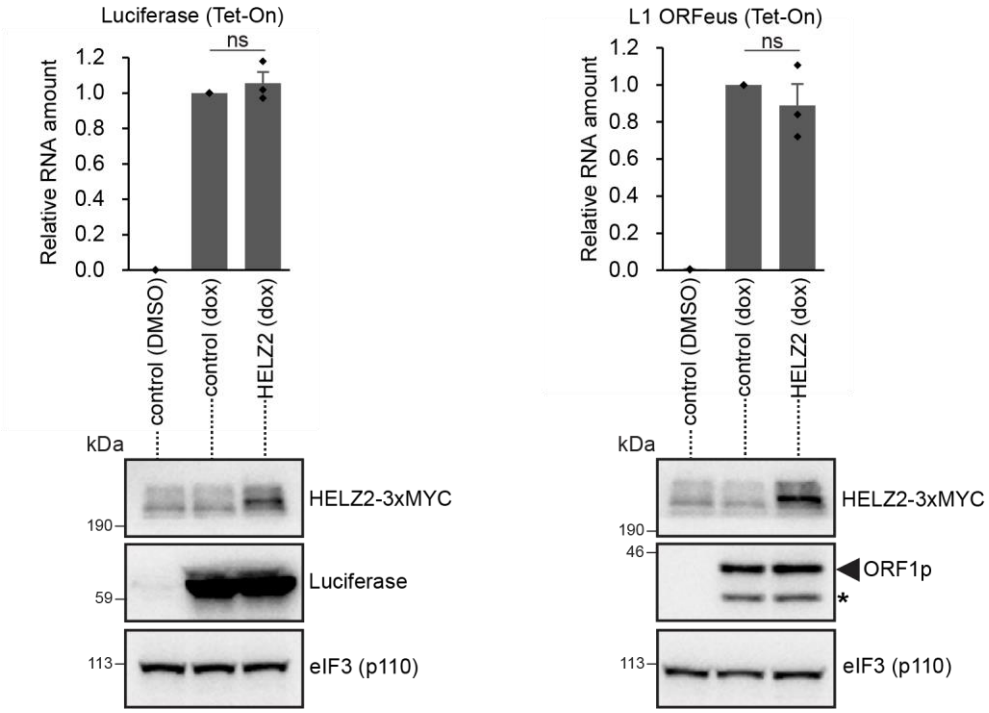


Fig. 4.10: *The effect of HELZ2 on doxycycline inducible (Tet-On) luciferase **(left)** or human L1 ORFeus **(right)** expression.* HeLa-JVM cells expressing inducible firefly luciferase (pSBtet-RN) or human L1 ORFeus (pDA093) were treated with vehicle (DMSO) or doxycycline (dox) and then transfected with either pCMV-3Tag-8-Barr (control) or pALAF015 (HELZ2). Cells were collected 48 hours post-transfection. Top: Luciferase and L1 levels were quantified using RT-qPCR (primer set: Luciferase and L1 [SV40], respectively) and normalized to *GAPDH* RNA levels (primer set: GAPDH). X-axis, construct name and whether cells were treated with vehicle (DMSO) or doxycycline (dox). Y-axis, RNA levels normalized to the inducible firefly luciferase (pSBtet-RN) or human L1 ORFeus (pDA093) + pCMV-3Tag-8-Barr control. Bottom: western blot analyses. An anti-MYC antibody was used to detect HELZ2, an anti-luciferase antibody was used to detect luciferase, and an anti-ORF1p antibody was used to detect ORF1p. Black arrowhead (middle right blot), the expected ORF1p band; asterisk (middle right blot), unexpected lower molecular weight ORF1p band. The eIF3 subunit (p110) served as a loading control. Pairwise comparisons: $p = 0.32^{ns}$ (Luciferase); and 0.28^{ns} (L1). Values represent the mean ± SEM of three independent biological replicates. The *p*-values were calculated using a one-way ANOVA followed by Bonferroni-Holm post-hoc tests; ns: not significant.

L1 RNA was reported to induce a type I IFN response previously (Mavragani et al., 2016; Zhao et al., 2018; Tunbak et al., 2020), which is consistent with our finding (**Fig. 3.3**). Since HELZ2 severely reduced L1 RNA levels, the effect of HELZ2 overexpression on L1-mediated IFN-α induction was examined. Strikingly, L1-dependent IFN-α induction was reduced to less than 5% of the control (pJM101/L1.3FLAG) upon HELZ2 overexpression (**Fig. 4.11**, compare the middle and right data graphs). Intriguingly, the IFN-α induction level was even lower than that observed in cells transfected with only the pCEP4 empty vector (**Fig. 4.11**, compare the left and right data graphs). Notably, co-expression of pCEP4 empty vector with HELZ2 also reduced the IFN-α level when compared to the pCEP4 only empty vector (**Fig. 4.11**), suggesting that HELZ2 overexpression may reduce not only ectopically expressed L1 RNA but also endogenous immunogenic RNAs, thereby reducing basal levels of IFN-α induction.

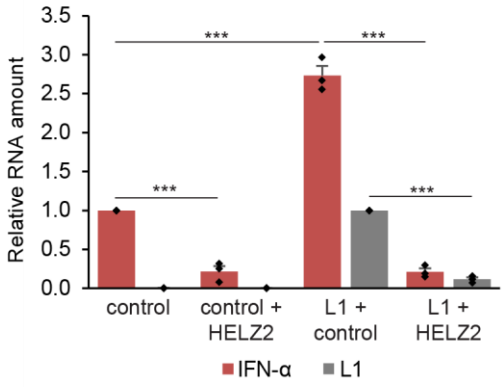


Fig. 4.11: *Expression of HELZ2 represses L1-induced IFN-α expression*. HEK293T cells were transfected with only the pCEP4 empty vector (control); or co-transfected with pCEP4 empty vector and pALAF015 (control + HELZ2); pJM101/L1.3FLAG and a pCEP4 empty vector (L1 + control); or pJM101/L1.3FLAG and pALAF015 (L1 + HELZ2). IFN-α (red bars) and L1 RNA (gray bars) levels were determined by RT-qPCR (using a primer set against IFNs [IFN-α, which amplifies IFN-α1 and IFN-α13] or the primer set *mneoI* [Alu or L1], respectively) and normalized to *ACTB* RNA levels (primer set: Beta-actin). The RNA levels then were normalized to the pCEP4 only (control) for IFN-α, and (L1 + control) for L1. L1 RNA pairwise comparison: (L1 + control *vs.* L1 + HELZ2), p = 1.4 x 10^{-10}***. IFN-α RNA pairwise comparisons: (control *vs.* control + HELZ2), p = 1.4 x 10^{-4}***; (control *vs*. L1 + control), p = 7.2 x 10^{-7}***; (L1 + control *vs.* L1 + HELZ2), p = 5.7 x 10^{-8}***. The values are reported as the mean ± SEM of three independent biological replicates. The p-values were calculated using a one-way ANOVA followed by Bonferroni-Holm post-hoc tests. ns: not significant; *** p<0.001.

Chapter 5

Discussion

5.1 Discussion

Previous literature showed that several ISG proteins inhibit L1 retrotransposition (see **chapter 1.4**: *"L1 retrotransposition inhibitors"* for a complete list) by destabilizing L1 RNA, L1 proteins, L1 RNPs, and perhaps L1 (-) strand cDNAs. In this study, String analysis of the IP-MS protein hits that were enriched in the WT ORF1p in comparison to the M8/RBM ORF1p mutant suggests that a network of ISG proteins may regulate L1 retrotransposition, supporting the idea that L1 is regulated through the innate immune response pathway (Yu et al., 2015; Mavragani et al., 2016; Zhao et al., 2018; Cecco et al., 2019; Simon et al., 2019; Tunbak et al., 2020). The ISG proteins in the network include 16 proteins that have not been reported to regulate L1 retrotransposition; I focused on three of these ISG proteins (HELZ2, HERC5, and OASL) and studied them in more detail.

Overexpression of HELZ2, HERC5, and OASL significantly reduced the retrotransposition of an engineered wild-type L1 (**Figs. 3.7a and 3.7b**) while localizing in the cytoplasm (**Fig. 3.9a**), suggesting a post-transcriptional inhibition. OASL was found to mainly impair ORF1p cytoplasmic formation (**Fig. 3.9**), while HERC5 destabilizes ORF1p but barely affects its cellular localization (**Figs. 3.7 and 3.9**). In comparison, HELZ2 reduces the steady state levels of engineered L1 RNA, ORF1p, and ORF1p cytoplasmic foci formation (**Figs. 3.7, 3.8, and 3.9**), possibly through RNA-dependent association with L1 RNP (**Fig. 4.6**). These results suggest that predominantly cytoplasmic ISG proteins may inhibit L1 retrotransposition at different steps in the L1 retrotransposition cycle to provide a multilayered defense against potential harm caused by L1.

HELZ2 is a poorly characterized protein that has a centrally-located RNB domain (Chu et al., 2017) that is flanked by two putative RNA helicase domains. RNB-containing proteins such as prokaryotic cold shock inducible protein RNase R and the yeast and human RNA exosome component Dis3 (Reis et al., 2013) are known to function in RNA quality control. Characterized RNB-containing proteins function to degrade highly structured RNAs through their concerted helicase and 3' to 5' exoribonuclease activities (Awano et al., 2010; Hossain

et al., 2015); thus, it is tempting to suggest that HELZ2 may function in a similar stepwise manner, where its helicase activity initially unwinds L1 RNA secondary structures to expose the single-stranded L1 RNA regions for degradation by the HELZ2 3' to 5' exoribonuclease activity (**Figs. 4.3c and 5.1**). Indeed, the HELZ2 helicase double mutant (WA1&2), but not a putative RNase-deficient mutant (dRNase), severely impaired the ability to inhibit L1 retrotransposition (**Figs. 4.3d, 4.3e, 4.4a, and 4.4b**). This suggests that the HELZ2 helicase activity may function upstream of the 3' to 5' exoribonuclease activity to degrade L1 RNA. It is possible that some regions of L1 RNA might be protected from HELZ2 degradation due to ORF1p RNA-binding because ORF1p-binding to L1 RNA is proposed to stabilize L1 RNAs (Khazina et al., 2011; Naufer et al., 2016), whereas other regions of L1 RNA that have complex RNA secondary structures may be targeted by HELZ2 for RNA destabilization to inhibit L1 retrotransposition.

CpG DNA methylation of the L1 5'UTR in the genome is previously shown to potently suppress endogenous L1 transcription (Bourc'his and Bestor, 2004; Coufal et al., 2009; Ewing et al., 2020). Krüppel-associated Box-containing Zinc-Finger Protein 93 (ZNF93) binds the genomic 5'UTR sequences of older L1s (e.g., members of the L1PA3 and L1PA4 subfamilies) to repress their transcription and deletion of these repressive sequences allowed the subsequent amplification of the L1PA2 and human-specific L1PA1 subfamilies in the human genome (Jacobs et al., 2014; Larson et al., 2018). Because the L1 5'UTR also contains multiple transcription factor binding sites for L1 expression and HELZ2 was initially found as a transcriptional co-activator of PPAR-*α* (Surapureddi et al., 2002) and PPAR-γ (Tomaru et al., 2006), it remains possible that HELZ2 may repress L1 transcription (Swergold, 1990; Tchenio et al., 2000; Yang et al., 2003; Athanikar et al., 2004; Sun et al., 2018).

Intriguingly, L1 RNAs containing 5'UTR sequences appear to be particularly susceptible to HELZ2-mediated RNA degradation (**Figs. 4.9 and 4.10**). This suggests a potential post-transcriptional mechanism by which RNA structures and/or RNA sequences within the 5'UTR might be targeted by host proteins to inhibit L1 retrotransposition. Cold shock and S1

domains that form an RNA-binding channel are usually found to flank RNB domains; however, prediction analysis based on an alignment with other RNB-containing proteins suggest that HELZ2 lacks the cold shock and S1 domains, as well as the conserved amino acids found in these domains (Chu et al., 2017). Thus, it remains unclear which domain of HELZ2 recognizes the 5'UTR region of L1 RNA.

A modest upregulation of IFN-α expression was observed upon the overexpression of an L1 construct; which was reported previously to contribute to inflammation, autoimmunity, and aging phenotypes (Mavragani et al., 2016; Zhao et al., 2018; Cecco et al., 2019; Simon et al., 2019; Ardeljan et al., 2020; Tunbak et al., 2020). The overexpression of an RT-deficient L1 caused a similar upregulation of type I IFN expression and was slightly more pronounced upon the overexpression of the ORF1p M8/RBM mutant, which compromises L1 RNP formation. These data suggest that not the intermediates generated during TPRT (e.g., L1 cDNAs), the L1-encoded proteins or the L1 RNP formation, but rather L1 RNA is responsible for the modest type I IFN upregulation observed in HEK293T cells. Consistently with L1 RNA reduction, HELZ2 overexpression severely reduced L1-mediated type I IFN upregulation to below the baseline levels when compared to the mock pCEP4 control (**Fig. 4.11**)**,** suggesting that HELZ2 may also destabilize endogenous immunogenic RNA(s). The data are in general agreement with other reports, which demonstrated that a relatively high L1 RNA expression levels can lead to the upregulation of the type I IFN response (Zhao et al., 2018; Ardeljan et al., 2020; Tunbak et al., 2020). Moreover, we also observed a modest increase in several cytokines in L1-transfected cells in comparison to the mock pCEP4-transfected cells (**Table 4**).

It is logical that L1s would benefit from not triggering an innate immune response as L1s lack an extracellular phase in their replication cycle. However, why the overexpression of the ORF1p M8/RBM mutant led to a more robust, yet modest, induction of type I IFN expression than the WT and RT-deficient L1s (**Fig. 3.3**) remains an unanswered question. One possible explanation is that ORF1p L1 RNA-binding and/or the sequestration of L1 RNPs within cytoplasmic foci effectively shields L1 RNAs from eliciting the interferon response and the

attenuated ability of the ORF1p M8/RBM mutant to bind L1 RNA could expose the L1 RNA as substrates for RNA sensors that can upregulate type I IFN expression (**Fig. 5.1**). The working model further implies that ORF1p binding to L1 RNA may attenuate the type I interferon response, which might lead to a reduction in the expression of inhibitory ISG proteins, suggesting another facet of ORF1p function in L1 retrotransposition. The data presented in this study is in agreement with a previous study which reported that depletion of the Human Silencing Hub (HUSH complex) facilitates the de-repression of primate-specific L1s and the resultant L1 double-stranded RNAs may drive physiological or autoinflammatory responses in human cells (Tunbak et al., 2020). Therefore, future studies are necessary to elucidate how L1 RNAs play important roles in innate immune activation and human autoimmune diseases, and how the host respond towards L1 intermediates' immunogenicity (Crow, 2010; Volkman and Stetson, 2014).

