LINE1: IMPLICATIONS IN THE ETIOLOGY OF HUMAN DISEASES, CLINICAL UTILITIES, AND PHARMACOLOGICAL TARGET FOR DISEASE TREATMENT

by

Mahwish Khalid

A Thesis Submitted to the Faculty of the

DEPARTMENT OF CELLULAR AND MOLECULAR MEDICINE

In Partial Fulfillment of the Requirements

For the Degree of

MASTER OF SCIENCE

In the Graduate College

THE UNIVERSITY OF ARIZONA

2017

STATEMENT BY AUTHOR

The thesis titled *Line 1: Implications in the etiology of human diseases, clinical utilities, and pharmacological target for disease treatment* prepared by *Mahwish Khalid* has been submitted in partial fulfillment of requirements for a master's degree at the University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this thesis are allowable without special permission, provided that an accurate acknowledgement of the source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: *Mahwish Khalid*

APPROVAL BY THESIS DIRECTOR

This thesis has been approved on the date shown below:

Dr. Kenneth Ramos, MD, PhD
Professor of Medicine

12/5/17
Date

Acknowledgements

The author would like to express her sincere gratitude to Dr. Pasano Bojang for his continuous support and his patience. Furthermore, the author would like to thank Dr. Kenneth Ramos, Dr. Stefano Guerra and Dr. Jean Wilson for their guidance and involvement.

Table of Contents

List of Figures

Page

List of Tables

Page

Abstract:

Long interspersed nuclear elements-1 (*Line-1* or *L1*) accounts for approximately 17% of the human genome. The majority of *L1*s are inactive, but ~100 remain retrotransposon competent (*RC-L1*) and retrotranspose through RNA intermediates to different locations of the genome. It is well established that *L1* is involved in both disease initiation and progression via retrotransposition dependent and independent mechanisms. Retrotransposed *L1* sequences disrupt loci (e.g. gene structure) in ways that lead to human disease, and activities of L1 si/piRNA, ORF1 and ORF2 proteins are implicated in the etiology and progression of human diseases, such as breast and colon cancer. Despite these implications, very little is known about pharmacological molecules that inhibit and reverse L1's harmful effects. The clinical utility of *L1* as a player in tumorigenesis and as a biomarker for disease initiation and progression is not thoroughly understood. In this review, we analyzed the life cycle of L1, its roles in disease initiation and progression, clinical utilities and potential as a pharmacological target and a biomarker for the diagnosis and treatment of human diseases, such as cancer.

1. Introduction

DNA transposons and retrotransposons are two classes of transposable elements (TE's) that comprise 45% of the human genome **(Lander et al., 2001)**. DNA transposons use a cut and paste mechanism to transpose into new genomic locations and do not require an RNA intermediate **(Muñoz-López and García-Pérez, 2010)**. Retrotransposons on the other hand use a copy and paste mechanism and require an RNA intermediate to retrotranspose into new genomic locations. Retrotransposons are divided into Long Interspersed Nuclear Elements (LINEs/ L1s), and Short Interspersed Nuclear Elements (SINEs) (e.g. Alu elements, tandem repeats and SINE-VNTR-Alus (SVA's) **(Deininger et al., 2003; Hancks and Kazazian, 2010; Smit, 1999)**. Human *L1* is ~6kb long and consists of two open reading frames (i.e. ORF1, ORF2), a 5' untranslated region (UTR) and a 3' UTR with poly (A) tail and signal **(Dombroski et al., 1991; Swergold, 1990)**. The 5'UTR acts as a bidirectional promoter to ORF1&2 in the sense orientation and to ORF0 & L1 siRNAs in the antisense orientation **(Denli et al., 2015; Yang and Kazazian, 2006).** *L1* functions primarily as an autonomous mobile element that propagates its DNA or other DNAs within the genome through an RNA intermediate, a process termed retrotransposition **(Kleckner, 1990).** A cycle of *L1* retrotransposition starts with the delocalization of the NuRD repressor complex **(Montoya-Durango et al., 2016)** and the removal of repressive marks such as histone-3 lysine-9 trimethylations (H3K9me3), histone-3 lysine-20 trimethylations (H3K20me3), and the addition of activating marks such as histone-3 lysine-4 trimethylation (H3k4m3) and histone-3 lysine-9 acetylation (H3K9Ac) from the L1 promoter **(Figure 1, step 1) (Montoya-Durango et al., 2009; Montoya-Durango et al., 2016; Teneng et al., 2011)**. Once activated, ORF1&2 and/or ORF0 & *L1* siRNAs are transcribed by the sense and anti-sense promoters (ASP), respectively **(Figure 1, step 2) (Denli et al., 2015; Yang and Kazazian, 2006)**. Transcribed mRNAs are exported into

the cytoplasm **(Figure 1, step 4)** where they are translated into ORF0, 1,& 2 proteins **(Figure 1, step 6).** ORF1 and ORF2 proteins in the cytoplasm exhibit a cis or trans preference and bind to L1 encoding mRNA or mRNA from other non-LTR retrotransposons such as: Alu and SVA to form a ribonucleoprotein (RNP) **(Figure 1, steps 8 & 9)**. Phosphorylation of ORF1 by protein kinase A (PKA) is essential for L1 retrotransposition **(Figure 1, step 11)(Bojang et al., 2013; Cook et al., 2015).** Finally, RNP is imported to the nucleus where target-primed reverse transcription (TPRT) occurs. This is where the endonuclease domain of ORF2 nicks genomic DNA to expose a 3' hydroxyl group (OH-) and the reverse transcriptase domain transcribes L1 mRNA into DNA and re-inserts it into a new genomic location **(Figure 1, step 12)(Wei et al., 2001)**. This completes retrotransposition and results in the reintegration of L1 as a full length or as a truncated sequence into the genome **(Figure 1, step 14)**.

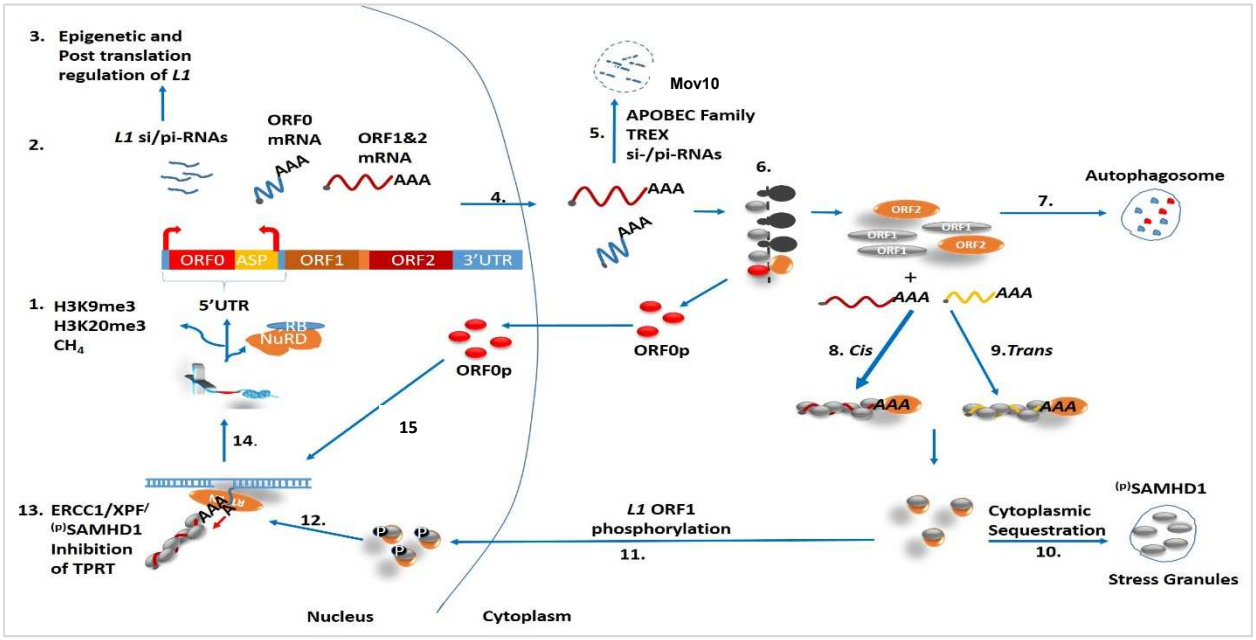

Figure 1: Life cycle of L1. Once epigenetic repressive marks and complexes are removed, L1 mRNA is transcribed and translated to form RNP in trans or cis. RNP is then imported into the nucleus where the DNA sequence is inserted into the genome (Step 14). There are also many

mechanisms that host cells use to inhibit L1 expression. Cytoplasmic L1 inhibition includes degradation of L1 RNA and proteins via the APOBEC family, Trex-1, si/piRNAs and by Mov10 (Step 5). Nuclear epigenetic regulation can also be induced by si/piRNAs (Step 3). Other mechanisms include degrading L1 RNA and proteins in the autophagasome (Step 7) or by sequestering the RNP within stress granules (Step 10). RNP that have translocated into the nucleus can be inhibited by ERCC1/ XPF which recognizes the cDNA in the process of TPRT, removes it, and restores the original target DNA sequence.