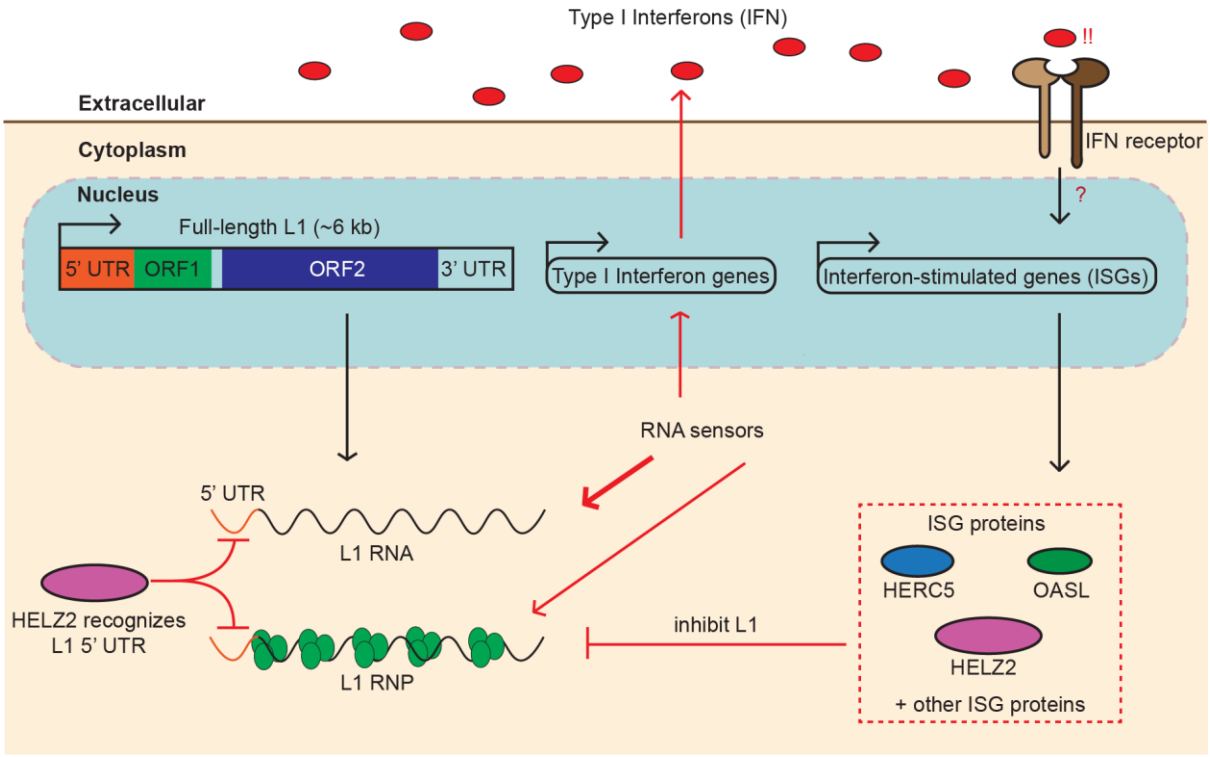


Fig. 5.1. A *working model for a negative feedback regulation between L1 expression and ISG proteins*. Cytoplasmic RNA sensors may detect L1 RNAs and/or RNPs to induce the secretion of type I interferons (IFNs); ORF1p RNA-binding may protect L1 RNA from being recognized by the sensors. IFN-binding to the extracellular IFN cell surface receptors then activates a signaling cascade, which induces the expression of ISGs, including *HERC5*, *OASL*, and *HELZ2*. These ISG proteins seem to inhibit L1 retrotransposition at different stages of an L1 retrotransposition life cycle. HELZ2 may recognize RNA sequences and/or RNA structures within the L1 5'UTR, independently of ORF1p RNA-binding, leading to the degradation of L1 RNA and subsequent reduction of the IFN response.

Materials and Methods

Cell lines and cell culture conditions

The human HeLa-JVM cervical cancer-derived (Moran et al., 1996), U-2 OS osteosarcoma-derived, and HEK293T embryonic carcinoma-derived cell lines were grown in Dulbecco's Modified Eagle Medium (DMEM) (Nissui, Tokyo, Japan) supplemented with 10% (volume/volume [v/v]) fetal bovine serum (FBS) (Gibco, Amarillo, Texas, United States or Capricorn Scientific, Ebsdorfergrund, Germany), 0.165% (weight/volume [w/v]) $NaHCO_3$, 100 U/mL penicillin G (Sigma-Aldrich, St. Louis, MO, United States), 100 µg/mL streptomycin (Sigma-Aldrich), and 2 mM L-glutamine (Sigma-Aldrich). HeLa-HA cells (Hulme et al., 2007) were grown in Minimum Essential Medium (MEM) (Gibco) supplemented with 10% (v/v) FBS (Capricorn Scientific), 0.165% (w/v) $NaHCO_3$, 100 U/mL penicillin G, 100 µg/mL streptomycin, 2 mM L-glutamine, and 1x MEM Non-Essential Amino Acids Solution (Nacalai, Kyoto, Japan). The cell lines were grown at 37°C in 100% humidified incubators supplied with 5% CO_2. The cell lines tested negative for mycoplasma contamination using a PCR-based method using the VenorGeM Classic Mycoplasma Detection Kit (Sigma-Aldrich). STR-genotyping was performed to confirm the identity of HeLa-JVM, HeLa-HA, U-2 OS, and HEK293T cells.

Plasmid construction

Creation of the ORF1p-FLAG mutant constructs: Briefly, the pJM101/L1.3FLAG (Beck et al., 2010; Moldovan and Moran, 2015) plasmid was used, unless otherwise indicated, to construct the plasmids in this study. Briefly, pJM101/L1.3FLAG DNA was used as a PCR template in conjunction with oligonucleotide primers containing the respective ORF1p mutations to generate the ORF1 mutants. The amplified PCR products and pJM101/L1.3FLAG plasmid DNA then were digested with *Not*I and *Age*I and ligated using the DNA Ligation Kit Mighty Mix (TaKaRa Bio, Shiga, Japan) at 16°C for 30 minutes. The resultant ligation products were transformed into *E. coli* XL1-Blue cells and plated on Luria

Broth (LB) agar plates containing 50 µg/mL ampicillin. The resultant plasmids then were sequenced from the *Not*I to *Age*I restriction sites to ensure the integrity of the mutants.

Creation of the mCherry/G3BP1, ISG fusion protein, and HELZ2 mutant constructs: The *G3BP1* cDNA sequence was amplified from a HeLa-JVM total cDNA library and concurrently inserted in-frame with an *mCherry*-coding sequence into a lentiviral vector (pCW) (Wang et al., 2014) using the in-Fusion Cloning Kit (TaKaRa Bio). The *HERC5*, *HELZ2*, *OASL*, *MOV10*, *IFIT1*, and *DDX60L* cDNAs were amplified from either a HeLa-JVM or HEK293T total cDNA library and inserted into the *Not*I and *Hind*III restriction sites in the pCMV-3Tag-9 vector (Agilent Technologies, Santa Clara, CA, United States) using either the Gibson Assembly Cloning Kit (New England Biolabs, Ipswich, MA, United States) or in-Fusion Cloning Kit. To generate *HELZ2* mutations, whole plasmid DNAs were amplified using primers harboring the intended mutations in separate reactions to avoid the formation of primer dimers. The template DNA then was digested with *Dpn*I at 37°C for 1 hour followed by heat inactivation at 80°C for 20 minutes. The PCR amplified DNA fragments were mixed, annealed, and transformed into *E*. *coli* XL1-Blue cells.

Plasmids used in this study

For mammalian cell experiments, plasmids were purified using the GenElute HP Plasmid Midiprep Kit (Sigma-Aldrich). All of the L1 expression plasmids contain a retrotransposition-competent L1 (L1.3, Genbank: L19088). The amino acid residues of ORF1p or ORF2p were counted from the first methionine of the L1.3 ORF1p and L1.3 ORF2p, respectively. The plasmids used in the study are listed below:

pCEP4 (Invitrogen): the mammalian expression vector backbone used for cloning pJM101/L1.3 and pJJ101/L1.3 variants.

phrGFP-C (Agilent technology): contains a humanized Renilla *GFP* gene whose expression is driven by a cytomegalovirus immediate-early (CMV) promoter.

pJM101/L1.3: was described previously (Sassaman et al., 1997; Wei et al., 2000). This plasmid contains the full-length L1.3, cloned into the pCEP4 vector plasmid. L1 expression is

driven by the CMV and L1.3 5'UTR promoters. The *mneoI* retrotransposition cassette was inserted into the L1.3 3'UTR as described previously (Moran et al., 1996).

pJM101/L1.3FLAG: was described previously (Moldovan and Moran, 2015). This plasmid is a derivative of pJM101/L1.3 that contains a single copy of the *FLAG* epitope tag fused in-frame to the 3' end of the L1.3 *ORF1* sequence.

pJM105/L1.3: was described previously (Wei et al., 2001). This plasmid is a derivative of pJM101/L1.3 that contains a D702A mutation in the ORF2p reverse transcriptase active site.

pTMF3: was described previously (Miyoshi et al., 2019). This plasmid is a derivative of pJM101/L1.3. A *T7 gene10* epitope tag was fused in-frame to the 3' end of the *ORF1* sequence and three copies of a *FLAG* epitope tag were fused to the 3' end of the *ORF2* sequence. This plasmid lacks the polypurine sequence in the L1 3'UTR.

pTMF3 ∆5UTR: is a derivative of pTMF3 that contains a deletion of the L1.3 5'UTR sequence.

pTMF3_M8_ORF1: is a derivative of pTMF3 that contains the M8 (RBM) mutations: R206A, R210A, and R211A in ORF1p, which impairs the ability of ORF1p to bind RNA (Khazina et al., 2011).

pL1(5&3UTRs)_Fluc: is a derivative of pTMF3 that contains a firefly luciferase gene in place of the L1.3-coding region.

pJJ101/L1.3: was described previously (Kopera et al., 2011). This plasmid is similar to pJM101/L1.3, but contains an *mblastI* retrotransposition indicator cassette within the L1.3 3'UTR.

pJJ105/L1.3: was described previously (Kopera et al., 2011). This plasmid is a derivative of pJJ101/L1.3 that contains a D702A mutation in the ORF2p reverse transcriptase active site.

pALAF001_L1.3FLAG_M1: is a derivative of pJM101/L1.3FLAG that contains the N157A and R159A mutations in ORF1p, which abolished ORF1p cytoplasmic foci formation (Goodier et al., 2007).

pALAF002_L1.3FLAG_M2: is a derivative of pJM101/L1.3FLAG that contains the R117A and E122A mutations in ORF1p, which are proposed to adversely affect ORF1p trimerization (Kammerer et al., 2005).

pALAF003_L1.3FLAG_M3: is a derivative of pJM101/L1.3FLAG that contains the N142A mutation in ORF1p, which is proposed to bind a chloride ion to stabilize ORF1p trimerization (Khazina et al., 2011).

pALAF004_L1.3FLAG_M4: is a derivative of pJM101/L1.3FLAG that contains the R135A mutation in ORF1p, which is proposed to bind a chloride ion to stabilize ORF1p trimerization (Khazina et al., 2011).

pALAF005_L1.3FLAG_M5: is a derivative of pJM101/L1.3FLAG that contains the E116A and D123A mutations in ORF1p, which are proposed to act as a binding site for host factors (Khazina et al., 2011).

pALAF006_L1.3FLAG_M6: is a derivative of pJM101/L1.3FLAG that contains the K137A and K140A mutations in ORF1p, which reduces the ability of ORF1p to bind L1 RNA (Khazina et al., 2011).

pALAF007_L1.3FLAG_M7: is a derivative of pJM101/L1.3FLAG that contains the R235A mutation in ORF1p, which reduces the ability of ORF1p to bind L1 RNA (Doucet et al., 2010).

pALAF008_L1.3FLAG_M8_(RBM): is a derivative of pJM101/L1.3FLAG that contains the R206A, R210A, and R211A mutations in ORF1p, which severely impair the ability of ORF1p to bind RNA (Khazina et al., 2011).

pALAF009_L1.3FLAG_M9: is a derivative of pJM101/L1.3FLAG that contains the R261A mutation in ORF1p, which reduces the ability of ORF1p to bind L1 RNA (Doucet et al., 2010).

pALAF010_L1.3FLAG_M10: is a derivative of pJM101/L1.3FLAG that contains the Y282A mutation in ORF1p, which is proposed to reduce nucleic chaperone activity (Doucet et al., 2010).

pALAF012_mCherry-G3BP1_pCW: contains the *mCherry* sequence fused in frame to a human *G3BP1* cDNA in a lentiviral expression vector, pCW (Wang et al., 2014). The puromycin resistant gene and reverse tetracycline-controlled trans-activator (rtTA) coding

regions are in-frame and are expressed by a human PGK promoter; puromycin and rtTA are separated by a self-cleaving T2A peptide so that each protein can be expressed from the bicistronic transcript. The *mCherry-G3BP1* cDNA is expressed from a doxycycline inducible (Tet-On) promoter. In the presence of doxycycline, rtTA can adopt an altered confirmation that allows it to bind the Tet-On promoter to allow *mCherry-G3BP1* expression.

pALAF015_hHELZ2L-3xMYC: contains the canonical human *HELZ2* long isoform cDNA (2649 bps) cloned into pCMV-3Tag-9 (Agilent Technologies), which allows the expression of a HELZ2-3xMYC fusion protein. The CMV promoter drives *HELZ2-3xMYC* expression.

pALAF016_hIFIT1-3xMYC: contains the human *IFIT1* cDNA cloned into pCMV-3Tag-9, which allows the expression of a hIFIT1-3xMYC fusion protein. The CMV promoter drives *IFIT1-3xMYC* expression.

pALAF021_hDDX60L-3xMYC: contains the human *DDX60L* cDNA cloned into pCMV-3Tag-9, which allows the expression of a hDDX60L-3xMYC fusion protein. The CMV promoter drives *DDX60L-3xMYC* expression.

pALAF022_hOASL-3xMYC: contains the human *OASL* cDNA cloned into pCMV-3Tag-9, which allows the expression of the OASL-3xMYC fusion protein. The CMV promoter drives *OASL-3xMYC* expression.

pALAF023_hHERC5-3xMYC: contains the human *HERC5* cDNA cloned into pCMV-3Tag-9, which allows the expression of a HERC5-3xMYC fusion protein. The CMV promoter drives *HERC5-3xMYC* expression.

pALAF024_hMOV10-3xMYC: contains the human *MOV10* cDNA cloned into pCMV-3Tag-9, which allows the expression of a MOV10-3xMYC fusion protein. The CMV promoter drives *MOV10-3xMYC* expression.

cepB-gfp-L1.3: was described previously (Miyoshi et al., 2019). The plasmid contains the full-length L1.3 with an *EGFP* retrotransposition reporter cassette, *mEGFPI*. L1.3 expression is augmented by the L1 5'UTR promoter. The plasmid backbone also contains a *blasticidin S-deaminase* (*BSD*) selectable marker driven by the SV40 early promoter.

cepB-gfp-L1.3RT(-) intronless: was described previously (Miyoshi et al., 2019). The plasmid is similar to cepB-gfp-L1.3RT(-) except that the intron in the *mEGFPI* retrotransposition cassette was removed, allowing EGFP expression in the absence of L1.3 retrotransposition.

cep99-gfp-L1.3: was described previously (Miyoshi et al., 2019). The plasmid is similar to cepB-gfp-L1.3 but contains the puromycin resistant gene instead of the blasticidin resistance gene as a selectable marker.

cep99-gfp-L1.3RT(-) intronless: was described previously (Miyoshi et al., 2019). The plasmid is similar to cep99-gfp-L1.3 except that it contains the D702A mutation in the ORF2p reverse transcriptase domain and the intron in the *mEGFPI* retrotransposition cassette was removed, allowing EGFP expression in the absence of L1.3 retrotransposition.

pALAF025_hHELZ2L-3xMYC_WA1: is a derivative of pALAF015_hHELZ2L-3xMYC that contains the K550A mutation in the Walker A motif of the N-terminal HELZ2 helicase domain, which is predicted to inactivate the ATP binding ability of the helicase domain (Walker et al., 1982).

pALAF026_hHELZ2L-3xMYC_WA2: is a derivative of pALAF015_hHELZ2L-3xMYC that contains the K2180A mutation in the Walker A motif of the carboxyl-terminal HELZ2 helicase domain, which is predicted to inactivate the ATP binding ability of the helicase domain (Walker et al., 1982).

pALAF027_hHELZ2L-3xMYC_WA1&2: is a derivative of pALAF015_hHELZ2L-3xMYC that contains the K550A and K2180A mutations in the Walker A motifs of both HELZ2 helicase domains (Walker et al., 1982).