1.1 Regulation of L1 activity in cells

Because of the mutagenic potential of L1, cells have developed a plethora of repressive mechanisms or checkpoints to inhibit L1 expression and retrotransposition in germ and somatic cells. L1 is normally silenced in all somatic cells and is predominately expressed in the germline. Thus, regulating L1 expression in the germline in crucial to maintaining genomic integrity through each generation **(Ishiuchi and Torres-Padilla, 2014)**. These repressive mechanisms used by cells to inhibit L1 include cytoplasmic RNA and protein degradation/inhibition by a series of mechanisms and nuclear regulation by epigenetic mechanisms and si/pi-RNA's, as described below.

1.2. ***Cytoplasmic regulation of L1 in germ cells***

Cytoplasmic regulation involves using post-transcriptional gene silencing (PTGS). This is where L1's own mRNA, produced by its bidirectional promotor, is used to silence its expression. For example, sense and antisense L1-RNAs are produced and can bind to each other to form double stranded RNAs (dsRNAs). These dsRNAs are then cut by dicer into small fragments known as

endogenous (endo)-siRNAs **(Soifer and Rossi, 2006).** Endo-siRNAs can then degrade L1 mRNA via an RNA interference (RNAi) mechanism **(Yang and Kazazian, 2006)**. The RNAi mechanism entails the siRNA association with the RNA induced silencing complex (RISC) and with AGO2 in the cytoplasm. This complex induces endonucleolytic cleavage to destroy target mRNA **(Figure 1, step 5) (Hammond et al., 2001; Schwarz et al., 2004)**. Endo-siRNA can also cause epigenetic repression of L1, discussed later in this review (Nuclear epigenetic regulation of L1 in germ cells). Another mechanism used to inhibit L1 involves the use of PIWI interacting RNA (piRNA) which are noncoding single stranded RNAs that use four mechanisms to suppress L1 in germ cells: (1) RNA degradation **(De Fazio et al., 2011)**, (2) translational inhibition **(Kiriakidou et al., 2007)**, (3) DNA methylation **(Sigurdsson et al., 2012)**, and (4) histone modification **(Pezic et al., 2014).** Mechanisms 1 and 2 inhibit L1 in the cytoplasm, while mechanisms 3 and 4 occur in the nucleus. Regulation begins with the transcription of piRNA transcripts from piclusters, which are genomic loci containing repetitive element sequences. In *Drosophila,* these transcripts are processed into piRNA intermediates or primary piRNAs within the perinuclear nuage region of the cytoplasm **(De Fazio et al., 2011; Zhang et al., 2012).** In the cytoplasm, piRNAs form complexes with PIWI proteins, and produce secondary piRNAs (silencing triggers) via a ping-pong amplification cycle **(De Fazio et al., 2011; Sigurdsson et al., 2012)**. The ping pong amplification cycle is an adaptive mechanism used to produce/amplify secondary piRNAs that target and degrade active L1 mRNA**(Czech and Hannon, 2016)**. This occurs by the piRNAs first guiding PIWI proteins to their target transcripts. The transposon transcripts that are complementary to these secondary piRNAs are then destroyed via endonucleolytic cleavage **(Le Thomas et al., 2013)**. Both siRNA and piRNA mechanisms are used to clear the cytoplasm of L1 mRNAs.

L1 mRNAs that escape endonucleolytic cleavage are inhibited during translation (second

mechanism used by piRNA). Normally, translation involves the recruitment of a eukaryotic translation initiation factor 4E (eIF4e). However, the PIWI proteins have similar binding domains as eIF4e. Thus, the PIWI complex inhibits eIF4e from binding and translating target mRNA sequences through competitive inhibition **(Kiriakidou et al., 2007).**

1.3. *Nuclear epigenetic regulation of L1 in germ cells*

Depending on the proteins they associate with, PIWI-piRNA complexes either stay in the cytoplasm to cause PTGS (described in Part 1.2), or localize to the nucleus **(Brennecke et al., 2007)**, to induce epigenetic regulation. In the embryonic germ line, PIWI proteins: Milli (or Piwil2) and Miwi2 (or Piwil4) direct L1 epigenetic regulation **(De Fazio et al., 2011)** and histone methylation **(Pezic et al., 2014)**. DNA methylation at L1 CpG sites and histone methylation/modification (fourth mechanism used by piRNA) occur via recruitment of DNA methyltransferases (DNMTs) by the PIWI-piRNA complex and an RNA-induced transcriptional silencing (RITC) complex **(Motamedi et al., 2004; Noma et al., 2004; Wang et al., 2003; Yang et al., 2007).** Like piRNA, siRNAs can also exert epigenetic regulation of L1 by inducing hypermethylation of the internal L1 5' promoter through a similar mechanism **(Chen et al., 2012; Li et al., 2013)**.

Epigenetic marks refer to the methylation of DNA, usually on CpG sites, or histone modification which dictate whether genes are switched "on" or "off" **(Ruthenburg et al., 2007)**. These epigenetic marks are transient and interact with many modifying enzymes within the nucleus. This results in different patterns of DNA methylation which are important for embryo development and differentiation programming of cells **(Anifandis et al., 2015; Teneng et al., 2011)**. Epigenetic marks such as H3K9me3 and H3K20me3 trimethylation are L1 repressive and result in

heterochromatin formation **(Brower-Toland et al., 2007; Teneng et al., 2011)** and transcriptional repression of L1. H3K4me3 and H3K9Ac are L1 activating marks **(Huang et al., 2013)** which result in euchromatin formation and transcriptional activation of L1. These marks have also been shown to occur in somatic tissue to reactivate L1 by environmental carcinogens **(Montoya-Durango et al., 2009; Teneng et al., 2011).** For example, Teneng et al., 2007 showed L1 activation by benzo (a) pyrene (BaP) in human cervical carcinoma (HeLa) cells. Additionally, L1 was also activated by 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) and UV irradiation in mouse vascular smooth muscle cells (mVSMC), human microvascular endothelial cells and in mouse embryonic kidney cells (mK4) **(Teneng et al., 2007)**.

1.4. *Other mechanisms for L1 inhibition in the nucleus and cytoplasm*

Other mechanisms used by the host cell to inhibit L1 expression and retrotransposition are: Trex-1/Apolipoprotein B (apo B)-editing catalytic subunit 1 (APOBEC1) mediated L1 mRNA degradation and sequestration, autophagy (metabolism) and phosphorylation of L1 proteins. APOBEC1 from the APOBEC3 (APO3) gene family catalyzes the deamination of cytosine to uracil within the L1 mRNA sequence **(Ikeda et al., 2011) (Figure 1, step 5).** Deamination is predicted to occur via nucleophilic attack which results in the placement of an oxygen double bond in place of an amine group on the fourth position of the cytosine nucleotide ring **(Harris and Liddament, 2004)**. These C-to-U point mutations on the mRNA are recognized, and target the mRNA for degradation by RNA degrading enzymes or endonucleases (**Houseley and Tollervey, 2009)**.

Furthermore, it has been reported that a host cellular protein known as Moloney leukemia virus type 10 protein (Mov10) **(Figure 1, step 5)** can associate with ORF1p and L1-RNA **(Li et al.,**

2013) in the RNP to cause degradation of L1-RNA via PTGS **(Figure 1, step 5)(Li et al., 2013)**. Another protein that interacts with ORF1p is Trex-1, an exonuclease that binds to prevent the accumulation of single-stranded DNA (ssDNA). Trex-1 interferes with L1-ORF1p binding to specific sequences of DNA and interferes with the expression of various genes **(Figure 1, step 5)(Neidhart et al., 2010).**

Host cells can also degrade L1 mRNA and proteins in the autophagosomes. L1 RNAs/proteins are selectively targeted by stress induced inhibition and degradation via autophagy receptors NDP52 **(Figure 1, step 7) (Guo et al., 2014).**

Another mechanism includes sequestering L1 mRNA and proteins into stress granules. This mechanism includes a SAM domain and HD domain containing protein 1 (SAMHD) which inhibits RNP formation by phosphorylation of elF2alpha, which induces cellular stress granule formation **(Figure 1, step 10) (Hu et al., 2015)**.

If L1 RNPs manage to cross the nuclear envelope and enter the nucleus, there are a series of host repair mechanisms that interfere with L1 integration. At or during reverse transcription, a series of mechanisms limit L1 retrotransposition, including the inhibition by ERCC1/ XPF complex. This complex is a heterodimer with 3' "flap" endonuclease activity **(Gasior et al., 2008).** It functions as a DNA repair protein during the mammalian nucleotide excision repair (NER) process **(de Laat et al., 1998)** and has been shown to limit L1 retrotransposition **(Gasior et al., 2008).** In mammals, ERCC1/ XPF recognizes the cDNA in the process of TPRT, removes it to restore the original target DNA sequence via DNA inter-strand cross-link repair **(Figure 1, step 13) (Gasior et al., 2008).**

Another mechanism of interest includes a new open reading frame in the antisense direction of L1 named ORF0 **(Denli et al., 2015)**. Structurally similar to genes, ORF0 contains a promoter, two

exons, and an intron that is transcribed and translated into ORF0 proteins **(Denli et al., 2015)**. ORF0 mRNA is capped and polyadenylated and its gene product is predominately nuclear **(Figure 1, step 15)(Denli et al., 2015)**. Functionally, ORF0 increases L1 mobility and hence its loss reduces retrotransposition of L1 in cultured cells **(Denli et al., 2015).** A large amount of ORF0 sequence is found within coding genes suggesting its mobilization through L1 retrotransposition **(Denli et al., 2015)**. In disease conditions, ORF0 localized in promyelocytic leukemia-adjacent nuclear bodies which are implicated in the initiation and/or progression of disease states **(Denli et al., 2015)**. Because ectopic L1 can retrotranspose independent of L1-ORF0, it is worth noting that ORF0 might not be necessary for retrotransposition, rather it is likely that it enhances retrotransposition.

It should be noted that all L1 inhibitory mechanisms are not present in every cell, but vary by cell type, and whether cells are embryonic or somatic in nature. If any of the above mechanisms fail, L1 can undergo a full cycle of retrotransposition that results in reintegration as full length or as a truncated insertion into the genome **(Figure 1, step 15).** These multilayered response mechanisms are in place to protect healthy somatic and germ cells from the damaging effects of L1. Our laboratory has shown that under disease conditions and during environmental insults, these regulatory mechanisms are lost resulting in a full cycle of retrotransposition **(Stribinskis and Ramos, 2006).** We have also shown that L1 influences disease progression and phenotype through mechanisms that are independent of L1 retrotransposition **(Bojang et al., 2013).** As continuous L1 retrotransposition is linked to genomic evolution and diversity of the human genome, it has also debunked all notions that our genome is static **(Deininger et al., 2003)**.

2. Implications of Line-1 in disease etiology and progression

L1 is directly or indirectly involved in tumor initiation and progression through retrotransposition dependent and independent mechanisms. Retrotransposition dependent mechanisms involve disruption and/or dysregulation of gene structure and function through insertion of L1 sequences or sequences of other elements such as Alu or SVAs. These new L1 insertions are recognized by the presence of a target site duplication sequences (TSD), which is a 4-8 bp sequence found flanking the L1 sequence **(Levin and Moran, 2011)**. The majority of L1 inserted sequences disrupt genes through the introduction of premature stop codons, deletion, alternative splice sites, polyadenylation sites, cis-regulatory elements, CpG island or alternative promoters. Retrotransposition independent mechanisms use L1 mRNA and proteins to perform functions that favor tumor growth. These involve L1 RNAs, or ORF1 and ORF2 proteins to function independently of the L1 retrotransposition pathway in cellular processes that enhance tumor growth. Such cellular processes include epithelial to mesenchymal transition (EMT), gene regulation through L1 mediated transcriptional interference, L1 derived regulatory RNA's, and L1 chimeric transcripts.

2.1. *Retrotransposition dependent mechanism*

L1 insertions can disrupt gene structure and function by seven mechanisms: (1) introduction of an early stop codon **(Schwahn et al., 1998)**, (2) deletions **(Narita et al., 1993)**, (3) aberrant splicing **(Samuelov et al., 2011),** (4) polyadenylation **(Kazazian et al., 1988)**, (5) hypermethylation of CpG sites **(Yoshida et al., 1998)**, (6) insertion of new promoter sequence **(Roman-Gomez et al., 2005)** and (7) change in gene regulatory networks.

L1 insertion was shown to introduce a stop codon into the retinitis pigmentosa type 2 (RP2) gene resulting in X-linked retinitis pigmentosa in one patient **(Schwahn et al., 1998)**. The stop codon occurred at a domain homologous to cofactor C, which plays an important functional role in RP2. A stop codon or sequence of "TAG" in a coding strand of DNA would be matched to a codon "ATC" on a second template DNA strand. Transcription of the template DNA strand or "ATC" codon would result in a stop codon "UAG" in the mRNA strand **(Figure 2)**. An early stop codon results in a truncated protein with some degree of loss of function. Schwahn et al., 1998 proposed that the introduction of this stop codon in the form of a missense mutation results in the accumulation of incorrectly folded proteins leading to the pathogenesis of RP2 **(Schwahn et al., 1998)**.

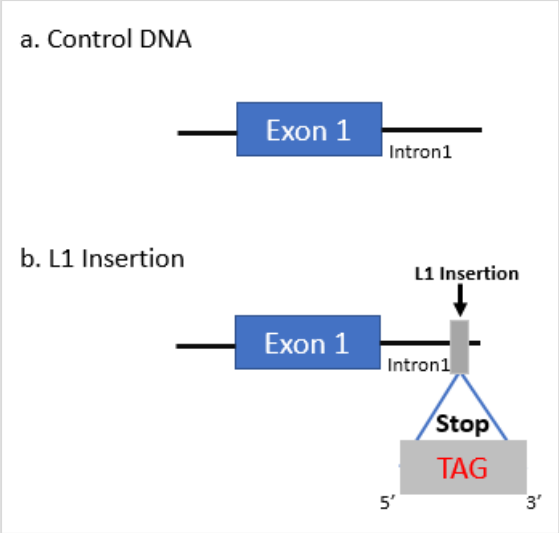

Figure 2: a. Control DNA without L1 insertion b. A depiction of an intronic L1 sequence and introduction of a stop codon "TAG" that results in transcription termination and a truncated protein.

In another study, Narita et al., 1993 identified L1 insertions in the 3' end of exon 44 in the Duchene Muscular Dystrophy (DMD) gene in two Japanese brothers. The insertion resulted in a deletion of two to four bases of exon 44 leading to skipping of exon 44 during mRNA splicing **(Narita et al., 1993)**. Exon skipping led to a stop codon (TGA) in codon 2182 **(Narita et al., 1993)** resulting in a truncated protein **(Figure 3)**. In another case, Matsuo et al., 1991, discovered an L1 insertion

that led to an intra-exon deletion of 53 base-pairs from exon 19. mRNA processing at this site resulted in skipping of exon 19 during splicing due to the introduction of new donor and acceptor sites **(Matsuo et al., 1991)**. This demonstrates a second mechanism by which L1's can disrupt gene structure and function through exon deletion.

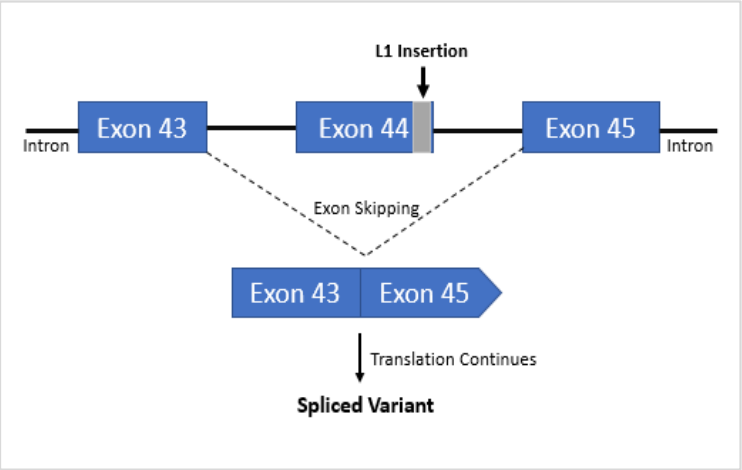

Figure 3: Diagram depicting an L1 insertion at the 3' end of exon 44 which results in a two to four base pair deletion, and skipping of exon 44 in the DMD gene. This results in the formation of a spliced variant form of the DMD protein.

Samuelov et al., 2011 identified an L1 insertion into intron 3 of the ABHD5 gene that results in Chanarin Dorfman Syndrome **(Samuelov et al., 2011)**. This L1 insertion created new donor and acceptor splicing sites before and after the L1 inserted sequence. During mRNA splicing, the inserted L1 sequence is added as a new exon into the mature transcript resulting in an alternative splice form of the ABHD5 protein **(Samuelov et al., 2011) (Figure 4)**. In chronic granulomatous disease (CGD), an L1 inserted sequence into intron 5 created a new exon in CYBB which results in the generation of a new spliced variant as described above **(Meischl et al., 2000)**. Thus, L1 insertions can disrupt gene structure and initiate disease by introducing aberrant splicing and is associated with the initiation of human disease.