pALAF028_hHELZ2L-3xMYC_WB1: is a derivative of pALAF015_hHELZ2L-3xMYC that contains the E668A mutation in the Walker B motif of the N-terminal helicase domain of HELZ2, which is predicted to inactivate the ATP hydrolysis activity of the helicase domain (Walker et al., 1982).

pALAF029_hHELZ2L-3xMYC_WB2: is a derivative of pALAF015_hHELZ2L-3xMYC that contains the E2361A mutation in the Walker B motif of the C-terminal helicase domain of

HELZ2, which is predicted to inactivate the ATP hydrolysis activity of the helicase domain (Walker et al., 1982).

pALAF030_hHELZ2L-3xMYC_dRNase: is a derivative of pALAF015_hHELZ2L-3xMYC that contains the D1346N, D1354N, and D1355N mutations in the RNB domain of HELZ2, which is predicted to inactivate the RNase activity of the RNB domain (Barbas et al., 2008).

pALAF071_hHELZ2L-3xFLAG: is a derivative of pALAF015_hHELZ2L-3xMYC where the *3xMYC* epitope tag was replaced with a *3xFLAG* epitope tag.

pALAF073_hHELZ2L-3xFLAG_dRNase: is a derivative of pALAF030_hHELZ2L-3xMYC_dRNase where the *3xMYC* epitope tag was replaced with a *3xFLAG* epitope tag.

psPAX2: is a lentivirus packaging vector that was a gift from Didier Trono (Addgene plasmid # 12260). The plasmid expresses the HIV-1 gag and pol proteins.

pMD2.G: is a lentivirus envelope expression vector that was a gift from Didier Trono (Addgene plasmid # 12259). The plasmid expresses a viral envelope protein and the vesicular stomatitis virus G glycoprotein (VSV-G).

pcDNA6: was described previously (Miyoshi et al., 2019). It is a derivative of pcDNA6/TR (Invitrogen, Carlsbad, CA, United States) and contains the *blasticidin S-deaminase* (*BSD*) selectable marker but lacks the TetR gene. This plasmid was made by Dr. John B. Moldovan (University of Michigan Medical School).

pCMV-3Tag-8-Barr: is a human *β-Arrestin* expression plasmid. The human *ARRB2* cDNA was cloned into pCMV-3Tag-8 (Agilent Technologies). The plasmid *contains three copies of a FLAG epitope tag fused in-frame to the 3' end of the ARRB2* cDNA. The CMV promoter drives *ARRB2-3xFLAG* expression.

Alu-neo^{Tet}: was described previously (Dewannieux et al., 2003). The plasmid contains an AluY element with the *neo*^{Tet} retrotransposition indicator cassette, which was inserted upstream of the Alu poly(A) tract. Alu expression is augmented by a 7SL promoter.

pTMO2F3_Alu: is a plasmid that co-expresses Alu and a monocistronic version of L1 ORF2p that contains the L1 5'UTR. The monocistronic *ORF2* coding sequence contains three

copies of an in-frame *FLAG* epitope tag sequence at its 3' end; the CMV promoter augments the expression of *ORF2-3xFLAG*. The plasmid also contains an AluY element whose expression is driven by a 7SL promoter. The Alu element contains the *neo*^{Tet} retrotransposition indicator cassette (Dewannieux et al., 2003), which was inserted upstream of the Alu poly(dA) tract. This arrangement allows the quantification of Alu retrotransposition efficiency by counting the resultant number of G418-resistant foci. This plasmid lacks the polypurine sequence in the L1 3'UTR.

pTMO2F3D145AD702A_Alu: is identical to pTMO2F3_Alu but contains the D145A and D702A mutations, which inactivate the ORF2p endonuclease and reverse transcriptase activities, respectively.

pTMO2H3_Alu: is a derivative of pTMO2F3_Alu plasmid where the *3xFLAG* epitope tag was replaced with three copies of *HA* epitope tag sequence.

pSBtet-RN: was a gift from Eric Kowarz (Kowarz et al., 2015; Ardeljan et al., 2020) (Addgene plasmid # 60503). The plasmid contains a firefly luciferase (*Fluc*) gene with an upstream Tet-On inducible promoter.

pDA093: was a gift from Kathleen Burns (Ardeljan et al., 2020) (Addgene plasmid # 131390). This plasmid is similar to pSBtet-RN but the luciferase gene was replaced with the human L1 *ORFeus* (*ORF1* and *ORF2*) sequence lacking the 5' or 3'UTR.

pCMV(CAT)T7-SB100: was a gift from Zsuzsanna Izsvak (Mátés et al., 2009) (Addgene plasmid # 34879). This plasmid contains a hyperactive variant of the *Sleeping Beauty* transposase, whose expression is driven by the CMV promoter.

Western blots

HeLa-JVM, U-2 OS, or HEK293T cells were seeded in a 6-well tissue culture plate (Greiner, Frickenhausen, Germany) at $2x10^5$ cells per well. On the following day, the cells were transfected with 1 µg of DNA (1 µg of an L1-expressing plasmid or 0.5 µg of the L1-expressing plasmid and 0.5 µg of either a pCMV-3Tag-8-Barr control or ISG-expressing plasmid) using 3 µL of FuGENE HD transfection reagent (Promega, Madison, WI, United

States) and 100 µL of Opti-MEM (Gibco) according to the protocol provided by the manufacturer. The medium was replaced with fresh DMEM approximately 24 hours post-transfection (day 1). The cells were harvested using 0.25% (v/v) trypsin (Gibco) at days 2 through 9 post-transfection (depending on the specific experiment). The transfected cells were enriched using 100 µg/mL of hygromycin B (Wako, Osaka, Japan), which was added to the media two days post-transfection and replaced with fresh DMEM containing hygromycin B daily. After collection by trypsinization, the cells were pelleted by centrifugation at 300 x *g* for 5 minutes. Then, the cells were washed twice with cold 1x PBS, flash-frozen in liquid nitrogen, and kept at -80°C.

For cell lysis, the cells were incubated in Radio-ImmunoPrecipitation Assay (RIPA) buffer (10 mM Tris-HCl [pH 7.5], 1 mM EDTA, 1% [v/v] TritonX-100, 0.1% [w/v] sodium deoxycholate, 0.1% [w/v] SDS, 140 mM NaCl, 1x cOmplete EDTA-free protease inhibitor cocktail [Roche, Mannheim, Germany]) at 4°C for 30 minutes. The cell debris was pelleted at 12,000 x *g* for 5 minutes and the supernatant was collected. The protein concentration was measured using the Protein Assay Dye Reagent Concentrate (Bio-Rad, Richmond, CA, United States) and all of the samples for each experiment were normalized to the same concentration. The protein lysate was mixed at an equal volume with 3x SDS sample buffer (187.5 mM Tris-HCl [pH 6.8], 30% [v/v]) glycerol, 6% [w/v] SDS, 0.3M DTT, 0.01% [w/v] bromophenol blue) and boiled at 105°C for 5 minutes. Twenty micrograms of total protein lysates for all samples were separated using sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE). Proteins on the gel were transferred onto Immobilon-P, 0.45 µm pore, polyvinylidene difluoride (PVDF) transfer membranes (Merck Millipore, Billerica, MA, United States) using 10 mM CAPS buffer (3-[cyclohexylamino]-1-propanesulfonic acid [pH 11]) in a Mini Trans-Blot Electrophoretic Transfer Cell tank (Bio-Rad) according to protocol provided by the manufacturer. The transfer was performed at 4°C at 50V for 16 hours. After the transfer was completed, the membrane was incubated with Tris-NaCl-Tween (TNT) buffer (0.1 M Tris-HCl [pH 7.5], 150 mM NaCl, 0.1% [v/v] Tween 20) containing 3% skim milk (Nacalai) for 30 minutes. The membranes then were washed with TNT buffer, cut

into strips, and incubated with the relevant primary antibodies in TNT buffer at 4˚C overnight. The next day, the membranes were washed four times with TNT buffer with five minutes interval at room temperature and incubated with HRP-conjugated secondary antibodies in TNT buffer containing 0.01% (w/v) SDS at room temperature for an hour. The membranes were washed four times with TNT buffer with five minutes interval at room temperature and the signals were detected with the Chemi-Lumi One L (Nacalai) or Chemi-Lumi One Super (Nacalai) chemiluminescence reagent using a LAS-3000 Imager (Fujifilm, Tokyo, Japan), LAS-4000 Imager (Fujifilm), or a FUSION Solo S Imager (Vilber-Lourmat, Marne-la-Vallee, France).

Primary antibodies and dilutions (in parentheses):

Please note: we tested two different anti-HELZ2 antibodies (Abcam [AB129781] and Affinity Biosciences [DF4285]), but the antibodies were not able to detect the endogenous HELZ2 protein in our experimental conditions.

Mouse monoclonal anti-FLAG M2 antibody (1/5000), (Sigma-Aldrich, F1804, 1.0 mg/mL, RRID: AB_262044)

Rabbit polyclonal anti-FLAG antibody (1/5000), (Sigma-Aldrich, F7425, ~0.8 mg/mL, RRID: AB_439687)

Mouse monoclonal anti-MYC antibody (1/5000), (Cell Signaling Technology, 9B11, RRID: AB_331783)

Rabbit polyclonal anti-PABPC1 antibody (1/5000), (Abcam, ab21060, 0.9 mg/mL, RRID: AB_777008)

Mouse monoclonal anti-GAPDH antibody (1/5000), (Millipore, MAB374, 1.0 mg/mL, RRID: AB_2107445)

Mouse monoclonal anti-Actin antibody (1/5000, diluted to 0.2 times of the original concentration), (Millipore, MAB1501R, RRID: AB_2223041)

Rabbit polyclonal anti-T7-tag antibody (1/5000), (Cell Signaling Technology, D9E1X, RRID: AB_2798161)

Goat polyclonal anti-Luciferase antibody (1/2000), (Promega, G7451, 1.0 mg/mL, RRID: AB_430862)

Mouse monoclonal anti-ORF1p (4H1) antibody (1/2000), (Millipore, MABC1152, 0.5 mg/mL)

Mouse monoclonal anti-eIF3 p110 (B-6) antibody (1/5000), (Santa Cruz Biotechnology, sc-74507, 0.2 mg/mL, RRID: AB_1122487)

Secondary antibodies and dilutions (in parentheses):

Sheep polyclonal anti-mouse HRP-conjugated Whole antibody (1/5000), (GE Healthcare, NA931-1ML, RRID: AB_772210)

Goat polyclonal anti-rabbit HRP-conjugated Whole antibody (1/5000), (Cell Signaling Technology, 7074, RRID: AB_2099233)

Donkey polyclonal anti-rabbit HRP-conjugated Whole antibody (1/5000), (GE Healthcare, NA934-1ML, RRID: AB_772206)

Donkey polyclonal anti-goat HRP-conjugated Whole antibody (1/5000), (Santa Cruz Biotechnology, sc-2020, 0.4 mg/mL, RRID: AB_631728)

Immunofluorescence

Cell transfection and fixation: HeLa-JVM or U-2 OS cells were plated on 18 mm glass coverslips (Matsunami Glass, Osaka, Japan) coated with Alcian Blue 8GX (Sigma-Aldrich) in 12-well tissue culture plates (Greiner) at $2.5x10^4$ cells per well in DMEM (with 1.0 µg/mL of doxycycline in mCherry-G3BP1-expressing U-2 OS cells). After 24 hours, the cells were transfected with 0.5 µg of plasmid DNA (0.5 µg of the L1-expressing plasmid [pJM101/L1.3FLAG, pALAF002, pALAF005, or pALAF008] or 0.25 µg of pJM101/L1.3FLAG and 0.25 µg of either a pCMV-3Tag-8-Barr control or ISG-expression plasmid) using 1.5 µL of FuGENE HD transfection reagent and 50 µL of Opti-MEM according to protocol provided by the manufacturer. Approximately 24 hours post-transfection, the medium was replaced with fresh DMEM and 1.0 µg/mL of doxycycline was added into the medium for mCherry-G3BP1-expressing U-2 OS cells. Approximately 48 hours post-transfection, the cells were washed with 1x PBS and fixed with 4% paraformaldehyde (PFA) at room temperature for 15

minutes. Prior to cell fixation, the cells were treated with DMSO (Sigma-Aldrich) or 0.5 mM sodium meta-arsenite (Sigma-Aldrich) for one hour. The fixed cells then were washed with 1x PBS three times and kept at 4°C until cell permeabilization.

Immunostaining: The resultant cells were permeabilized with 0.2% (v/v) Triton X-100 and 0.5% (v/v) normal donkey serum (NDS) for 5 minutes. The cells were washed once with 1x PBS and twice with PBST (1x PBS and 0.1% [v/v] Tween 20) following permeabilization. The primary antibodies (1/1000 dilution in PBST) containing 0.5% (v/v) NDS were applied onto the coverslip and incubated for 45 minutes at room temperature. The cells were washed with PBST three times after the primary antibody incubation. The secondary antibodies (1/250 dilution in PBST) containing 0.5% (v/v) NDS and 0.1 µg/mL of 4', 6-diamidino-2-phenylindole (DAPI) were applied onto the coverslip and incubated for 45 minutes at room temperature. The cells were washed with PBST three times followed by multiple rinses with water. The excess liquid was removed, and the glass coverslips were fixed on glass slides with 3 µL of VECTASHIELD (Vector Laboratories, Burlingame, CA, United States).

Immunofluorescence: Images were captured using the DeltaVision Elite microscope with DeltaVision softWoRx 5.5 software (Cytiva, Marlborough, MA, United States). Six z-stack images with 1 µm thickness difference were captured and projected into a single image with the max intensity for each image. For ORF1p-FLAG probed with the Alexa 488-conjugated antibody or MYC-tagged proteins probed with the Cy5-conjugated antibody, the FITC/AF488 or Cy5/AF647 channel was used, respectively. mCherry-G3BP1 fluorescence was detected through the mCherry/AF594 channel. In the ORF1p foci counting experiments, the same signal intensity threshold was applied to all samples and only cells with visible ORF1p signals were counted as positive cells. Only cells that displayed clear cytoplasmic ORF1p signals with foci distinguishable from the background were counted as an L1 foci-positive cells.

Primary antibodies and dilutions (in parentheses):

Mouse monoclonal anti-FLAG M2 antibody (1/1000), (Sigma-Aldrich, F3165, 3.8 - 4.2 mg/mL, RRID: AB_259529)

Rabbit polyclonal anti-FLAG antibody (1/1000), (Sigma-Aldrich, F7425, ~0.8 mg/mL, RRID: AB_439687)

Mouse monoclonal anti-MYC antibody (1/1000), (Cell Signaling Technology, 9B11, RRID: AB_331783)

Secondary antibodies and dilutions (in parentheses):

Donkey anti-mouse polyclonal Alexa Fluor 488 IgG (H+L) (1/250), (Thermo Fisher Scientific, A-21202, 2.0 mg/mL, RRID: AB_141607)

Donkey anti-rabbit polyclonal Alexa Fluor 488 IgG (H+L) (1/250), (Thermo Fisher Scientific, A-21206, 2.0 mg/mL, RRID: AB_2535792)

Goat polyclonal anti-mouse Cy5 (1/250), (Jackson ImmunoResearch Labs, 115-175-146, RRID: AB_2338713)

Lentiviral transduction

HEK293FT cells were plated in a 10-cm tissue culture dish at $1x10^6$ cells per plate. On the following day, the cells were transfected with 5 µg plasmid DNA (2.5 µg of pALAF012, 1.875 µg of psPAX2, and 0.625 µg of pMD2.G) using 15 µL of 1 mg/mL transfection grade linear *polyethylenimine hydrochloride (MW 40,000) (PEI-MAX-40K) (Polysciences, Warrington, PA, United States) in 500 µL of Opti-MEM*. Approximately 24 hours post-transfection, the medium was replaced with fresh DMEM. The medium containing the virus was collected 48 hours post-transfection and filtered through a 0.45 µm polyethersulfone (PES) filter (Merck Millipore).