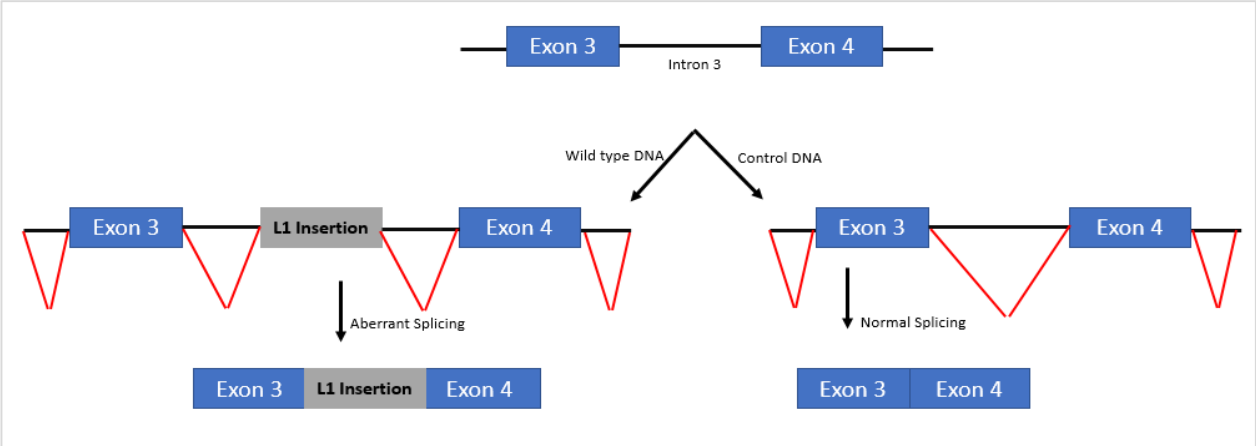

Figure 4: Diagram depicting L1 insertion and aberrant splicing of ABHD5 gene. An intronic L1 insertion introduces new acceptor and donor splice sites both before and after the L1 insertion. This results in the L1 sequence being inserted as a new exon in the mature transcript which results in a spliced variant form of the protein. On the right is a depiction of a control DNA without an L1 insertion for comparison. Adapted and modified from **Samuelov et al. 2011.**

L1 insertions have also been shown to introduce new 3' polyadenylation tracts (Mechanism 4) that disrupt gene function resulting in the initiation of tumorigenesis. The first case documented was an L1 insertion into exon 14 of Factor VIII in two patients resulting in Hemophilia A **(Kazazian et al., 1988).** A second occurred in exon 16 of APC gene resulting in the initiation of colon cancer **(Miki et al., 1992)**. These L1 insertion events include a chain of adenines within the coding DNA strand which is complementary to a chain of thymine's on the template DNA strand. After transcription and processing, a 5'cap and poly A tail are added to the mature RNA transcript **(Figure 5) (Brown, 2002).** When the translation machinery reaches the internal poly A tract within the mRNA strand, translation is terminated leading to a truncated protein **(Savas et al., 2006)**. In the case of the two patients with Hemophilia A, no Factor VIII activity was detected **(Kazazian et**

al., 1988), while in the second case, an inactive APC protein product was detected **(Miki et al., 1992)**.

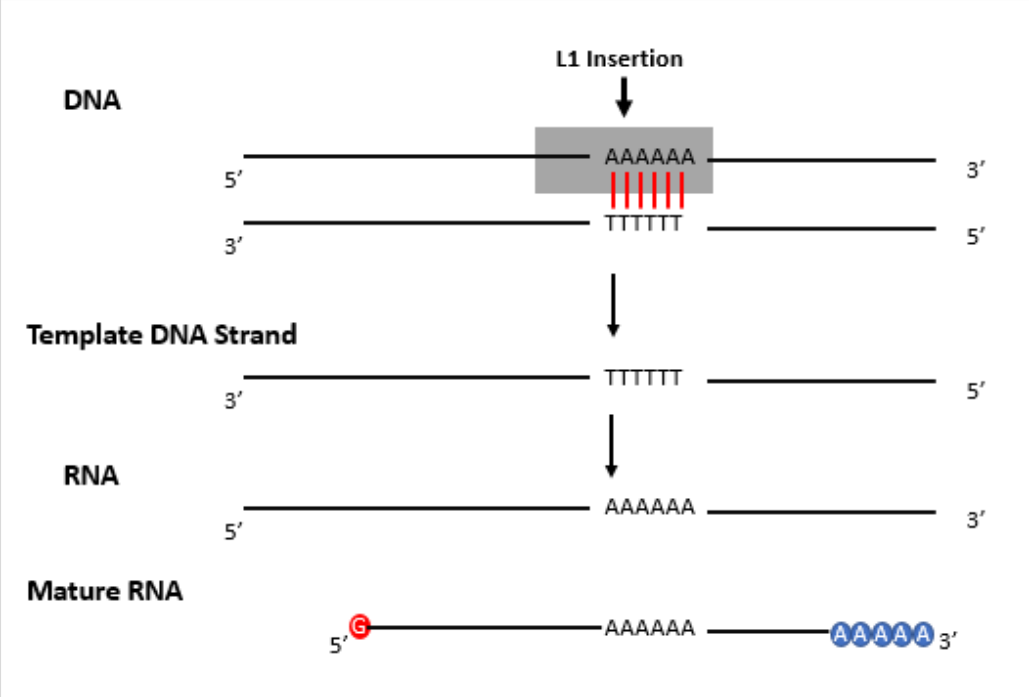

Figure 5: Schematic representation of an L1 inserted 3' polyadenylation tract in the coding strand of DNA and its resulting mRNA transcript. The L1 inserted poly A tract is complementary to a chain of thymine's on the template DNA strand. When this template DNA strand is transcribed it results in an internal poly A tract in the pre-RNA transcript. After processing of the pre-RNA, a 5' cap and poly A tail are added to the mature RNA transcript.

A fifth mechanism that can be used to disrupt gene structure is the introduction of a CpG rich sequence into genomic loci. L1 sequences contain many CpG sites **(Miousse and Koturbash, 2015; Teneng et al., 2011)**. Thus, L1 insertions can repress gene transcription by introducing CpG islands which can in turn be methylated to introduce heterochromatic formation **(Montoya-Durango et al., 2009; Montoya-Durango et al., 2016)**. The development of T-cell lymphoblastic

leukemias (T-ALLs) has been linked to hypermethylation of CpG sites introduced through insertion of L1 sequences **(Han and Boeke, 2005; Roman-Gomez et al., 2005)**. This demonstrates another mechanism by which L1s can modify gene function and contribute to the etiology of disease.

A sixth mechanism includes insertion of an L1 sequence to serve as a new promoter/enhancer of genes. Shukla et al., 2013 showed that an L1 insertion increased Suppression of tumorigenicity 18 (ST18) expression in hepatocellular carcinoma (HCC) by moving the repressive marks further away from the ST18 enhancer **(Shukla et al., 2013).** As inserted L1 sequences contain binding sites for transcription factors, this could thereby introduce a gain-of-function mutation in the ST18 gene **(Shukla et al., 2013).** This confirms L1's ability to function as an enhancer**.** L1 insertions also provide new promoters in the sense and antisense direction which may alter the expression or regulation of genes **(Shukla et al., 2013; Speek, 2001)**. For example, mouse L1 sequences act as alternative promoters to 13 % of genes involved in the regulation of developmental programs **(Ting et al., 1992)**.

L1 insertions have also been implicated in changing the regulatory network of genes (Mechanism 7). Teneng et al., 2011 described a previously unknown association of L1 and Wnt signaling in regard to nephrogenesis **(Teneng et al., 2011)**. The L1 sequence contains nine SOX/LEF binding sites **(Shukla et al., 2013)**. L1 retrotransposition scatters these **s**equences within the genome and interferes by re-wiring the genes for a different function.

All these different insertions of L1 sequences have been implicated in the etiology of human diseases especially cancers of epithelial origin. The table below summarizes L1 retrotransposition mediated disease initiations **(Table 1)**.

Disease	Gene	Mechanism of Insertion	Location	Reference
Retinitis Pigmentosa	RP2	Early Stop Codon	Intron 1	(Schwahn et al., 1998)
XLDCM	DMD	Deletion	Exon 44	(Narita et al., 1993)
Chanarin Dorfman Syndrome	ABHD5	Aberrant Splicing	Intron 3	(Samuelov et al., 2011)
Hemophilia A	Factor VIII	Polyadenylation	Exon 14	(Kazazian et al., 1988)
Colon Cancer	APC		Exon 16	(Miki et al., 1992)
T-Cell Lymphoblastic Leukemias	*	Hypermethylation of CpG site	*	(Roman-Gomez et al., 2005)
Lung Cancer	ST18	New Promoter Insertion	*	(Shukla et al., 2013)
*	*	Change in Gene Regulatory Network	*	(Kuwabara et al., 2009)

Table 1: L1 insertions into the following gene locations are associated with disease. The mechanism of L1 insertions such as, early stop codon, deletion, aberrant splicing, polyadenylation, hypermethylation of CpG site, new promoter insertion and change of gene regulatory network were identified, and location of L1 insertions into introns or exons within specific genes were recorded from different sources. (*) These are shown in cell-culture experiments.

2.2. *Retrotransposition independent mechanisms*

L1 uses an RNA intermediate to retrotranspose within the genome. Not all RNAs/proteins are used for mobilization, and L1 mRNAs and proteins are sometimes co-opted to perform or take part in cellular functions independent of retrotransposition that result in initiation and progression of human diseases. These cellular processes include epithelial to mesenchymal transition (EMT), gene regulation through L1 mediated transcriptional interference, L1 derived regulatory RNA's, and L1 chimeric transcripts.

L1 mediated transcriptional interference

Transcriptional interference (TI) consists of a series of mechanisms by which pre-existing L1 sequences in genes can disrupt the transcriptional process. Intronic L1 sequences in genes can produce transcriptional interference by intron retention, exogenization and polyadenylation **(Kaer et al., 2011)**. Pre-existing L1 sequences may also interfere with the activity of RNA polymerase II and result in early transcription termination **(Han et al., 2004)**. These sequences may also provide alternative splicing sites, producing different splicing variants that may be associated with cancer. For example, BCL2L gene is subject to alternative splicing to produce BCL-XL (long isoform) mRNA with anti-apoptotic activity, and BCL-XS mRNA with apoptotic activity **(Boise et al., 1993)**. BCL-XL mRNA allows cancer cells to escape cell-death, a hallmark of tumorigenesis. Thus, cancer cells can commandeer splicing variants within L1 sequences found in many genes for their own advantage **(Chen and Weiss, 2015).**

L1 derived regulatory RNA's

Many miRNAs originate from L1 sequences **(Ohms et al., 2014),** and the 3' UTR of L1 also contains miRNA target sites **(Spengler et al., 2013).** L1 miRNA target sites can sequester miRNA's by competing for endogenous RNAs (ceRNA) **(Cheng et al., 2015)**. ceRNAs are associated with many different cancers such as lung, prostate, gastric, breast, and endometrial cancers **(Cheng et al., 2015)**. An imbalance in ceRNA caused by alterations in miRNA, miRNA/ceRNA abundance or binding affinity may lead to cancer initiation and progression **(Cheng et al., 2015)**. For example, Ohms et al., 2014 showed that L1 activity controls the expression of *let-7* miRNA **(Ohms et al., 2014)**. Usually, mature *let-7* miRNA is involved in regulating oncogenic pathways such as the JAK and c-MYC pathway, by binding and suppressing

the expression of cancer related genes**(Thammaiah and Jayaram, 2016)**. However, in breast cancer it is predicted that L1 mRNA binds and forms complexes with RNA binding proteins such as, hnRNPA1. The L1-hnRNPA1 complex binds to pre-*let-7* miRNA and inhibits its processing into mature miRNA by KH-type splicing regulatory protein (KSRP) **(Ohms et al., 2014)**. In accord with these findings, L1 derived regulatory RNAs have been shown to disrupt *let-7* expression, thereby activating oncogenic pathways associated with breast cancer **(Ohms et al., 2014)**, and which may be associated with other epithelial cancers.

L1 chimeric transcripts

L1 chimeric mRNAs are formed by the transcription of L1 sequences that contain bidirectional promoter activity **(Rodić and Burns, 2013; Speek, 2001)**. L1 chimeric transcripts (LCTs) can cause gene activation, alter chromatin structure **(Ashe et al., 1997; Bolland et al., 2004; Gribnau et al., 2000)**, and contribute to human transcriptome diversity **(Kim et al., 2016).** A study by Cruckskank and Tufarell, 2009 isolated L1-antisense promotor (ASP) containing chimeric transcripts that are unable to retrotranspose, but can impose transcriptional activation **(Cruickshanks and Tufarelli, 2009)**. Chimeric transcripts have been implicated in cancers such as L1-ASP driven transcript in oesophageal adenocarcinoma and breast and colon cancers **(Cruickshanks and Tufarelli, 2009).** MCF-7 breast cancer cells and colon cancer cells, compared to normal, contained increased numbers of L1-chimeric transcripts **(Cruickshanks and Tufarelli, 2009)**. L1 activation has been associated with increased chimeric transcripts and thus can be used to predict cancer status **(Santourlidis et al., 1999).**

L1 enhances epithelial to mesenchymal transition (EMT)

Epithelial to mesenchymal transition (EMT) is a process by which epithelial characteristics such as differentiation, apical-basal polarity, and cell adhesion are lost and mesenchymal characteristics such as increased migratory capacity, resistance to apoptosis, and invasiveness and dysregulation of extracellular matrix components are gained **(Bonnans et al., 2014; Kalluri and Weinberg, 2010; Lee et al., 2012).** Recent studies have focused on the role of L1 in EMT **(Bojang et al., 2013)**. Our laboratory has shown that L1 can regulate EMT in HepG2 cells independent of RT activity **(Bojang et al., 2013)**. In this study, we showed that forced expression of wild type or RT mutant L1 decreased levels of epithelial markers such as E-Cadherin, N-Cadherin and Zonula occluding, while the levels of mesenchymal markers such as vimentin increased **(Bojang et al., 2013)**. Similarly, Apostolou et al., 2015 observed decreases in E-Cadherin and N-Cadherin levels post L1-ORF2 transfection and up-regulation of vimentin **(Apostolou et al., 2015)**, which strongly support L1 as a key player in EMT. Although the exact mechanism of how L1 mediates EMT is unknown, interpretation of the current data suggest L1 may directly potentiate the EMT process **(Lamouille et al., 2014)**. In Apostolou et al., 2015 suppression of ORF2 regulates the expression SOX2 transcription factor, whose nuclear translocation is suppressed by transforming growth factor β (TGF-β), a master regulator of the EMT process **(Apostolou et al., 2015; Xu et al., 2009; Zhu et al., 2013).** This study also showed, ORF2 over expression upregulated the expression of Snail 1 and 2 which are transcription factors known to down regulate epithelial markers such as E-cadherin **(Apostolou et al., 2015; Batlle et al., 2000),** further supporting L1s role as an important player in the EMT process Change to (**Reyes-Reyes., 2017).**

3. Clinical utilities of Line-1

L1 is associated with tumor initiation and progression through various mechanisms (Reviewed in Part 2). Thus, the aim of this review is to identify the potential of using L1 as a tool for early detection and prevention of cancer. Currently, there is a lack of previous studies that utilize L1 as an early biomarker for cancer risk, and is an area that still needs to be investigated.

3.1. *Biomarker*

In almost all somatic cells, L1 is silenced by DNA methylation at the promoter region and is re-activated during disease initiation and/or progression to signal genomic disturbance **(Rangasamy et al., 2015).** L1 methylation, DNA/RNA sequences and L1 encoded proteins can be detected using techniques such as quantitative real-time PCR (qPCR) **(Sunami et al., 2008)**, enzyme-linked immunosorbent assay (ELISA), pyrosequencing **(Aparicio et al., 2009)** and western blotting **(Asch et al., 1996).** Using serum or plasma samples, L1 methylation, RNA/DNA and protein can be screened and used as a diagnostic tool in cancer for cancer initiation and progression **(Figure 6)**.

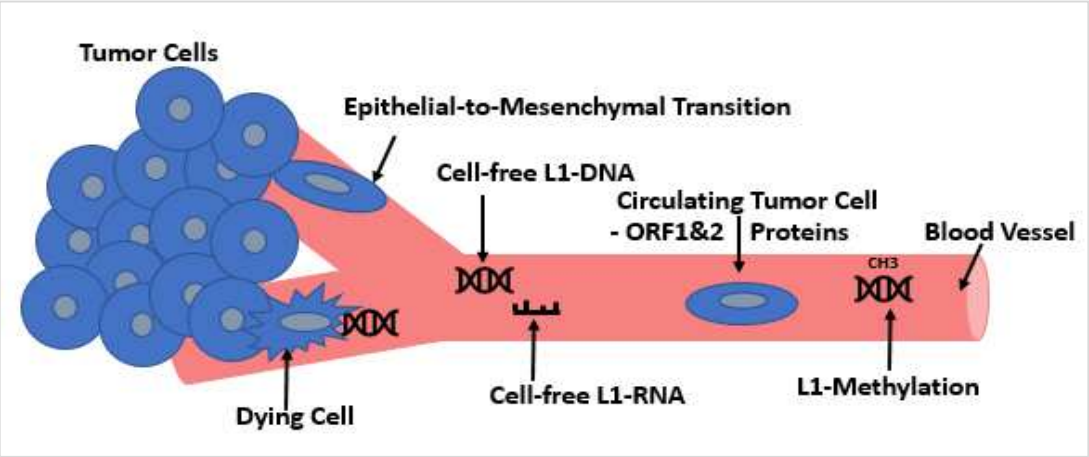

Figure 6: Schematic representation of the tumor microenvironment. Upper left corner shows a cell during epithelial-to-mesenchymal transition which results in the release of a circulating tumor

cell. Circulating L1 elements such as L1-DNA, L1-RNA are shown released from an apoptotic cell (lower left corner). The image indicates the potential of screening and measuring circulating L1-DNA, L1-RNA, L1 proteins and L1-DNA methylation status as a potential biomarker for cancer initiation and progression. Adapted and modified from **Wyatt et al. 2015**.