To generate the inducible mCherry-G3BP1-expressing U-2 OS cell line, $2x10^5$ cells per well were plated in a 6-well tissue culture plate. On the next day, the medium was replaced with virus-containing medium supplemented with 8 µg/mL of polybrene (Sigma-Aldrich). Approximately 24 hours post-viral treatment, the medium was replaced with fresh DMEM. From the second day post-viral treatment onwards, the media was replaced with fresh DMEM containing 1 µg/mL puromycin every three days until the non-transduced cells were dead.

Construction of cell lines expressing Tet-On Luciferase and human L1 ORFeus

HeLa-JVM cells were plated in 6-well plates at $2x10^5$ cells per well. On the following day, the cells were transfected with 500 ng of plasmid DNA (pSBtet-RN or pDA093) and 50 ng of a sleeping beauty plasmid (pCMV[CAT]T7-SB100) using 2.0 µL of FuGENE HD transfection reagent and 100 µL of Opti-MEM according to the protocol provided by the manufactures. After ~24 hours, the medium was replaced with fresh DMEM. G418 (Nacalai) selection (500 µg/mL) began ~48 hours post-transfection for 1 week; the G418 containing media was replaced daily. Five percent of the total living cells were transferred into 10-cm tissue culture dishes and the media was replaced daily with 500 µg/mL G418 until the cells reached ~90% confluency. The cells then were trypsinized and resuspended in PBS containing 2% (v/v) FBS and dTomato-positive cells were sorted using a BD FACSAria III flow cytometer with BD FACSDiva Software v.6.1.3 (BD Biosciences, San Jose, CA, United States) to obtain clonal cell lines. Western blotting was used to screen the resultant cell lines for doxycycline dosage-dependent expression of Luciferase or human L1 ORFeus.

L1 and Alu retrotransposition assays

L1 or Alu cultured cell retrotransposition assays were performed as described with modifications (Moran et al., 1996; Ostertag et al., 2000; Wei et al., 2000; Dewannieux et al., 2003; Goodier et al., 2015; Kopera et al., 2016).

In retrotransposition assays using the *mneoI* retrotransposition indicator cassette, $2x10^5$ HeLa-JVM or HeLa-HA cells per well were seeded in 6-well tissue culture plates. On the following day, the cells were transfected with 1 µg of DNA (0.5 µg of pJM101L1.3/FLAG or its variants and 0.5 µg of phrGFP-C for the L1 retrotransposition assay) or 1 µg of DNA (0.5 µg of pTMO2F3_Alu or phrGFP-C and 0.5 µg of pCMV-3Tag-8-Barr control, pALAF015 [HELZ2], or pALAF024 [MOV10] for the Alu retrotransposition assay) using 3 µL FuGENE HD and 100 µL of Opti-MEM according to the protocol provided by the manufacturer. The medium was replaced with fresh DMEM (HeLa-JVM) or MEM (HeLa-HA), respectively ~24 hours post-transfection (day 1). On day 3 post-transfection, to check transfection efficiency,

each duplicate was collected, fixed with 0.5% paraformaldehyde, and subjected to flow cytometry analysis using BD Accuri C6 Plus Flow Cytometer (BD Biosciences). The FITC channel was used to determine the number of hrGFP-expressing cells out of 10, 000 cells as a transfection efficiency control. The medium in the remaining transfectants was replaced daily with fresh DMEM or MEM containing 500 µg/mL G418 from day 3 onwards. The resultant colonies were fixed at day 10-14 post-transfection using the fixation solution (1x PBS containing 0.2% [v/v] glutaraldehyde and 2% [v/v] formaldehyde). The cells were stained with 0.1% (w/v) crystal violet. The resultant number of foci were counted and normalized to the transfection efficiency. Please note: the HEK293T cells are G418-resistant and could not be used in *mneoI* based retrotransposition assays.

In retrotransposition assays using the *mblastI* retrotransposition indicator cassette, $5x10^4$ HeLa-JVM cells per well were seeded in 6-well tissue culture plates. After ~24 hours, the cells were transfected with 1 µg of DNA (0.5 µg of pJJ101/L1.3 and 0.5 µg of an ISG-expressing plasmid or pCMV-3Tag-8-Barr) using 3 µL of FuGENE HD in 100 µL of Opti-MEM. For the viability control, $5x10^3$ HeLa-JVM cells per well were seeded in 6-well tissue culture plates. After ~24 hours, the cells were transfected with 1 µg of DNA (0.5 µg of pcDNA6 and 0.5 µg of an ISG-expressing plasmid or pCMV-3Tag-8-Barr) using 3 µL of FuGENE HD in 100 µL of Opti-MEM. Approximately 24 hours post-transfection (day 1), the medium was changed with fresh DMEM. Blasticidin selection (10 µg/mL of blasticidin S HCl) began from day 4 post-transfection and the media containing blasticidin was replaced every three days until day 8-10. The resultant colonies were fixed using the fixation solution and stained with 0.1% (w/v) crystal violet. The resultant number of foci were counted and normalized to the resultant number of pcDNA6-transfected foci.

In retrotransposition assays using the *mEGFPI* retrotransposition indicator cassette, $2x10^5$ HeLa-JVM or HEK293T cells per well were seeded in 6-well tissue culture plates. On the next day, the cells were transfected with 1 µg of DNA (0.5 µg of cepB-gfp-L1.3 or cepB-gfp-L1.3RT[-] intronless and 0.5 µg of a pCMV-3Tag-8-Barr control or ISG-expressing plasmid) using 3 µL of FuGENE HD in 100 µL of Opti-MEM. Approximately 24 hours post-

transfection (day 1), the medium was replaced with fresh DMEM. Transfected cells were selected using 10 µg/mL blasticidin S HCl from day 2 post-transfection, changing the media every three days. The cells were collected on day 7-8 post-transfection and the resultant EGFP positive cells were analyzed using BD Accuri C6 Plus Software v.1.0.23.1 (BD Biosciences). The FITC channel was used to count the EGFP positive cells out of 30,000 cells. The number of the EGFP-positive cells was normalized to the transfection efficiency measured by counting the number of cepB-gfp-L1.3RT(-) intronless GFP-positive cells.

siRNA treatment

HeLa-JVM cells were plated in 6-well tissue culture plates at $1x10^5$ cells per well. After ~24 hours, 25 nM of a Dharmacon siRNA mixture (non-targeting control: ON-TARGETplus Non-targeting Pool, D-001810-10-0020; HELZ2: ON-TARGETplus HELZ2 siRNA SMARTpool, L-019109-00-0005; or MOV10: ON-TARGETplus MOV10 siRNA SMARTpool, L-014162-00-0005) were transfected using 3.75 µL of Lipofectamine RNAiMAX (Thermo Fisher Scientific, Waltham, MA, United States). Approximately 24 hours post-siRNA treatment (day 1), the medium was replaced with fresh DMEM and the cells were transfected with 0.5 µg of cepB-gfp-L1.3 or cepB-gfp-L1.3RT(-) intronless using 1.5 µL of FuGENE HD in 100 µL of Opti-MEM. Transfected cells were selected using 10 µg/mL blasticidin S HCl from day 3 post-transfection with media changes every three days. On day 8 post-transfection, the cells were harvested, washed with cold 1x PBS twice, and analyzed for EGFP expression using BD Accuri C6 Plus Flow Cytometer out of 30,000 cells. The number of the EGFP-positive cells was normalized to the transfection efficiency measured by counting the number of cepB-gfp-L1.3RT(-) intronless GFP-positive cells.

Immunoprecipitation of L1 ORF1p

Immunoprecipitation for IP-MS:

HeLa-JVM cells were plated in 15-cm tissue culture dishes containing DMEM medium at $2.5x10^6$ cells per dish. Three 15-cm tissue culture dishes were used for each sample preparation. After ~24 hours, the cells were transfected with 10 µg of an L1-expressing plasmid (pJM101/L1.3, pJM101/L1.3FLAG, or pALAF008) using 30 µL of FuGENE HD (Promega) in 1,000 µL of Opti-MEM. On the following day (day 1), the medium was replaced with fresh DMEM. From day 2 post-transfection onwards, the medium was replaced daily with fresh DMEM containing 100 µg/ml hygromycin B. On day 6 post-transfection, the cells were harvested using trypsin, washed with 1x cold PBS twice, flash-frozen with liquid nitrogen, and stored at -80°C.

For IP reactions, one hundred fifty microliters of Dynabeads Protein G (Invitrogen) was washed twice with PBS containing 0.5% (w/v) BSA and 0.1% (v/v) Triton X-100. For each sample, the beads were incubated with 15 µg of mouse monoclonal anti-FLAG M2 antibody (Sigma-Aldrich, F1804, RRID: AB_262044) in 1 mL of PBS containing 0.5% (w/v) BSA and 0.1% (v/v) Triton X-100 at 4°C for 2 hours. After incubation, the antibody-conjugated beads were washed with PBS containing 0.5% (w/v) BSA and 0.1% (v/v) Triton X-100 twice. The beads were resuspended in Lysis150 buffer (20 mM Tris-HCl [pH 7.5], 2.5 mM $MgCl_2$, 150 mM KCl, 0.5% (v/v) IGEPAL CA-630, 1 mM DTT) containing 0.2 mM phenylmethylsulfonyl fluoride (PMSF) and 1x cOmplete EDTA-free protease inhibitor cocktail before immunoprecipitation. Each cell pellet was lysed using the Lysis150 buffer containing 0.2 mM PMSF and 1x cOmplete EDTA-free protease inhibitor cocktail. The resuspended cell pellets were incubated at 4°C for 30 minutes and centrifuged at 12,000 x *g* for 5 minutes to pellet the cell debris. The supernatant was collected and incubated with antibody non-conjugated Dynabeads Protein G at 4°C for 2 hours with gentle rotation to remove non-specific protein binding. The Dynabeads were removed and the protein concentration in the pre-cleared cell lysates was quantified using Protein Assay Dye Reagent Concentrate. The same total amount of protein was used for each immunoprecipitation. Dynabeads Protein G conjugated

to the anti-FLAG antibody was added to the supernatant and incubated at 4°C for 3 hours with gentle rotation. The beads were then washed five times with 200 µL of the Lysis150 buffer. The ORF1p-FLAG protein complex bound was eluted using 200 µg/mL of 3xFLAG peptide (Sigma-Aldrich) in the Lysis150 buffer containing 0.2 mM PMSF and 1x cOmplete EDTA-free protease inhibitor cocktail by incubation at 4°C for 1 hour with gentle rotation. This step was repeated once, and the protein was precipitated overnight by adding three times the volume of cold acetone to the resultant eluate. The protein was pelleted at 12,000 x *g* at 4°C for 30 minutes, resuspended in 1x SDS sample buffer and boiled at 105°C for 5 minutes.

Immunoprecipitation for western blotting:

HEK293T cells were plated in 10-cm tissue culture dishes at $3x10^6$ cells per dish. Approximately 24 hours after plating, the cells were transfected with 4 µg of pJM101/L1.3FLAG or pJM101/L1.3 and 2 µg of ISG-expressing plasmid (pALAF015, pALAF016, pALAF021, pALAF022, pALAF023, or pALAF024) using 18 µL of 1 mg/mL PEI-MAX-40K in 500 µL of Opti-MEM. Approximately 24 hours post-transfection, the media was changed with fresh DMEM. From day 2 post-transfection onwards, the medium was replaced daily with fresh DMEM containing 100 µg/ml hygromycin B. On day 4 post-transfection, the cells were harvested with pipetting, washed with 1x cold PBS twice, flash-frozen with liquid nitrogen, and stored at -80°C for subsequent experiments.

For each sample, ten microliters of the Dynabeads Protein G were incubated with 1 µg of anti-FLAG M2 antibody in 50 µL of PBS containing 0.5% (w/v) BSA and 0.1% (v/v) Triton X-100 at 4°C for 2 hours. After incubation, the antibody-conjugated beads were washed with PBS containing 0.5% (w/v) BSA and 0.1% (v/v) Triton X-100 twice. The beads were resuspended in Lysis150 buffer containing 0.2 mM PMSF and 1x cOmplete EDTA-free protease inhibitor cocktail before immunoprecipitation. Each cell pellet was lysed in 500 µL of the Lysis150 buffer containing 0.2 mM PMSF and 1x cOmplete EDTA-free protease inhibitor cocktail. The resuspended cell pellets were incubated at 4°C for 1 hour and centrifuged at 12,000 x *g* for 5 minutes to pellet the cell debris. The supernatant was

collected and 10 µL of the supernatant was saved as input. Anti-FLAG antibody-conjugated Dynabeads were added to the samples and incubated at 4°C for 4 hours with gentle rotation.

The RNase treatment for HELZ2-expressed samples was performed after removal of the cell lysate using 20 µg/mL of RNase A (Nippongene, Tokyo, Japan) in 100 µL of the Lysis150 buffer for five minutes at 37°C. The beads then were washed four times with 100 µL of the Lysis150 buffer. The beads were resuspended directly in 1x SDS sample buffer and boiled at 105°C for 5 minutes except for the HELZ2-expressed samples, where the ORF1p-FLAG protein complex was eluted using 20 µL of the Lysis150 buffer containing 0.2 mM PMSF, 1x cOmplete EDTA-free protease inhibitor cocktail, and 200 µg/mL 3xFLAG peptide by incubation at 4°C for 1 hour with gentle rotation. The eluted protein was resuspended in 1x SDS sample buffer and boiled at 105°C for 5 minutes.

Label-free quantification (LFQ) of LC-MS/MS results

Mass spectrometry analysis was performed by the proteomics facility in the Graduate School of Biostudies at Kyoto University. After SDS-PAGE and visualization of the gel using PlusOne Silver Staining Kit, Protein (Cytiva) according to the protocol provided by the manufacturer, the entire gel lane from each sample was excised into 15 components. The silver stain was then removed, and the excised gel slices were incubated with sequencing-grade modified trypsin (Promega) to extract the peptides. The purified peptides then were subjected to liquid chromatography-tandem mass spectrometry (LC-MS/MS) on nano-Advance (AMR, Tokyo, Japan) and Q Exactive Plus (Thermo Fisher Scientific) using Xcalibur 3.1 (Thermo Fisher Scientific), Paradigm Home v.2.0.4 R4 B22 (Bruker Daltonics, Billerica, MA, United States), and Cycle Composer v.1.6.0 (CTC Analytics AG, Zwingen, Switzerland) for mass spectrometry acquisition. LFQ analyses on the resultant datasets were performed using Proteome Discoverer 2.3 (Thermo Fisher Scientific) with the peptide hits identified in the No Tag (pJM101/L1.3), WT (pJM101/L1.3FLAG), and M8 (RBM) (pALAF008) samples. Data are available via ProteomeXchange with identifier PXD038851. Briefly, the human UniProt Knowledgebase (UniProtKB:

https://www.uniprot.org/help/uniprotkb) database was used for protein identification and the Mascot Server 2.7.0 database (Matrix Science: https://www.matrixscience.com) was used as the search engine. The Protein Validator Node of Proteome Discoverer 2.3 calculated high (<0.01), medium (0.01≤ and <0.05), or low (0.05≤) false discovery rates (FDRs) with the peptide hits to generate the protein FDR confidence score. Both unique and razor peptides were used for identification of the best associated protein group with those peptides in the analysis. Razor peptides are shared in multiple protein groups and assigned to the protein group with the largest number of total peptides when combined with the unique peptides. Triplicate data from 15 gel strips of WT L1 and M8 (RBM) L1 protein lists were grouped respectively to obtain each group abundance using the Precursor Ions Quantifier nodes of Proteome Discoverer 2.3. The abundances were normalized with the ORF1p peptides. This analysis was followed by a comparison of the grouped abundances between WT L1 (grouped) *vs.* M8 (RBM) L1 (grouped) to calculate the abundance ratio, where the upper and lower limits of the ratios were set to 1000 and 0.001, respectively. The *p*-values of the abundance ratios were calculated using the Tukey Honestly Significant Difference test (post-hoc) after an analysis of variance (ANOVA) test. The volcano plot depicts the resultant $\log_2$ abundance ratios (WT ORF1p-FLAG *vs*. M8 ORF1p-FLAG) on the x-axis and the $-\log_{10}$ *p*-values of the abundance ratios on the y-axis. A threshold of >0.5 for $\log_2$ abundance ratios was used for the GO term analysis.