L1-DNA methylation

DNA methylation involves placement of a methyl group (-CH3) on the fifth position of the cytosine nucleotide ring **(Klose and Bird, 2006)**. Increased methylation is referred to as hypermethylation, while a loss of these methylation marks is known as hypomethylation. Pattamadilok et al., 2008 measured L1 methylation levels in ovarian cancer tissue samples and found that, of 27 patients with increased hypermethylation, only one had died, while out of 17 patients with increased hypomethylation, 12 had died **(Pattamadilok et al., 2008)**. Similarly, Ogino et al., 2008 measured L1 methylation levels of 643 colon cancer patients by pyrosequencing and found that increased L1 hypomethylation correlated with a decline in the five-year survival rate of these patients **(Ogino et al., 2008)**. Primary and secondary glioblastoma multiforme (GBMs, WHO grade 4) showed increased hypomethylation compared to normal brain tissue and low-grade gliomas (WHO grade 2) **(Ohka et al., 2011)**. Together, these findings suggest that L1 hypomethylation is associated with cancer progression and poor clinical outcome. Thus, L1 DNA methylation may have potential as a biomarker for cancer initiation and progression.

L1-DNA

DNA from normal apoptotic cells is uniform in size, whereas DNA from cancer cells vary in length **(Wang et al., 2003)**. DNA from circulating tumor cancer cells or ones released from cancer cells

can thus be differentiated from DNA from apoptotic cells by size and abundance. L1 sequences constitute 17% of the human genome which make it a compelling target to be used as a biomarker. In fact, Sunami et al., 2008 developed a technique to quantify circulating L1-DNA using a 300bp L1-DNA sequence by quantitative real-time PCR (qPCR) in breast cancer patients. They tested breast cancer patient serum and detected higher levels of long L1-DNA compared to normal breast tissue **(Sunami et al., 2008)**. Additionally, they found that L1-DNA levels correlated with tumor size **(Sunami et al., 2008)**. This demonstrates the feasibility of using circulating L1-DNA as a biomarker in the diagnosis of cancer progression **(Sunami et al., 2008)**.

L1-RNA

L1 uses an RNA intermediate to retrotranspose into new genomic locations. As mentioned previously, not all RNAs are involved in mobilization, some transcripts are involved in other functions (Reviewed in Part 2.1.). L1-RNA is detectable and can be traced back to the L1 it is transcribed from. Deininger and Belancio, 2016 described a method for tracing L1 mRNAs to actively transcribing L1 within the genome **(Deininger and Belancio, 2016)**. Furthermore, proteins encoded by these L1 transcripts, ORF1p and ORF2p are associated with many cancer types (Reviewed in L1-ORF1p & ORF2p). Thus, this indicates the potential of measuring L1-RNA as a biomarker for disease initiation and/or progression.

L1-ORF1p & ORF2p

ORF1 and ORF2 are L1 gene products necessary for retrotransposition. It is also important to note that L1 proteins expressed in cancer cells are detectable **(Rodić et al., 2014),** and thus have potential to be used as screening or diagnostic tools in cancer progression. L1-ORF2p is

overexpressed in transformed cell lines and epithelial cancer tissues (colon, prostate, lung and breast cancer) with no expression in non-transformed cells **(De Luca et al., 2016)**. The downregulation of L1-ORF2p by RNAi restores differentiation and reduces proliferation in cancer cells **(Oricchio et al., 2007; Sciamanna et al., 2005).** ORF1p expression also occurs in many invasive tumors **(Reyes-Reyes et al., 2016; Rodić et al., 2014)** such as breast cancer **(Harris et al., 2010)** and is associated with poor prognostic outcome **(Harris et al., 2010)**. L1-ORF1p upregulation has also been associated with cell proliferation, while downregulation with tumorigenesis **(Feng et al., 2013).** L1 induced proliferation in cancer cells may occur by interfering with signal transduction pathways, but the mechanism is unknown. Reyes-Reyes et al., 2016 identified an association between TGF-β signaling and the expression of ORF1p in hepatocellular carcinoma (HCC) **(Reyes-Reyes et al., 2016)**. Similarly, induction of EMT via TGF-β signaling and an increase in ORF1p and ORF2p was also observed in human epithelial cells (BEAS-2B) **(Reyes-Reyes et al., 2017)**. These findings indicate that L1 contributes to malignancy phenotypes in cancer cells via TGF-β signaling and may be involved in other signal transduction pathways. Furthermore, nuclear expression of both L1-ORF1p & ORF2p compared to cytoplasmic expression is associated with poor patient survival **(Chen et al., 2012)**. Thus, L1-ORF1p and ORF2p may be used as biomarkers for cancer progression, and translocation of L1 proteins into the nucleus may serve as a risk indicator of poor survival.

It should be noted that each cell-type may demonstrate different levels of expression of L1-DNA, L1-RNA, or L1-ORF1 & 2 proteins. Our lab and others have found that L1 is reactivated in tumors of epithelial origin **(Teneng et al., 2007)**. The use of these L1 elements as potential biomarkers further opens new avenues for diagnostic and therapeutic intervention but must be tailored to the tumors whose initiation and expression are associated with L1 reactivation.

4. Targeting line-1 for treatment of human diseases

The activities of L1 play a prominent role in the initiation and progression of epithelial derived tumors and tumors initiated by environmental carcinogens. In addition, certain drugs that target the epigenetic machinery are prone to reactivate L1 during treatments. Currently, no drug is available that targets the activities of L1. Drugs can be used to target the reverse transcriptase activity of L1ORF2, phosphorylation of L1-ORF1, regulatory sequence of L1 such as the g-quadruplex sequence within L1s 3'UTR and genes that regulate L1 expression, such as the aryl hydrocarbon receptor.

4.1. *Reverse transcriptase inhibitors*

L1 retrotransposition events are correlated with genomic alterations and the onset of tumorigenesis **(Spadafora, 2015).** Reverse transcriptase (RT) is a key player in retrotransposition and functions by transcribing L1 mRNA at the integration sites **(Spadafora, 2015).** Sinibaldi-Vallebona et al., 2011 showed that targeting the RT-activity with efavirenz and nevirapine promotes differentiation and reduces proliferation of cancer cells in culture and animal models **(Sinibaldi-Vallebona et al., 2011)**. Furthermore, mice xenografted with human tumorigenic cells treated with efavirenz showed delayed tumor progression that reversed upon discontinuation **(Sciamanna et al., 2005)**. Together, these findings suggest the potential of targeting L1 RT activity with RT inhibitory drugs alone, or in combination therapy to treat tumors with aberrant L1 expression.

4.2. *Kinase inhibitors*

L1 encoded ORF1 protein is required for retrotransposition **(Cook et al., 2015).** ORF1p phosphorylation is also required for L1 retrotransposition by protein kinase A (PKA) **(Bojang et**

al., 2013; Cook et al., 2015). ORF1 contain target sites for proline-directed protein kinase (PDPK) phosphorylation and a mutation of these sites has been associated with inhibition of retrotransposition **(Cook et al., 2015)**. Currently, there are no kinase inhibitors specifically designed to target PKA on the market, so there is room for further drug development studies focusing on targeting and inhibiting PKA. Thus, we suggest that targeting L1 ORF1 phosphorylation with kinase inhibitors may be a novel method for inhibiting L1 activity in cancer.

4.3. *AHR inhibitors*

Aryl hydrocarbon receptor (AHR) is a ligand activated transcription factor **(Teneng et al., 2007; Reyes-Reyes et al. 2016).** Previous studies have shown an association between increased levels of AHR ligand activated by benzo (a) pyrene (BaP) and higher levels of ORF1 mRNA in Hela cells **(Teneng et al., 2007)** and TGF-β1 mRNA expression HepG2 cells **(Reyes-Reyes et al., 2016)**. Thus, AHR signaling may be associated with the activation of L1. Furthermore, AHR has been found to be overexpressed in breast cancer **(Powell et al. 2013).** Lai et al., 2014 studied the effects of biseugenol, a novel AHR inhibitor, in gastric cancer cells (MKN45)**(Lai et al., 2014)**. They found that biseugenol impeded cancer growth and inhibited EMT **(Lai et al., 2014)**. These findings together suggest that targeting AHR with AHR inhibitors has the potential to inhibit the activities of L1 in tumors.