ORF1p crystal structure analysis

The crystal structure images of ORF1p and the mutations were created using UCSF ChimeraX software 1.2.5 for Windows (Pettersen et al., 2021) based on the 2ykp pdb file (Khazina et al., 2011).

GO term analysis

The proteins were first filtered before the analyses, followed by removal of protein groups with medium and low FDR confidence scores and those that lacked detectable peptide peaks in either the WT ORF1p-FLAG or M8 ORF1p-FLAG samples. The remaining proteins with UniProt accession numbers were converted to official gene symbols using the UniProt Retrieve/ID mapping tool (https://www.uniprot.org/id-mapping); the unmapped UniProt accession numbers were converted to gene symbols manually. UniProt accession numbers that do not map to any gene symbols were excluded from the analysis. In the case of different UniProt accession numbers that map to the same gene symbol, only the UniProt accession number with the highest $\log_2$ abundance ratio value (WT *vs*. M8 [RBM] L1) was included in the analysis. This filtration process resulted in a total number of 1437 genes (Source Data 2). Among the 1437 genes, genes with values of a >0.5 $\log_2$ abundance ratio (WT *vs*. M8 [RBM] L1) were used in the DAVID (Huang et al., 2009; Sherman et al., 2022) 2021 gene ontology (https://david.ncifcrf.gov/) analyses to obtain the "Functional annotations of UniProt Keyword Biological Processes" GO terms that are shown in **Fig. 3.2a** and **Table 2**.

GSEA preranked analysis

GSEA 4.2.3 for Windows software was used for the analysis (Subramanian et al., 2005) (http://www.broad.mit.edu/GSEA). The 1437 genes described in the GO term analysis paragraph were included in the GSEA preranked analysis using the $\log_2$ abundance ratios of WT *vs*. M8 (RBM) of the respective protein hits (Source Data 2). The GSEA preranked analysis was performed using the hallmark gene sets from GSEA Molecular Signatures Database v7.5.1 on Human Gene Symbol with Remapping v7.5 Chip platform.

ImageJ quantification of western blot band intensity

Using the ImageJ 1.5.2a for Windows software tool (Schneider et al., 2012), identical sized rectangles were drawn for each band. The area of intensity of the bands were generated using Plot Lanes function and calculated using a wand (tracing) tool. The intensity of each ORF1p-T7 band was normalized to that of the GAPDH band with respective samples. The values were displayed as ratios in comparison to the leftmost band in the western blot image (pTMF3 and pCMV-3Tag-8-Barr control co-transfected cells).

Bio-Plex cytokine assay

To collect culture supernatants, $2x10^5$ HEK293T cells per well were seeded in 6-well plates. Approximately 24 hours after seeding, the cells were transfected with 1 µg of an L1-expressing plasmid or pCEP4 using 3 µL of FuGENE HD transfection reagent and 100 µL of Opti-MEM according to the manufacturer's instruction. The media was replaced with fresh DMEM at ~24 hours (day 1) and ~72 hours (day 3) post-transfection. The culture supernatants were collected at ~96 hours (24 hours post-day 3 media change) and ~120 hours (48 hours post-day 3 media change) post-transfection. For the polyinosinic:polycytidylic acid (poly[I:C]) transfection, HEK293T cells were transfected with 5 µg/mL of High Molecular Weight Poly(I:C) (InvivoGen, San Diego, California, United States) using 3.75 µL of Lipofectamine RNAiMax (Thermo Fisher Scientific) in 1 mL of culture media in 6-well plates. The culture supernatants were collected at ~24 hours post poly(I:C) transfection. All culture supernatants were centrifuged at 500 x *g* for 5 minutes to remove cell debris, flash-frozen in liquid nitrogen, and stored at -80˚C. The same batch of DMEM was used in all cell cultures for this cytokine assay. Bio-Plex 200, a multiplex cytokine array system (Bio-Rad), was used to quantify the basal levels of cytokines in DMEM medium and the secreted cytokines and chemokines in the collected culture supernatants according to the protocol provided by the manufacturer. The Bio-Plex Pro Human Inflammation Panel 1 37-Plex includes 37 cytokines and chemokines (APRIL, BAFF, CD30, CD163, Chitinase-3, gp130, IFN-α2, IFN-β, IFN-γ, IL-2, IL-6Ra, IL-8, IL-10, IL-11, IL-12 (p40) , IL-12 (p70), IL-19,

IL-20, IL-22, IL-26, IL-27 (p28), IL-28A, IL-29, IL-32, IL-34, IL-35, LIGHT, MMP-1, MMP-2, MMP-3, Osteocalcin, Osteopontin, Pentraxin-3, TNF-R1, TNF-R2, TSLP, TWEAK). Data acquisition and analyses were performed using Bio-Plex Manager software version 5.0 (Bio-Rad).

RNA extraction and RT-qPCR

HeLa-JVM or HEK293T at $2x10^5$ cells per well were seeded in 6-well tissue culture plates. On the following day, the cells were transfected with 1 µg of DNA (1 µg of an L1-expressing plasmid or 0.5 µg of the L1-expressing plasmid and 0.5 µg of a pCMV-3Tag-8-Barr control or an ISG-expressing plasmid). Approximately 24 hours post-transfection (day 1), the medium was replaced with fresh DMEM. On day 2 (HeLa-JVM and HeLa-HA) or day 4 (HEK293T) post-transfection, the cells were washed with 1x PBS and 0.9 mL TRIzol was added directly to each well. The RNA extractions were performed according to the protocol provided by the manufacturer. The cells were lysed with TRIzol and transferred into new 1.5 mL tubes. One hundred eighty microliters of chloroform was added into each tube and shaken vigorously for 15 seconds. After incubation at room temperature for 5 minutes, the samples were centrifuged at 12,000 x *g* for 15 minutes at 4°C. Three hundred sixty microliters of the upper layer were transferred into a new 1.5 mL tube and 400 µL of 100% isopropanol was added to precipitate the RNA. The samples were incubated at room temperature for 10 minutes. Next, RNA was pelleted at 12,000 x *g* for 30 minutes. The purified RNA then was washed with 75% cold ethanol and centrifuged at 10,000 x *g* for 5 minutes. The RNA pellet was dried at room temperature. Once dried, 30 µL of RNase-free H_2O was added and incubated at 55°C for 10 minutes to dissolve RNA. The resultant RNA was then treated with RNase-free DNase Set (QIAGEN) according to the protocol provided by the manufacturer with some minor modifications. Five microliters of DNase I (15 K units, TaKaRa Bio), 0.2 U/µL of ribonuclease inhibitor (porcine liver) (TaKaRa Bio) in 44.5 µL of the RNase-free Buffer RDD was added to each sample. The samples were incubated at room temperature for 15 minutes and the RNA then was pelleted after ethanol precipitation

(incubation at -20°C overnight in 240 µL of 100% ethanol and 8 µL of 3M NaOAc [pH 5.2]). The RNA pellets were washed with 75% cold ethanol, dried at room temperature, resuspended in RNase-free water, and incubated at 75°C for 10 minutes to inactivate the DNase I. One microgram of total RNA was used as a template in reverse transcription reactions using 0.2 mM dNTP (TakaRa Bio), 1 U/µL ribonuclease inhibitor (porcine liver) (TaKaRa Bio), 0.25 U/µL AMV reverse transcriptase XL (TaKaRa Bio), and 0.125 µM of an oligo (dT) primer (Invitrogen) according to the protocol provided by the manufacturer unless stated otherwise. Two negative controls were included for all instances: no reverse transcriptase (reverse transcriptase was excluded during cDNA synthesis) and no template (cDNA was replaced with RNase-free water). The reverse transcription reaction was performed as follows: 30°C for 10 minutes, 42°C for 30 minutes, and 95°C for 5 minutes. Prime Script MMLV reverse transcriptase (TaKaRa Bio) and 0.125 µM of the oligo (dT) primer for RNA-IP experiments (see below) or a HELZ2 specific primer (HELZ2_R) for HELZ2 RNA quantification were used to reverse transcribe instead. RNA was incubated at 65°C for 5 minutes before the addition of Prime Script MMLV reverse transcriptase and the reverse transcription was performed as follows: 42°C for 60 minutes followed by 70°C for 15 minutes. RT-qPCR was performed using Luna Universal qPCR Master Mix (New England Biolabs). Amplification was performed using StepOnePlus Real-Time PCR System (Applied Biosystems) using the following parameters: 15 seconds at 95°C; followed by 40 cycles of denaturation (95°C for 15 seconds) and amplification (60°C for 60 seconds). Technical duplicates were made for each sample. Quantification of cDNA for each reaction was determined by comparing the cycle threshold (Ct) with a standard curve generated from one of the samples using StepOne Software v2.2. All Ct readings fall within the range of the standard curve generated.

Primers used for RT-qPCR:

HLTF_F: 5'-GTGCATGCTGCAGTACAGA-3'

HLTF_R: 5'-GCTGTTCCCAGAATGGTGGA-3'

SMC2_F: 5'-GCTTTTTGCTGGGCATCTCC-3'

SMC2_R: 5’-ACCAGCCTGCCCATTTTTGT-3’

L1 (SV40)_F: 5’-TCCAGACATGATAAGATACATTGATGAG-3’

L1 (SV40)_R: 5’-GCAATAGCATCACAAATTTCACAAA-3’

Luciferase_F: 5’-CGAGGCTACAAACGCTCTCA-3’

Luciferase_R: 5’-CAGGATGCTCTCCAGTTCGG-3’

IFN-α _F: 5’-CTGAATGACTTGGAAGCCTG-3’

IFN-α _R: 5’-ATTTCTGCTCTGACAACCTC-3’

HELZ2_F: 5’-GAGAAGGTGGTTCTTCTCGGAG-3’

HELZ2_R: 5’-CTCATGCATGCGGTACTGAG-3’

MOV10_F: 5’-CGTACCGGAAACAGGTGGAG-3’

MOV10_R: 5’- TGAACCCACCTTCAAGTCCTTG-3’

mneoI (Alu or L1)_F: 5’- ACCGGACAGGTCGGTCTTG-3’

mneoI (Alu or L1)_R: 5’- CTGGGCACAACAGACAATCG-3’

Beta-actin_F: 5’-CCTTTTTTGTCCCCCAACTTG-3’

Beta-actin_R: 5’-TGGCTGCCTCCACCCA-3’

GAPDH_F: 5’-GGAGTCCCTGCCACACTCAG-3’

GAPDH_R: 5’-GGTCTACATGGCAACTGTGAGG-3’

Oligo (dT): 5’-TTTTTTTTTTTTTTTTTTTTVN-3’

RNA-IP

RNA immunoprecipitation (RNA-IP) experiments were carried out as described previously with some modifications (Doucet et al., 2015). HeLa-JVM cells were plated in 10-cm tissue culture dishes at $1.5x10^6$ cells per dish. On the following day (day 0), the cells were transfected with 5 µg of plasmid DNA (pJM101/L1.3, pJM101/L1.3FLAG or pALAF008_M8) using 15 µL of PEI-MAX-40K in 500 µL of Opti-MEM. Approximately 24 hours post-transfection (day 1), the medium was replaced with fresh DMEM. On the following day (day 2), the medium was replaced daily with fresh DMEM containing 100 µg/mL hygromycin B and the cells were collected at day 5 post-transfection. The whole cell

extracts were prepared by incubation in the Lysis150 buffer containing 0.2 mM PMSF and 1x cOmplete EDTA-free protease inhibitor cocktail for one hour at 4°C. The lysate was separated from the insoluble fraction by centrifugation at 12, 000 x *g* for five minutes and transferred to a new tube. Ten microliters of the lysate were saved as the input fraction. Prior to immunoprecipitation, the anti-FLAG antibody-conjugated beads were prepared as described in "immunoprecipitation and western blotting" section of the Methods. The cleared lysate (input) was incubated with the anti-FLAG antibody-conjugated beads for 5 hours at 4°C. The beads were then washed four times with 150 µL of Lysis150 buffer without protease inhibitors. The RNA extraction was performed as described in "RNA extraction and RT-qPCR" in the Methods section with a slight modification: 200 µg/mL glycogen was added to the immunoprecipitated RNA fraction before ethanol precipitation. All of the RNA samples were resuspended in 30 µL of RNase-free water. Five microliters (one sixth) of the extracted RNA from the input and IP fractions were used to synthesize cDNA using PrimeScript MMLV reverse transcriptase as described in the previous section. The ORF1p-associated RNA values were calculated by dividing the cDNA amount in the IP fraction by that in the input.

In vitro **RNase assay**

The RNase assay was performed based on RNase II assay by Barbas A., et. al. (Barbas et al., 2008) with several modifications. To produce HELZ2-3xFLAG proteins, HEK293T cells were seeded on two 10-cm dishes at ~$5x10^6$ cells per dish in DMEM. Approximately 24 hours after cell seeding, the cells were transfected with 10 µg of plasmid pCEP4, pALAF071 (HELZ2-3xFLAG), or pALAF073 (HELZ2-3xFLAG_dRNase), which was preincubated in 500 µL of Opti-MEM with 30 µL of PEI-MAX-40K (1 mg/mL) for 10 minutes at room temperature. The medium was replaced with fresh DMEM ~24 hours post-transfection. Approximately 48 hours post-transfection, the transfected cells were collected, washed with 1x cold PBS, flash-frozen with liquid nitrogen, and stored at -80˚C.

To purify the recombinant proteins, 20 µL of Dynabeads Protein G was washed twice with PBS containing 0.5% (w/v) BSA and 0.1% (v/v) Triton X-100 followed by conjugation

with 2 µg of mouse monoclonal anti-FLAG M2 antibody (Sigma-Aldrich, F1804) in PBS containing 0.5% (w/v) BSA and 0.1% (v/v) Triton X-100 at 4°C for 1 hour. After conjugation, the beads were washed with PBS containing 0.5% (w/v) BSA and 0.1% (v/v) Triton X-100 twice and resuspended in RIPA buffer (10 mM Tris-HCl [pH 7.5], 1 mM EDTA, 1% [v/v] TritonX-100, 0.1% [w/v] sodium deoxycholate, 0.1% SDS [w/v], 140 mM NaCl). Each frozen cell pellet was lysed using 1 mL RIPA buffer containing 0.2 mM PMSF and 1x cOmplete EDTA-free protease inhibitor cocktail. The resuspended cell pellets were incubated at 4°C for 1 hour and centrifuged at 12,000 x *g* for 5 minutes to pellet the cell debris. The supernatant (~1 mL) was collected and incubated with anti-FLAG antibody-conjugated Dynabeads Protein G at 4°C for 3 hours with gentle rotation. The beads were then washed five times with 200 µL of RIPA buffer. The HELZ2-3xFLAG protein was eluted using 20 µL of 200 µg/mL of 3xFLAG peptide (Sigma-Aldrich) in RIPA buffer containing 0.2 mM PMSF and 1x cOmplete EDTA-free protease inhibitor cocktail by incubation at 4°C for 1 hour with gentle rotation. The eluate fraction (~20 µL) was mixed with 40 µL of TBST (20 mMTris-HCl [pH 8], 150 mM NaCl, 0.05% [v/v] Tween 20) containing 90% glycerol, subjected to SDS-PAGE, visualized by PlusOne Silver Staining Kit, and analyzed by western blotting using an anti-FLAG antibody (Sigma-Aldrich, F1804).

For the detection of 3’ to 5’ RNase activity, 2 µL of the purified HELZ2-3xFLAG or HELZ2 3xFLAG_dRNase were incubated with a single-strand poly$(A)_{30}$ RNA oligonucleotide labeled with IRDye800 at its 5’ end (poly[rA30], Integrated DNA Technologies [IDT]) for 0, 5, 10, and 60 minutes at 37˚C in 50 µL of the RNase buffer (20 mM Tris-HCl [pH 8], 100 mM KCl, 1 mM MgCl2, 1 mM DTT, 6 nM poly[rA30]). Only the labeled probe without recombinant protein served as a control to indicate the full-length poly$(A)_{30}$ RNA. Fifty microliters of 2x RNA loading dye (47.5% [v/v] formamide, 20 mM EDTA, 0.1% [w/v] Orange G) were added to the reaction and the resultant single-stranded (ss) RNA products were separated in a 5% acrylamide/TBE gel (45 mM Tris-HCl [pH 7.5], 45 mM boric acid, 2 mM EDTA, 5% acrylamide/bisacrylamide [37.5:1]). The gel image was captured by an Odyssey CLx imaging system (LI-COR Biosciences, Lincoln, NE, United States).

Statistics and reproducibility

All western blots and immunofluorescence were independently replicated three times to ensure reproducibility. The RNase assay experiment in **Fig. 4.3c** was performed twice with similar results observed. The rest of the experimental replicate numbers are indicated in the figure legends. One-way ANOVA followed by Bonferroni-Holm post hoc tests were performed for all statistical analyses unless stated otherwise in the figure legends. All analyses were performed using online website statistical calculator ASTATSA 2016 (https://www.astatsa.com/) or GraphPad Prism version 9.0.0 for Windows (GraphPad Software, San Diego, California, United States; www.graphpad.com). The numbers of biological replicates are indicated in the figure legends. Where applicable, data are always shown as the mean ± standard errors of the means (SEM). The exact *p*-value of each pair was indicated in the figure legends. ns: not significant; * $p<0.05$; ** $p<0.01$; *** $p<0.001$).

References

Adney, E.M., Ochmann, M.T., Sil, S., Truong, D.M., Mita, P., Wang, X., Kahler, D.J., Fenyö, D., Holt, L.J., and Boeke, J.D. (2019) Comprehensive Scanning Mutagenesis of Human Retrotransposon LINE-1 Identifies Motifs Essential for Function. Genetics **213**, 1401–1414.

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

Alexandrova, E.A., Olovnikov, I.A., Malakhova, G.V., Zabolotneva, A.A., Suntsova, M.V., Dmitriev, S.E., and Buzdin, A.A. (2012) Sense transcripts originated from an internal part of the human retrotransposon LINE-1 5′ UTR. Gene **511**, 46–53.

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.

Amblar, M., Barbas, A., Fialho, A.M., and Arraiano, C.M. (2006) Characterization of the Functional Domains of Escherichia coli RNase II. J. Mol. Biol. **360**, 921–933.

Anderson, P., and Kedersha, N. (2002) Stressful initiations. J. Cell Sci. **115**, 3227–3234.

Ardeljan, D., Steranka, J.P., Liu, C., Li, Z., Taylor, M.S., Payer, L.M., Gorbounov, M., Sarnecki, J.S., Deshpande, V., Hruban, R.H., et al. (2020) Cell fitness screens reveal a conflict between LINE-1 retrotransposition and DNA replication. Nat. Struct. Mol. Biol. **27**, 168–178.

Ariumi, Y. (2016) Guardian of the Human Genome: Host Defense Mechanisms against LINE-1 Retrotransposition. Front. Chem. **4**, 28.

Arjan-Odedra, S., Swanson, C.M., Sherer, N.M., Wolinsky, S.M., and Malim, M.H. (2012) Endogenous MOV10 inhibits the retrotransposition of endogenous retroelements but not the replication of exogenous retroviruses. Retrovirology **9**, 53.

Athanasiadis, A., Rich, A., and Maas, S. (2004) Widespread A-to-I RNA Editing of Alu-Containing mRNAs in the Human Transcriptome. PLOS Biol. **2**, e391.

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.

Awano, N., Rajagopal, V., Arbing, M., Patel, S., Hunt, J., Inouye, M., and Phadtare, S. (2010) Escherichia coli RNase R Has Dual Activities, Helicase and RNase. J. Bacteriol. **192**, 1344–1352.

Balaj, L., Lessard, R., Dai, L., Cho, Y.-J., Pomeroy, S.L., Breakefield, X.O., and Skog, J. (2011) Tumour microvesicles contain retrotransposon elements and amplified oncogene sequences. Nat. Commun. **2**, 180.

Bannert, N., Hofmann, H., Block, A., and Hohn, O. (2018) HERVs New Role in Cancer: From Accused Perpetrators to Cheerful Protectors. Front. Microbiol. **9**, 178.

Barbas, A., Matos, R.G., Amblar, M., López-Viñas, E., Gomez-Puertas, P., and Arraiano, C.M. (2008) New insights into the mechanism of RNA degradation by ribonuclease II:

identification of the residue responsible for setting the RNase II end product. J. Biol. Chem. **283**, 13070–13076.

Bartsch, K., Knittler, K., Borowski, C., Rudnik, S., Damme, M., Aden, K., Spehlmann, M.E., Frey, N., Saftig, P., Chalaris, A., et al. (2017) Absence of RNase H2 triggers generation of immunogenic micronuclei removed by autophagy. Hum. Mol. Genet. **26**, 3960–3972.

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., 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.

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

Benitez-Guijarro, M., Lopez-Ruiz, C., Tarnauskaitė, Ž., Murina, O., Mian Mohammad, M., Williams, T.C., Fluteau, A., Sanchez, L., Vilar-Astasio, R., Garcia-Canadas, M., et al. (2018) RNase H2, mutated in Aicardi-Goutières syndrome, promotes LINE-1 retrotransposition. EMBO J. **37**, e98506.

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

Bogerd, H.P., Wiegand, H.L., Hulme, A.E., Garcia-Perez, J.L., O'Shea, K.S., Moran, J.V., and Cullen, B.R. (2006b) Cellular inhibitors of long interspersed element 1 and Alu retrotransposition. Proc. Natl. Acad. Sci. U. S. A. **103**, 8780–8785.

Boissinot, S., Entezam, A., Young, L., Munson, P.J., and Furano, A.V. (2004) The insertional history of an active family of L1 retrotransposons in humans. Genome Res. **14**, 1221–1231.

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

Brégnard, C., Guerra, J., Déjardin, S., Passalacqua, F., Benkirane, M., and Laguette, N. (2016) Upregulated LINE-1 Activity in the Fanconi Anemia Cancer Susceptibility Syndrome Leads to Spontaneous Pro-inflammatory Cytokine Production. EBioMedicine **8**, 184–194.

Briggs, E.M., McKerrow, W., Mita, P., Boeke, J.D., Logan, S.K., and Fenyö, D. (2021) RIP-seq reveals LINE-1 ORF1p association with p-body enriched mRNAs. Mob. DNA **12**, 5.

Britten, R.J., and Kohne, D.E. (1968) Repeated sequences in DNA. Hundreds of thousands of copies of DNA sequences have been incorporated into the genomes of higher organisms. Science **161**, 529–540.

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

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. Mech. Dis. **15**, 51–70.

Carter, V., LaCava, J., Taylor, M.S., Liang, S.Y., Mustelin, C., Ukadike, K.C., Bengtsson, A., Lood, C., and Mustelin, T. (2020) High prevalence and disease correlation of autoantibodies against p40 encoded by long interspersed nuclear elements (LINE-1) in systemic lupus erythematosus. Arthritis Rheumatol. Hoboken NJ **72**, 89–99.

Cecco, M.D., 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) LINE-1 derepression in senescent cells triggers interferon and inflammaging. Nature **566**, 73–78.

Chen, H., Lilley, C.E., Yu, Q., Lee, D.V., Chou, J., Narvaiza, I., Landau, N.R., and Weitzman, M.D. (2006) APOBEC3A Is a Potent Inhibitor of Adeno-Associated Virus and Retrotransposons. Curr. Biol. **16**, 480–485.

Chiappinelli, K.B., Strissel, P.L., Desrichard, A., Li, H., Henke, C., Akman, B., Hein, A., Rote, N.S., Cope, L.M., Snyder, A., et al. (2015) Inhibiting DNA methylation causes an interferon response in cancer via dsRNA including endogenous retroviruses. Cell **162**, 974–986.

Chiu, Y.-L., Witkowska, H.E., Hall, S.C., Santiago, M., Soros, V.B., Esnault, C., Heidmann, T., and Greene, W.C. (2006) High-molecular-mass APOBEC3G complexes restrict Alu retrotransposition. Proc. Natl. Acad. Sci. U. S. A. **103**, 15588–15593.

Choi, J., Hwang, S.-Y., and Ahn, K. (2018) Interplay between RNASEH2 and MOV10 controls LINE-1 retrotransposition. Nucleic Acids Res. **46**, 1912–1926.

Chu, L.-Y., Hsieh, T.-J., Golzarroshan, B., Chen, Y.-P., Agrawal, S., and Yuan, H.S. (2017) Structural insights into RNA unwinding and degradation by RNase R. Nucleic Acids Res. **45**, 12015–12024.

Chung, H., Calis, J.J.A., Wu, X., Sun, T., Yu, Y., Sarbanes, S.L., Dao Thi, V.L., Shilvock, A.R., Hoffmann, H.-H., Rosenberg, B.R., et al. (2018) Human ADAR1 Prevents Endogenous RNA from Triggering Translational Shutdown. Cell **172**, 811-824.e14.

Cohen, L., Cui, N., Cai, Y., Garden, P.M., Li, X., Weitz, D.A., and Walt, D.R. (2020) Single Molecule Protein Detection with Attomolar Sensitivity Using Droplet Digital Enzyme-Linked Immunosorbent Assay. ACS Nano **14**, 9491–9501.

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.

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.

Crow, M.K. (2010) Long interspersed nuclear elements (LINE-1): Potential triggers of systemic autoimmune disease. Autoimmunity **43**, 7–16.

Crow, M.K. (2020) Reactivity of IgG With the p40 Protein Encoded by the Long Interspersed Nuclear Element 1 Retroelement: Comment on the Article by Carter et al. Arthritis Rheumatol. **72**, 374–376.

Crow, M.K., Olferiev, M., and Kirou, K.A. (2019) Type I Interferons in Autoimmune Disease. Annu. Rev. Pathol. Mech. Dis. **14**, 369–393.

Crow, Y.J., Chase, D.S., Schmidt, J.L., Szynkiewicz, M., Forte, G.M.A., Gornall, H.L., Oojageer, A., Anderson, B., Pizzino, A., Helman, G., et al. (2015) Characterization of Human Disease Phenotypes Associated with Mutations in TREX1, RNASEH2A, RNASEH2B, RNASEH2C, SAMHD1, ADAR, and IFIH1. Am. J. Med. Genet. A. **0**, 296–312.

Cuellar, T.L., Herzner, A.-M., Zhang, X., Goyal, Y., Watanabe, C., Friedman, B.A., Janakiraman, V., Durinck, S., Stinson, J., Arnott, D., et al. (2017) ---Silencing of retrotransposons by SETDB1 inhibits the interferon response in acute myeloid leukemia--. J. Cell Biol. **216**, 3535–3549.

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.

Deininger, P., Morales, M.E., White, T.B., Baddoo, M., Hedges, D.J., Servant, G., Srivastav, S., Smither, M.E., Concha, M., DeHaro, D.L., et al. (2017) A comprehensive approach to expression of L1 loci. Nucleic Acids Res. **45**, e31.

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

Dmitriev, S.E., Andreev, D.E., Terenin, I.M., Olovnikov, I.A., Prassolov, V.S., Merrick, W.C., and Shatsky, I.N. (2007) Efficient translation initiation directed by the 900-nucleotide-long and GC-rich 5' untranslated region of the human retrotransposon LINE-1 mRNA is strictly cap dependent rather than internal ribosome entry site mediated. Mol. Cell. Biol. **27**, 4685–4697.

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.

Dombroski, B.A., Feng, Q., Mathias, S.L., Sassaman, D.M., Scott, A.F., Kazazian, H.H., 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.

Dong, B., and Silverman, R.H. (1995) 2-5A-dependent RNase Molecules Dimerize during Activation by 2-5A. J. Biol. Chem. **270**, 4133–4137.

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

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., Rodić, N., Sharma, R., Darbari, I., Abril, G., Choi, J.A., Young Ahn, J., Cheng, Y., Anders, R.A., Burns, K.H., et al. (2015) LINE-1 expression and retrotransposition in Barrett's esophagus and esophageal carcinoma. Proc. Natl. Acad. Sci. U. S. A. **112**, E4894-4900.

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

Ewing, A.D., and Kazazian, H.H. (2010) High-throughput sequencing reveals extensive variation in human-specific L1 content in individual human genomes. Genome Res. **20**, 1262–1270.

Ewing, A.D., Smits, N., Sanchez-Luque, F.J., Faivre, J., Brennan, P.M., Richardson, S.R., Cheetham, S.W., and Faulkner, G.J. (2020) Nanopore Sequencing Enables Comprehensive Transposable Element Epigenomic Profiling. Mol. Cell **80**, 915-928.e5.

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., and Boeke, J.D. (1996) Human L1 Retrotransposon Encodes a Conserved Endonuclease Required for Retrotransposition. Cell **87**, 905–916.

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.

Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Heger, A., Hetherington, K., Holm, L., Mistry, J., et al. (2014) Pfam: the protein families database. Nucleic Acids Res. **42**, D222–D230.

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

Frassinelli, L., Orecchini, E., Al-Wardat, S., Tripodi, M., Mancone, C., Doria, M., Galardi, S., Ciafrè, S.A., and Michienzi, A. (2021) The RNA editing enzyme ADAR2 restricts L1 mobility. RNA Biol. **18**, 75–87.

Frazão, C., McVey, C.E., Amblar, M., Barbas, A., Vonrhein, C., Arraiano, C.M., and Carrondo, M.A. (2006) Unravelling the dynamics of RNA degradation by ribonuclease II and its RNA-bound complex. Nature **443**, 110–114.

Friedli, M., and Trono, D. (2015) The Developmental Control of Transposable Elements and the Evolution of Higher Species. Annu. Rev. Cell Dev. Biol. **31**, 429–451.

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**, e2022751118.

Gallois-Montbrun, S., Kramer, B., Swanson, C.M., Byers, H., Lynham, S., Ward, M., and Malim, M.H. (2007) Antiviral Protein APOBEC3G Localizes to Ribonucleoprotein Complexes Found in P Bodies and Stress Granules. J. Virol. **81**, 2165–2178.

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

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.

Gasior, S.L., Wakeman, T.P., Xu, B., and Deininger, P.L. (2006) The Human LINE-1 Retrotransposon Creates DNA Double-strand Breaks. J. Mol. Biol. **357**, 1383–1393.

George, C.X., John, L., and Samuel, C.E. (2014) An RNA Editor, Adenosine Deaminase Acting on Double-Stranded RNA (ADAR1). J. Interferon Cytokine Res. **34**, 437–446.

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

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.

Goldstone, D.C., Ennis-Adeniran, V., Hedden, J.J., Groom, H.C.T., Rice, G.I., Christodoulou, E., Walker, P.A., Kelly, G., Haire, L.F., Yap, M.W., et al. (2011) HIV-1 restriction factor SAMHD1 is a deoxynucleoside triphosphate triphosphohydrolase. Nature **480**, 379–382.