4.4. *G-quadruplex targeting drugs*

L1 can form secondary structures known as G-quadruplex DNA or G4 structures. G4 structures are formed by folding of guanine-rich sequences **(Howell et al. 1997, Lexa et al.2014)** which are held by hoogsteen base pairing **(Bochman et al. 2012).** Sahakyan et al., 2017, stabilized G4

structures by targeting them with small molecule ligands. They observed a 30% reduction in L1 retrotransposition when G4 structures were mutated or deleted **(Sahakyan et al., 2017)**. Thus, targeting G-quadruplex structures via drugs such as berberine may be used to curtail L1 activity in tumors.

5. Concluding Remarks

L1 retrotransposition events are correlated with initiation and progression of many diseases, especially cancers of epithelial origin (Table 1). The life cycle of L1 (Figure 1) which is described as a retrotransposition dependent mode can be tracked to determine disease initiation and progression. Also, because L1 elements can be detected in circulating blood; they can be used as a novel biomarker for cancer initiation and progression. Furthermore, this review has addressed different ways of inhibiting L1 activity as a potential therapeutic option.

For the future, a clinical study analyzing the effects of these drugs (Reviewed in Part 4) on L1 associated tumors should be conducted, as our main goal is to alleviate L1s harmful effects and to increase a patient's survival and quality of life. Also, as mentioned previously there are currently no kinase inhibitors that specifically target PKA on the market. Thus, there is need a for further drug development studies that focus specifically on targeting PKA.

Prospectively, the long-term effects of some drugs should also be analyzed, as the molecules they target may play a role in other cellular processes. For example, AHR is also associated with neurogenesis, so theoretically inhibiting AHR with AHR inhibitors may disrupt neuron development and/or interfere with other signaling pathways.

Also, the potential of using L1 as a biomarker for cancer initiation and progression has been shown in this review. Thus, a randomized control clinical trial is needed to look at the efficacy of L1 in the early diagnosis of L1 associated cancers.

Furthermore, future research should focus on ORF0 which is known to be involved in gene regulatory changes by producing chimeric transcripts that have the potential of encoding novel proteins. So, understanding how ORF0 transcripts may contribute to both normal or evolutionary terms and disease states still needs to be investigated.

As mentioned previously L1 is associated with EMT but the mechanism of how L1 contributes to cancer related phenotypes such as proliferation and dedifferentiation in cancer cells is still unknown. Also, previous studies have focused only on the expression levels of ORF1 and ORF2 proteins in different cancers, but the role these proteins play in signal transduction is an area that still needs to be studied. Furthermore, as growth factors play an important role in cancer cell proliferation, it can be speculated that ORF1 and ORF2 may play a role in cancer cell proliferation via the phosphatidylinositol 3-kinase (P13K) and MAP kinase signaling pathways.

References

Anifandis, G., Messini, C., Dafopoulos, K., and Messinis, I. (2015). Genes and conditions controlling mammalian pre-and post-implantation embryo development. Current genomics *16*, 32-46.

Aparicio, A., North, B., Barske, L., Wang, X., Bollati, V., Weisenberger, D., Yoo, C., Tannir, N., Horne, E., and Groshen, S. (2009). LINE-1 methylation in plasma DNA as a biomarker of activity of DNA methylation inhibitors in patients with solid tumors. Epigenetics *4*, 176-184.

Apostolou, P., Toloudi, M., Chatziioannou, M., Kourtidou, E., Mimikakou, G., Vlachou, I., Chlichlia, A., and Papasotiriou, I. (2015). Involvement of retrotransposon L1 in stemness and cellular plasticity. Cell communication & adhesion *22*, 1-7.

Asch, H.L., Eliacin, E., Fanning, T.G., Connolly, J.L., Bratthauer, G., and Asch, B.B. (1996). Comparative expression of the LINE-1 p40 protein in human breast carcinomas and normal breast tissues. Oncology Research Featuring Preclinical and Clinical Cancer Therapeutics *8*, 239-247.

Ashe, H.L., Monks, J., Wijgerde, M., Fraser, P., and Proudfoot, N.J. (1997). Intergenic transcription and transinduction of the human β-globin locus. Genes & development *11*, 2494-2509.

Batlle, E., Sancho, E., Francí, C., Domínguez, D., Monfar, M., Baulida, J., and de Herreros, A.G. (2000). The transcription factor snail is a repressor of E-cadherin gene expression in epithelial tumour cells. Nature cell biology *2*, 84-89.

Boise, L.H., González-Garcia, M., Postema, C.E., Ding, L., Lindsten, T., Turka, L.A., Mao, X., Nuñez, G., and Thompson, C.B. (1993). bcl-x, a bcl-2-related gene that functions as a dominant regulator of apoptotic cell death. cell *74*, 597-608.

Bojang, P., Roberts, R.A., Anderton, M.J., and Ramos, K.S. (2013). Reprogramming of the HepG2 genome by long interspersed nuclear element□1. Molecular oncology *7*, 812-825.

Bolland, D.J., Wood, A.L., Johnston, C.M., Bunting, S.F., Morgan, G., Chakalova, L., Fraser, P.J., and Corcoran, A.E. (2004). Antisense intergenic transcription in V (D) J recombination. Nature immunology *5*, 630-637.

Bonnans, C., Chou, J., and Werb, Z. (2014). Remodelling the extracellular matrix in development and disease. Nature reviews Molecular cell biology *15*, 786-801.

Brennecke, J., Aravin, A.A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., and Hannon, G.J. (2007). Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. Cell *128*, 1089-1103.

Brower-Toland, B., Findley, S.D., Jiang, L., Liu, L., Yin, H., Dus, M., Zhou, P., Elgin, S.C., and Lin, H. (2007). Drosophila PIWI associates with chromatin and interacts directly with HP1a. Genes & development *21*, 2300-2311.

Brown, T.A. (2002). Synthesis and processing of RNA.

Chen, J., and Weiss, W. (2015). Alternative splicing in cancer: implications for biology and therapy. Oncogene *34*, 1-14.

Chen, L., Dahlstrom, J.E., Lee, S.-H., and Rangasamy, D. (2012). Naturally occurring endo-siRNA silences LINE-1 retrotransposons in human cells through DNA methylation. Epigenetics *7*, 758-771.

Cheng, D.-L., Xiang, Y.-Y., Ji, L.-j., and Lu, X.-J. (2015). Competing endogenous RNA interplay in cancer: mechanism, methodology, and perspectives. Tumor Biology *36*, 479-488.

Cook, P.R., Jones, C.E., and Furano, A.V. (2015). Phosphorylation of ORF1p is required for L1 retrotransposition. Proceedings of the National Academy of Sciences *112*, 4298-4303.

Cruickshanks, H.A., and Tufarelli, C. (2009). Isolation of cancer-specific chimeric transcripts induced by hypomethylation of the LINE-1 antisense promoter. Genomics *94*, 397-406.

Czech, B., and Hannon, G.J. (2016). One loop to rule them all: the ping-pong cycle and piRNA-guided silencing. Trends in biochemical sciences *41*, 324-337.

De Fazio, S., Bartonicek, N., Di Giacomo, M., Abreu-Goodger, C., Sankar, A., Funaya, C., Antony, C., Moreira, P.N., Enright, A.J., and O'Carroll, D. (2011). The endonuclease activity of Mili fuels piRNA amplification that silences LINE1 elements. Nature *480*, 259-263.

de Laat, W.L., Appeldoorn, E., Jaspers, N.G., and Hoeijmakers, J.H. (1998). DNA structural elements required for ERCC1-XPF endonuclease activity. Journal of Biological Chemistry *273*, 7835-7842.

De Luca, C., Guadagni, F., Sinibaldi-Vallebona, P., Sentinelli, S., Gallucci, M., Hoffmann, A., Schumann, G.G., Spadafora, C., and Sciamanna, I. (2016). Enhanced expression of LINE-1-encoded ORF2 protein in early stages of colon and prostate transformation. Oncotarget *7*, 4048.

Deininger, P., and Belancio, V.P. (2016). Detection of LINE-1 RNAs by northern blot. Transposons and Retrotransposons: Methods and Protocols, 223-236.

Deininger, P.L., Moran, J.V., Batzer, M.A., and Kazazian, H.H. (2003). Mobile elements and mammalian genome evolution. Current opinion in genetics & development *13*, 651-658.

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

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

Feng, F., Lu, Y.-Y., Zhang, F., Gao, X.-D., Zhang, C.-F., Meredith, A., Xu, Z.-X., Yang, Y.-T., Chang, X.-J., and Wang, H. (2013). Long interspersed nuclear element ORF-1 protein promotes proliferation and resistance to chemotherapy in hepatocellular carcinoma. World Journal of Gastroenterology: WJG *19*, 1068.