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

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

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

Goodier, J.L., Cheung, L.E., and Kazazian, H.H. (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. (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. (2015) The Broad-Spectrum Antiviral Protein ZAP Restricts Human Retrotransposition. PLoS Genet. **11**.

Grandi, N., and Tramontano, E. (2018) Human Endogenous Retroviruses Are Ancient Acquired Elements Still Shaping Innate Immune Responses. Front. Immunol. **9**, 2039.

Griffin, G.K., Wu, J., Iracheta-Vellve, A., Patti, J.C., Hsu, J., Davis, T., Dele-Oni, D., Du, P.P., Halawi, A.G., Ishizuka, J.J., et al. (2021) Epigenetic silencing by SETDB1 suppresses tumour intrinsic immunogenicity. Nature **595**, 309–314.

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., Martin, L., Ware, C., Blish, C.A., Chang, H.Y., et al. (2015) Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. Nature **522**, 221–225.

Hancks, D.C., and Kazazian, H.H. (2010) SVA retrotransposons: Evolution and genetic instability. Semin. Cancer Biol. **20**, 234–245.

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

Hancks, D.C., and Kazazian, H.H., Jr. (2012) Active human retrotransposons: variation and disease. Curr. Opin. Genet. Dev. **22**, 191–203.

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

Harris, R.S., Bishop, K.N., Sheehy, A.M., Craig, H.M., Petersen-Mahrt, S.K., Watt, I.N., Neuberger, M.S., and Malim, M.H. (2003) DNA Deamination Mediates Innate Immunity to Retroviral Infection. Cell **113**, 803–809.

Helman, E., Lawrence, M.S., Stewart, C., Sougnez, C., Getz, G., and Meyerson, M. (2014) Somatic retrotransposition in human cancer revealed by whole-genome and exome sequencing. Genome Res. **24**, 1053–1063.

Herrmann, A., Wittmann, S., Thomas, D., Shepard, C.N., Kim, B., Ferreirós, N., and Gramberg, T. (2018) The SAMHD1-mediated block of LINE-1 retroelements is regulated by phosphorylation. Mob. DNA **9**, 11.

Hickman, A.B., and Dyda, F. (2015) Mechanisms of DNA Transposition. Microbiol. Spectr. **3**, MDNA3-0034–2014.

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

Hopfner, K.-P., and Hornung, V. (2020) Molecular mechanisms and cellular functions of cGAS–STING signalling. Nat. Rev. Mol. Cell Biol. **21**, 501–521.

Horn, A.V., Klawitter, S., Held, U., Berger, A., Jaguva Vasudevan, A.A., Bock, A., Hofmann, H., Hanschmann, K.-M.O., Trösemeier, J.-H., Flory, E., et al. (2014) Human LINE-1 restriction by APOBEC3C is deaminase independent and mediated by an ORF1p

interaction that affects LINE reverse transcriptase activity. Nucleic Acids Res. **42**, 396–416.

Hossain, S.T., Malhotra, A., and Deutscher, M.P. (2015) The Helicase Activity of Ribonuclease R Is Essential for Efficient Nuclease Activity. J. Biol. Chem. **290**, 15697–15706.

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, C.R., Schneider, A.M., Lu, Y., Niranjan, T., Shen, P., Robinson, M.A., Steranka, J.P., Valle, D., Civin, C.I., Wang, T., et al. (2010) Mobile interspersed repeats are major structural variants in the human genome. Cell **141**, 1171–1182.

Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. **4**, 44–57.

Hubbard, N.W., Ames, J.M., Maurano, M., Chu, L.H., Somfleth, K.Y., Gokhale, N.S., Werner, M., Snyder, J.M., Lichauco, K., Savan, R., et al. (2022) ADAR1 mutation causes ZBP1-dependent immunopathology. Nature **607**, 769–775.

Hulme, A.E., Bogerd, H.P., Cullen, B.R., and Moran, J.V. (2007) Selective inhibition of Alu retrotransposition by APOBEC3G. Gene **390**, 199–205.

Hung, T., Pratt, G.A., Sundararaman, B., Townsend, M.J., Chaivorapol, C., Bhangale, T., Graham, R.R., Ortmann, W., Criswell, L.A., Yeo, G.W., et al. (2015) The Ro60 autoantigen binds endogenous retroelements and regulates inflammatory gene expression. Science **350**, 455–459.

van den Hurk, J.A.J.M., Meij, I.C., Seleme, M. del 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.

Ikeda, T., Abd El Galil, K.H., Tokunaga, K., Maeda, K., Sata, T., Sakaguchi, N., Heidmann, T., and Koito, A. (2011) Intrinsic restriction activity by apolipoprotein B mRNA editing enzyme APOBEC1 against the mobility of autonomous retrotransposons. Nucleic Acids Res. **39**, 5538–5554.

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.

Ivashkiv, L.B., and Donlin, L.T. (2014) Regulation of type I interferon responses. Nat. Rev. Immunol. **14**, 36–49.

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

Jacobs, F.M.J., 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.

Janeway, C.A. (1992) The immune system evolved to discriminate infectious nonself from noninfectious self. Immunol. Today **13**, 11–16.

Jones, P.A., Ohtani, H., Chakravarthy, A., and De Carvalho, D.D. (2019) Epigenetic therapy in immune-oncology. Nat. Rev. Cancer **19**, 151–161.

Jurka, J., Bao, W., and Kojima, K.K. (2011) Families of transposable elements, population structure and the origin of species. Biol. Direct **6**, 44.

Kaikkonen, M.U., and Adelman, K. (2018) Emerging Roles of Non-Coding RNA Transcription. Trends Biochem. Sci. **43**, 654–667.

Kammerer, R.A., Kostrewa, D., Progias, P., Honnappa, S., Avila, D., Lustig, A., Winkler, F.K., Pieters, J., and Steinmetz, M.O. (2005) A conserved trimerization motif controls the topology of short coiled coils. Proc. Natl. Acad. Sci. **102**, 13891–13896.

Kaneko-Ishino, T., and Ishino, F. (2010) Retrotransposon silencing by DNA methylation contributed to the evolution of placentation and genomic imprinting in mammals. Dev. Growth Differ. **52**, 533–543.

Kaneko-Ishino, T., and Ishino, F. (2012) The role of genes domesticated from LTR retrotransposons and retroviruses in mammals. Front. Microbiol. **3**, 262.

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

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.

Katano-Toki, A., Satoh, T., Tomaru, T., Yoshino, S., Ishizuka, T., Ishii, S., Ozawa, A., Shibusawa, N., Tsuchiya, T., Saito, T., et al. (2013) THRAP3 Interacts with HELZ2 and Plays a Novel Role in Adipocyte Differentiation. Mol. Endocrinol. **27**, 769–780.

Kawamura, Y., Sanchez Calle, A., Yamamoto, Y., Sato, T.-A., and Ochiya, T. (2019) Extracellular vesicles mediate the horizontal transfer of an active LINE-1 retrotransposon. J. Extracell. Vesicles **8**, 1643214.

Kazazian, H.H. (2011) Mobile DNA: Finding Treasure in Junk (FT Press).

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

Kazazian, H.H., 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.

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

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**, e34960.

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

Kinomoto, M., Kanno, T., Shimura, M., Ishizaka, Y., Kojima, A., Kurata, T., Sata, T., and Tokunaga, K. (2007) All APOBEC3 family proteins differentially inhibit LINE-1 retrotransposition. Nucleic Acids Res. **35**, 2955–2964.

Kolosha, V.O., and Martin, S.L. (1997) In vitro properties of the first ORF protein from mouse LINE-1 support its role in ribonucleoprotein particle formation during retrotransposition. Proc. Natl. Acad. Sci. U. S. A. **94**, 10155–10160.

Kolosha, V.O., and Martin, S.L. (2003) High-affinity, non-sequence-specific RNA binding by the open reading frame 1 (ORF1) protein from long interspersed nuclear element 1 (LINE-1). J. Biol. Chem. **278**, 8112–8117.

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. **108**, 20345–20350.

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. Clifton NJ **1400**, 139–156.

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

Kubo, S., Seleme, M.D.C., Soifer, H.S., Perez, J.L.G., Moran, J.V., Kazazian, H.H., and Kasahara, N. (2006) L1 retrotransposition in nondividing and primary human somatic cells. Proc. Natl. Acad. Sci. 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. (2011) Initial impact of the sequencing of the human genome. Nature **470**, 187–197.

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., 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.

Levanon, E.Y., Eisenberg, E., Yelin, R., Nemzer, S., Hallegger, M., Shemesh, R., Fligelman, Z.Y., Shoshan, A., Pollock, S.R., Sztybel, D., et al. (2004) Systematic identification of

abundant A-to-I editing sites in the human transcriptome. Nat. Biotechnol. **22**, 1001–1005.

Li, P., Du, J., Goodier, J.L., Hou, J., Kang, J., Kazazian, H.H., Zhao, K., and Yu, X.-F. (2017) Aicardi–Goutières syndrome protein TREX1 suppresses L1 and maintains genome integrity through exonuclease-independent ORF1p depletion. Nucleic Acids Res. **45**, 4619–4631.

Li, X., Zhang, J., Jia, R., Cheng, V., Xu, X., Qiao, W., Guo, F., Liang, C., and Cen, S. (2013) The MOV10 Helicase Inhibits LINE-1 Mobility. J. Biol. Chem. **288**, 21148–21160.

Lindič, N., Budič, M., Petan, T., Knisbacher, B.A., Levanon, E.Y., and Lovšin, N. (2013) Differential inhibition of LINE1 and LINE2 retrotransposition by vertebrate AID/APOBEC proteins. Retrovirology **10**, 156.

Liu, C., Zhang, Y., Liu, C.C., and Schatz, D.G. (2022) Structural insights into the evolution of the RAG recombinase. Nat. Rev. Immunol. **22**, 353–370.

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.

Lovšin, N., and Peterlin, B.M. (2009) APOBEC3 Proteins Inhibit LINE-1 Retrotransposition in the Absence of ORF1p Binding. Ann. N. Y. Acad. Sci. **1178**, 268–275.

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.

Luqman-Fatah, A., and Miyoshi, T. (2022) Human LINE-1 retrotransposons: impacts on the genome and regulation by host factors. Genes Genet. Syst. **advpub**, 22–00038.

MacDuff, D.A., Demorest, Z.L., and Harris, R.S. (2009) AID can restrict L1 retrotransposition suggesting a dual role in innate and adaptive immunity. Nucleic Acids Res. **37**, 1854–1867.

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

Maharana, S., Wang, J., Papadopoulos, D.K., Richter, D., Pozniakovsky, A., Poser, I., Bickle, M., Rizk, S., Guillén-Boixet, J., Franzmann, T.M., et al. (2018) RNA buffers the phase separation behavior of prion-like RNA binding proteins. Science **360**, 918–921.

Maharana, S., Kretschmer, S., Hunger, S., Yan, X., Kuster, D., Traikov, S., Zillinger, T., Gentzel, M., Elangovan, S., Dasgupta, P., et al. (2022) SAMHD1 controls innate immunity by regulating condensation of immunogenic self RNA. Mol. Cell **82**, 3712-3728.e10.

Mandal, P.K., Ewing, A.D., Hancks, D.C., and Kazazian, H.H. (2013) Enrichment of processed pseudogene transcripts in L1-ribonucleoprotein particles. Hum. Mol. Genet. **22**, 3730–3748.

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

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., 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., 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.

Mátés, L., Chuah, M.K.L., Belay, E., Jerchow, B., Manoj, N., Acosta-Sanchez, A., Grzela, D.P., Schmitt, A., Becker, K., Matrai, J., et al. (2009) Molecular evolution of a novel hyperactive Sleeping Beauty transposase enables robust stable gene transfer in vertebrates. Nat. Genet. **41**, 753–761.

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.

Matzinger, P. (2002) The Danger Model: A Renewed Sense of Self. Science **296**, 301–305.

Mavragani, C.P., Sagalovskiy, I., Guo, Q., Nezos, A., Kapsogeorgou, E.K., Lu, P., Liang Zhou, J., Kirou, K.A., Seshan, S.V., Moutsopoulos, H.M., et al. (2016) Expression of Long Interspersed Nuclear Element 1 Retroelements and Induction of Type I Interferon in Patients With Systemic Autoimmune Disease. Arthritis Rheumatol. Hoboken NJ **68**, 2686–2696.

Mavragani, C.P., Nezos, A., Sagalovskiy, I., Seshan, S., Kirou, K.A., and Crow, M.K. (2018) Defective regulation of L1 endogenous retroelements in primary Sjogren's syndrome and systemic lupus erythematosus: Role of methylating enzymes. J. Autoimmun. **88**, 75–82.

McClintock, B. (1950) The Origin and Behavior of Mutable Loci in Maize. Proc. Natl. Acad. Sci. U. S. A. **36**, 344–355.

Mehdipour, P., Marhon, S.A., Ettayebi, I., Chakravarthy, A., Hosseini, A., Wang, Y., de Castro, F.A., Loo Yau, H., Ishak, C., Abelson, S., et al. (2020) Epigenetic therapy induces transcription of inverted SINEs and ADAR1 dependency. Nature **588**, 169–173.

Metzner, M., Jäck, H.-M., and Wabl, M. (2012) LINE-1 Retroelements Complexed and Inhibited by Activation Induced Cytidine Deaminase. PLoS ONE **7**, e49358.

Mi, S., Lee, X., Li, X., Veldman, G.M., Finnerty, H., Racie, L., LaVallie, E., Tang, X.Y., Edouard, P., Howes, S., et al. (2000) Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. Nature **403**, 785–789.

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.

Miller, J.M., and Enemark, E.J. (2016) Fundamental Characteristics of AAA+ Protein Family Structure and Function. Archaea **2016**.

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**, e30058.

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

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.e12.

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

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.

Montoya-Durango, D.E., Liu, Y., Teneng, I., Kalbfleisch, T., Lacy, M.E., Steffen, M.C., and Ramos, K.S. (2009) Epigenetic Control of Mammalian LINE-1 Retrotransposon by Retinoblastoma Proteins. Mutat. Res. **665**, 20–28.

Montoya-Durango, D.E., Ramos, K.A., Bojang, P., Ruiz, L., Ramos, I.N., and Ramos, K.S. (2016) LINE-1 silencing by retinoblastoma proteins is effected through the nucleosomal and remodeling deacetylase multiprotein complex. BMC Cancer **16**, 38.

Moran, J.V., Holmes, S.E., Naas, T.P., DeBerardinis, R.J., Boeke, J.D., and Kazazian, H.H. (1996) High Frequency Retrotransposition in Cultured Mammalian Cells. Cell **87**, 917–927.

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

Motwani, M., Pesiridis, S., and Fitzgerald, K.A. (2019) DNA sensing by the cGAS–STING pathway in health and disease. Nat. Rev. Genet. **20**, 657–674.

Muckenfuss, H., Hamdorf, M., Held, U., Perković, M., Löwer, J., Cichutek, K., Flory, E., Schumann, G.G., and Münk, C. (2006) APOBEC3 Proteins Inhibit Human LINE-1 Retrotransposition. J. Biol. Chem. **281**, 22161–22172.

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

Myers, J.S., Vincent, B.J., Udall, H., Watkins, W.S., Morrish, T.A., Kilroy, G.E., Swergold, G.D., Henke, J., Henke, L., Moran, J.V., et al. (2002) A Comprehensive Analysis of Recently Integrated Human Ta L1 Elements. Am. J. Hum. Genet. **71**, 312–326.

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.

Nichols, P.J., Bevers, S., Henen, M., Kieft, J.S., Vicens, Q., and Vögeli, B. (2021) Recognition of non-CpG repeats in Alu and ribosomal RNAs by the Z-RNA binding domain of ADAR1 induces A-Z junctions. Nat. Commun. **12**, 793.

Niewiadomska, A.M., Tian, C., Tan, L., Wang, T., Sarkis, P.T.N., and Yu, X.-F. (2007) Differential Inhibition of Long Interspersed Element 1 by APOBEC3 Does Not Correlate with High-Molecular-Mass-Complex Formation or P-Body Association. J. Virol. **81**, 9577–9583.

Nurk, S., Koren, S., Rhie, A., Rautiainen, M., Bzikadze, A.V., Mikheenko, A., Vollger, M.R., Altemose, N., Uralsky, L., Gershman, A., et al. (2022) The complete sequence of a human genome. Science **376**, 44–53.

Olovnikov, I.A., Adyanova, Z.V., Galimov, E.R., Andreev, D.E., Terenin, I.M., Ivanov, D.S., Prassolov, V.S., and Dmitriev, S.E. (2007) Key role of the internal 5′-UTR segment in the transcription activity of the human L1 retrotransposon. Mol. Biol. **41**, 453–458.

Ono, R., Nakamura, K., Inoue, K., Naruse, M., Usami, T., Wakisaka-Saito, N., Hino, T., Suzuki-Migishima, R., Ogonuki, N., Miki, H., et al. (2006) Deletion of Peg10, an imprinted gene acquired from a retrotransposon, causes early embryonic lethality. Nat. Genet. **38**, 101–106.

Onomoto, K., Onoguchi, K., and Yoneyama, M. (2021) Regulation of RIG-I-like receptor-mediated signaling: interaction between host and viral factors. Cell. Mol. Immunol. **18**, 539–555.

Orecchini, E., Doria, M., Antonioni, A., Galardi, S., Ciafrè, S.A., Frassinelli, L., Mancone, C., Montaldo, C., Tripodi, M., and Michienzi, A. (2017) ADAR1 restricts LINE-1 retrotransposition. Nucleic Acids Res. **45**, 155–168.

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

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.

Pace, J.K., and Feschotte, C. (2007) The evolutionary history of human DNA transposons: Evidence for intense activity in the primate lineage. Genome Res. **17**, 422–432.

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

Percharde, M., Lin, C.-J., Yin, Y., Guan, J., Peixoto, G.A., Bulut-Karslioglu, A., Biechele, S., Huang, B., Shen, X., and Ramalho-Santos, M. (2018) A LINE1-Nucleolin Partnership Regulates Early Development and ESC Identity. Cell **174**, 391-405.e19.

Pereira, G.C., Sanchez, L., Schaughency, P.M., Rubio-Roldán, A., Choi, J.A., Planet, E., Batra, R., Turelli, P., Trono, D., Ostrow, L.W., et al. (2018) Properties of LINE-1 proteins and repeat element expression in the context of amyotrophic lateral sclerosis. Mob. DNA **9**, 35.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Meng, E.C., Couch, G.S., Croll, T.I., Morris, J.H., and Ferrin, T.E. (2021) UCSF ChimeraX: Structure visualization for researchers, educators, and developers. Protein Sci. **30**, 70–82.

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

Pickeral, O.K., Makałowski, W., Boguski, M.S., and Boeke, J.D. (2000) Frequent Human Genomic DNA Transduction Driven by LINE-1 Retrotransposition. Genome Res. **10**, 411–415.

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

Rehwinkel, J., and Gack, M.U. (2020) RIG-I-like receptors: their regulation and roles in RNA sensing. Nat. Rev. Immunol. **20**, 537–551.

Reis, F.P., Pobre, V., Silva, I.J., Malecki, M., and Arraiano, C.M. (2013) The RNase II/RNB family of exoribonucleases: putting the 'Dis' in disease. WIREs RNA **4**, 607–615.

Ricci, M., Peona, V., Guichard, E., Taccioli, C., and Boattini, A. (2018) Transposable Elements Activity is Positively Related to Rate of Speciation in Mammals. J. Mol. Evol. **86**, 303–310.

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.

Richardson, S.R., Doucet, A.J., Kopera, H.C., Moldovan, J.B., Garcia-Pérez, 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., Muñoz-Lopez, M., Jesuadian, J.S., Kempen, M.-J.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.

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

Rodić, 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.

Rodriguez-Martin, B., Alvarez, E.G., Baez-Ortega, A., Zamora, J., Supek, F., Demeulemeester, J., Santamarina, M., Ju, Y.S., Temes, J., Garcia-Souto, D., et al. (2020) Pan-cancer analysis of whole genomes identifies driver rearrangements promoted by LINE-1 retrotransposition. Nat. Genet. **52**, 306–319.

Roulois, D., Yau, H.L., Singhania, R., Wang, Y., Danesh, A., Shen, S.Y., Han, H., Liang, G., Pugh, T.J., Jones, P.A., et al. (2015) DNA-demethylating agents target colorectal cancer cells by inducing viral mimicry by endogenous transcripts. Cell **162**, 961–973.

Rowe, H.M., Jakobsson, J., Mesnard, D., Rougemont, J., Reynard, S., Aktas, T., Maillard, P.V., Layard-Liesching, H., Verp, S., Marquis, J., et al. (2010) KAP1 controls endogenous retroviruses in embryonic stem cells. Nature **463**, 237–240.

Rusinova, I., Forster, S., Yu, S., Kannan, A., Masse, M., Cumming, H., Chapman, R., and Hertzog, P.J. (2013) INTERFEROME v2.0: an updated database of annotated interferon-regulated genes. Nucleic Acids Res. **41**, D1040–D1046.

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

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.

Schmitt, K., Richter, C., Backes, C., Meese, E., Ruprecht, K., and Mayer, J. (2013) Comprehensive Analysis of Human Endogenous Retrovirus Group HERV-W Locus Transcription in Multiple Sclerosis Brain Lesions by High-Throughput Amplicon Sequencing. J. Virol. **87**, 13837–13852.

Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012) NIH Image to ImageJ: 25 years of image analysis. Nat. Methods **9**, 671–675.

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

Seczynska, M., Bloor, S., Cuesta, S.M., and Lehner, P.J. (2022) Genome surveillance by HUSH-mediated silencing of intronless mobile elements. Nature **601**, 440–445.

Sekita, Y., Wagatsuma, H., Nakamura, K., Ono, R., Kagami, M., Wakisaka, N., Hino, T., Suzuki-Migishima, R., Kohda, T., Ogura, A., et al. (2008) Role of retrotransposon-derived imprinted gene, Rtl1, in the feto-maternal interface of mouse placenta. Nat. Genet. **40**, 243–248.

Serrato-Capuchina, A., and Matute, D.R. (2018) The Role of Transposable Elements in Speciation. Genes **9**, 254.

Shen, J.Z., Qiu, Z., Wu, Q., Finlay, D., Garcia, G., Sun, D., Rantala, J., Barshop, W., Hope, J.L., Gimple, R.C., et al. (2021) FBXO44 promotes DNA replication-coupled repetitive element silencing in cancer cells. Cell **184**, 352-369.e23.

Sherman, B.T., Hao, M., Qiu, J., Jiao, X., Baseler, M.W., Lane, H.C., Imamichi, T., and Chang, W. (2022) DAVID: a web server for functional enrichment analysis and

functional annotation of gene lists (2021 update). Nucleic Acids Res. **50**, W216–W221.

Shi, X., Seluanov, A., and Gorbunova, V. (2007) Cell Divisions Are Required for L1 Retrotransposition. Mol. Cell. Biol. **27**, 1264–1270.

Shukla, R., Upton, K.R., Muñoz-Lopez, M., Gerhardt, D.J., Fisher, M.E., Nguyen, T., Brennan, P.M., Baillie, J.K., Collino, A., Ghisletti, S., et al. (2013) Endogenous Retrotransposition Activates Oncogenic Pathways in Hepatocellular Carcinoma. Cell **153**, 101–111.

Simon, M., Van Meter, M., Ablaeva, J., Ke, Z., Gonzalez, R.S., Taguchi, T., De Cecco, M., Leonova, K.I., Kogan, V., Helfand, S.L., et al. (2019) LINE1 Derepression in Aged Wild-Type and SIRT6-Deficient Mice Drives Inflammation. Cell Metab. **29**, 871-885.e5.

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.

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.

Soper, S.F.C., van der Heijden, G.W., Hardiman, T.C., Goodheart, M., Martin, S.L., de Boer, P., and Bortvin, A. (2008) Mouse maelstrom, a component of nuage, is essential for spermatogenesis and transposon repression in meiosis. Dev. Cell **15**, 285–297.

Speek, M. (2001) Antisense Promoter of Human L1 Retrotransposon Drives Transcription of Adjacent Cellular Genes. Mol. Cell. Biol. **21**, 1973–1985.

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.

Stetson, D.B., Ko, J.S., Heidmann, T., and Medzhitov, R. (2008) Trex1 prevents cell-intrinsic initiation of autoimmunity. Cell **134**, 587–598.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005) Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. U. S. A. **102**, 15545–15550.

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

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

Surapureddi, S., Yu, S., Bu, H., Hashimoto, T., Yeldandi, A.V., Kashireddy, P., Cherkaoui-Malki, M., Qi, C., Zhu, Y.-J., Rao, M.S., et al. (2002) Identification of a transcriptionally active peroxisome proliferator-activated receptor α-interacting

cofactor complex in rat liver and characterization of PRIC285 as a coactivator. Proc. Natl. Acad. Sci. U. S. A. **99**, 11836–11841.

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., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N.T., Morris, J.H., Bork, P., et al. (2019) STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. **47**, D607–D613.

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

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.

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.

Thomas, C.A., Tejwani, L., Trujillo, C.A., Negraes, P.D., Herai, R.H., Mesci, P., Macia, A., Crow, Y.J., and Muotri, A.R. (2017) Modeling of TREX1-dependent autoimmune disease using human stem cells highlights L1 accumulation as a source of neuroinflammation. Cell Stem Cell **21**, 319-331.e8.

Tokuyama, M., Kong, Y., Song, E., Jayewickreme, T., Kang, I., and Iwasaki, A. (2018) ERVmap analysis reveals genome-wide transcription of human endogenous retroviruses. Proc. Natl. Acad. Sci. U. S. A. **115**, 12565–12572.

Tomaru, T., Satoh, T., Yoshino, S., Ishizuka, T., Hashimoto, K., Monden, T., Yamada, M., and Mori, M. (2006) Isolation and Characterization of a Transcriptional Cofactor and Its Novel Isoform that Bind the Deoxyribonucleic Acid-Binding Domain of Peroxisome Proliferator-Activated Receptor-γ. Endocrinology **147**, 377–388.

Tsokos, G.C., Lo, M.S., Reis, P.C., and Sullivan, K.E. (2016) New insights into the immunopathogenesis of systemic lupus erythematosus. Nat. Rev. Rheumatol. **12**, 716–730.

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.

Tunbak, H., Enriquez-Gasca, R., Tie, C.H.C., Gould, P.A., Mlcochova, P., Gupta, R.K., Fernandes, L., Holt, J., van der Veen, A.G., Giampazolias, E., et al. (2020) The HUSH complex is a gatekeeper of type I interferon through epigenetic regulation of LINE-1s. Nat. Commun. **11**, 5387.

Turelli, P., Vianin, S., and Trono, D. (2004) The Innate Antiretroviral Factor APOBEC3G Does Not Affect Human LINE-1 Retrotransposition in a Cell Culture Assay. J. Biol. Chem. **279**, 43371–43373.

Ueda, M.T., Kryukov, K., Mitsuhashi, S., Mitsuhashi, H., Imanishi, T., and Nakagawa, S. (2020) Comprehensive genomic analysis reveals dynamic evolution of endogenous retroviruses that code for retroviral-like protein domains. Mob. DNA **11**, 29.

Ukadike, K.C., and Mustelin, T. (2021) Implications of Endogenous Retroelements in the Etiopathogenesis of Systemic Lupus Erythematosus. J. Clin. Med. **10**.

Van Meter, M., Kashyap, M., Rezazadeh, S., Geneva, A.J., Morello, T.D., Seluanov, A., and Gorbunova, V. (2014) SIRT6 represses LINE1 retrotransposons by ribosylating KAP1 but this repression fails with stress and age. Nat. Commun. **5**, 5011.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., et al. (2001) The sequence of the human genome. Science **291**, 1304–1351.

Vieira, V.C., and Soares, M.A. (2013) The Role of Cytidine Deaminases on Innate Immune Responses against Human Viral Infections. BioMed Res. Int. **2013**, e683095.

Volkman, H.E., and Stetson, D.B. (2014) The enemy within: endogenous retroelements and autoimmune disease. Nat. Immunol. **15**, 415–422.

Walker, J.E., Saraste, M., Runswick, M.J., and Gay, N.J. (1982) Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO J. **1**, 945–951.

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

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.

Waring, M., and Britten, R.J. (1966) Nucleotide sequence repetition: a rapidly reassociating fraction of mouse DNA. Science **154**, 791–794.

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

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.

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

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

White, T.E., Brandariz-Nuñez, A., Han, K., Sawyer, S.L., Kim, B., and Diaz-Griffero, F. (2016) Modulation of LINE-1 retrotransposition by a human SAMHD1 polymorphism. Virol. Rep. **6**, 53–60.

Wichroski, M.J., Robb, G.B., and Rana, T.M. (2006) Human Retroviral Host Restriction Factors APOBEC3G and APOBEC3F Localize to mRNA Processing Bodies. PLoS Pathog. **2**, e41.

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.

Woodcock, D.M., Lawler, C.B., Linsenmeyer, M.E., Doherty, J.P., and Warren, W.D. (1997) Asymmetric Methylation in the Hypermethylated CpG Promoter Region of the Human L1 Retrotransposon. J. Biol. Chem. **272**, 7810–7816.

Xie, Y., Mates, L., Ivics, Z., Izsvák, Z., Martin, S.L., and An, W. (2013) Cell division promotes efficient retrotransposition in a stable L1 reporter cell line. Mob. DNA **4**, 10.

Xue, B., Sechi, L.A., and Kelvin, D.J. (2020) Human Endogenous Retrovirus K (HML-2) in Health and Disease. Front. Microbiol. **11**, 1690.

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

Yang, Z., Boffelli, D., Boonmark, N., Schwartz, K., and Lawn, R. (1998) Apolipoprotein(a) Gene Enhancer Resides within a LINE Element. J. Biol. Chem. **273**, 891–897.

Yu, Q., Carbone, C.J., Katlinskaya, Y.V., Zheng, H., Zheng, K., Luo, M., Wang, P.J., Greenberg, R.A., and Fuchs, S.Y. (2015) Type I Interferon Controls Propagation of Long Interspersed Element-1. J. Biol. Chem. **290**, 10191–10199.

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.

Zhang, T., Yin, C., Fedorov, A., Qiao, L., Bao, H., Beknazarov, N., Wang, S., Gautam, A., Williams, R.M., Crawford, J.C., et al. (2022) ADAR1 masks the cancer immunotherapeutic promise of ZBP1-driven necroptosis. Nature **606**, 594–602.

Zhang, Y., Cheng, T.C., Huang, G., Lu, Q., Surleac, M.D., Mandell, J.D., Pontarotti, P., Petrescu, A.J., Xu, A., Xiong, Y., et al. (2019) Transposon molecular domestication and the evolution of the RAG recombinase. Nature **569**, 79–84.

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-Goutières Syndrome-Related SAMHD1. Cell Rep. **4**, 1108–1115.

Zhao, K., Du, J., Peng, Y., Li, P., Wang, S., Wang, Y., Hou, J., Kang, J., Zheng, W., Hua, S., et al. (2018) LINE1 contributes to autoimmunity through both RIG-I- and MDA5-mediated RNA sensing pathways. J. Autoimmun. **90**, 105–115.

Zhao, Y., Oreskovic, E., Zhang, Q., Lu, Q., Gilman, A., Lin, Y.S., He, J., Zheng, Z., Lu, J.Y., Lee, J., et al. (2021) Transposon-triggered innate immune response confers cancer resistance to the blind mole rat. Nat. Immunol. **22**, 1219–1230.