Gasior, S.L., Roy-Engel, A.M., and Deininger, P.L. (2008). ERCC1/XPF limits L1 retrotransposition. DNA repair *7*, 983-989.

Gribnau, J., Diderich, K., Pruzina, S., Calzolari, R., and Fraser, P. (2000). Intergenic transcription and developmental remodeling of chromatin subdomains in the human β-globin locus. Molecular cell *5*, 377-386.

Guo, H., Chitiprolu, M., Gagnon, D., Meng, L., Perez-Iratxeta, C., Lagace, D., and Gibbings, D. (2014). Autophagy supports genomic stability by degrading retrotransposon RNA. Nature communications *5*, ncomms6276.

Hammond, S.M., Boettcher, S., Caudy, A.A., Kobayashi, R., and Hannon, G.J. (2001). Argonaute2, a link between genetic and biochemical analyses of RNAi. Science *293*, 1146-1150.

Han, J.S., and Boeke, J.D. (2005). LINE□1 retrotransposons: Modulators of quantity and quality of mammalian gene expression? Bioessays *27*, 775-784.

Han, J.S., Szak, S.T., and Boeke, J.D. (2004). Transcriptional disruption by the L1 retrotransposon and implications for mammalian transcriptomes. Nature *429*, 268-274.
Hancks, D.C., and Kazazian, H.H. (2010). SVA retrotransposons: Evolution and genetic instability. In Seminars in cancer biology (Elsevier), pp. 234-245.
Harris, C.R., Normart, R., Yang, Q., Stevenson, E., Haffty, B.G., Ganesan, S., Cordon-Cardo, C., Levine, A.J., and Tang, L.H. (2010). Association of nuclear localization of a long interspersed nuclear element-1 protein in breast tumors with poor prognostic outcomes. Genes & cancer *1*, 115-124.
Harris, R.S., and Liddament, M.T. (2004). Retroviral restriction by APOBEC proteins. Nature Reviews Immunology *4*, 868-877.
Houseley, J., and Tollervey, D. (2009). The many pathways of RNA degradation. Cell *136*, 763-776.
Hu, S., Li, J., Xu, F., Mei, S., Le Duff, Y., Yin, L., Pang, X., Cen, S., Jin, Q., and Liang, C. (2015). SAMHD1 inhibits LINE-1 retrotransposition by promoting stress granule formation. PLoS genetics *11*, e1005367.
Huang, X.A., Yin, H., Sweeney, S., Raha, D., Snyder, M., and Lin, H. (2013). A major epigenetic programming mechanism guided by piRNAs. Developmental cell *24*, 502-516.
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 research *39*, 5538-5554.
Ishiuchi, T., and Torres-Padilla, M.-E. (2014). LINEing germ and embryonic stem cells' silencing of retrotransposons. Genes & development *28*, 1381-1383.
Kaer, K., Branovets, J., Hallikma, A., Nigumann, P., and Speek, M. (2011). Intronic L1 retrotransposons and nested genes cause transcriptional interference by inducing intron retention, exonization and cryptic polyadenylation. PLoS One *6*, e26099.
Kalluri, R., and Weinberg, R.A. (2010). The basics of epithelial-mesenchymal transition. The Journal of clinical investigation *120*, 1786.
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.
Kim, S., Kim, Y.-J., and Han, K. (2016). Analysis of L1-chimeric transcripts derived from bidirectional promoter of human-specific L1. Genes & Genomics *38*, 69-79.
Kiriakidou, M., Tan, G.S., Lamprinaki, S., De Planell-Saguer, M., Nelson, P.T., and Mourelatos, Z. (2007). An mRNA m 7 G cap binding-like motif within human Ago2 represses translation. Cell *129*, 1141-1151.
Kleckner, N. (1990). Regulation of transposition in bacteria. Annual review of cell biology *6*, 297-327.
Klose, R.J., and Bird, A.P. (2006). Genomic DNA methylation: the mark and its mediators. Trends in biochemical sciences *31*, 89-97.
Kuwabara, T., Hsieh, J., Muotri, A., Yeo, G., Warashina, M., Lie, D.C., Moore, L., Nakashima, K., Asashima, M., and Gage, F.H. (2009). Wnt-mediated activation of NeuroD1 and retro-elements during adult neurogenesis. Nature neuroscience *12*, 1097-1105.
Lai, D.-W., Liu, S.-H., Karlsson, A.I., Lee, W.-J., Wang, K.-B., Chen, Y.-C., Shen, C.-C., Wu, S.-M., Liu, C.-Y., and Tien, H.-R. (2014). The novel Aryl hydrocarbon receptor inhibitor

biseugenol inhibits gastric tumor growth and peritoneal dissemination. Oncotarget *5*, 7788.
Lamouille, S., Xu, J., and Derynck, R. (2014). Molecular mechanisms of epithelial–mesenchymal transition. Nature reviews Molecular cell biology *15*, 178-196.
Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., and FitzHugh, W. (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860-921.
Le Thomas, A., Rogers, A.K., Webster, A., Marinov, G.K., Liao, S.E., Perkins, E.M., Hur, J.K., Aravin, A.A., and Tóth, K.F. (2013). Piwi induces piRNA-guided transcriptional silencing and establishment of a repressive chromatin state. Genes & development *27*, 390-399.
Lee, E., Iskow, R., Yang, L., Gokcumen, O., Haseley, P., Luquette, L.J., Lohr, J.G., Harris, C.C., Ding, L., and Wilson, R.K. (2012). Landscape of somatic retrotransposition in human cancers. Science *337*, 967-971.
Levin, H.L., and Moran, J.V. (2011). Dynamic interactions between transposable elements and their hosts. Nature Reviews Genetics *12*, 615-627.
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. Journal of Biological Chemistry *288*, 21148-21160.
Matsuo, M., Masumura, T., Nishio, H., Nakajima, T., Kitoh, Y., Takumi, T., Koga, J., and Nakamura, H. (1991). Exon skipping during splicing of dystrophin mRNA precursor due to an intraexon deletion in the dystrophin gene of Duchenne muscular dystrophy kobe. Journal of Clinical Investigation *87*, 2127.
Meischl, C., de Boer, M., Åhlin, A., and Roos, D. (2000). A new exon created by intronic insertion of a rearranged LINE-1 element as the cause of chronic granulomatous disease. European Journal of Human Genetics *8*.
Miki, Y., Nishisho, I., Horii, A., Miyoshi, Y., Utsunomiya, J., Kinzler, K.W., Vogelstein, B., and Nakamura, Y. (1992). Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer. Cancer research *52*, 643-645.
Miousse, I.R., and Koturbash, I. (2015). The fine LINE: methylation drawing the cancer landscape. BioMed research international *2015*.
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. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis *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.
Motamedi, M.R., Verdel, A., Colmenares, S.U., Gerber, S.A., Gygi, S.P., and Moazed, D. (2004). Two RNAi complexes, RITS and RDRC, physically interact and localize to noncoding centromeric RNAs. Cell *119*, 789-802.
Muñoz-López, M., and García-Pérez, J.L. (2010). DNA transposons: nature and applications in genomics. Current genomics *11*, 115-128.
Narita, N., Nishio, H., Kitoh, Y., Ishikawa, Y., Ishikawa, Y., Minami, R., Nakamura, H., and Matsuo, M. (1993). Insertion of a 5'truncated L1 element into the 3'end of exon 44 of the

dystrophin gene resulted in skipping of the exon during splicing in a case of Duchenne muscular dystrophy. Journal of Clinical Investigation *91*, 1862.
Neidhart, M., Karouzakis, E., Schumann, G.G., Gay, R.E., and Gay, S. (2010). Trex□1 deficiency in rheumatoid arthritis synovial fibroblasts. Arthritis & Rheumatology *62*, 2673-2679.
Noma, K.-i., Sugiyama, T., Cam, H., Verdel, A., Zofall, M., Jia, S., Moazed, D., and Grewal, S.I. (2004). RITS acts in cis to promote RNA interference–mediated transcriptional and post-transcriptional silencing. Nature genetics *36*, 1174-1180.
Ogino, S., Nosho, K., Kirkner, G.J., Kawasaki, T., Chan, A.T., Schernhammer, E.S., Giovannucci, E.L., and Fuchs, C.S. (2008). A cohort study of tumoral LINE-1 hypomethylation and prognosis in colon cancer. JNCI: Journal of the National Cancer Institute *100*, 1734-1738.
Ohka, F., Natsume, A., Motomura, K., Kishida, Y., Kondo, Y., Abe, T., Nakasu, Y., Namba, H., Wakai, K., and Fukui, T. (2011). The global DNA methylation surrogate LINE-1 methylation is correlated with MGMT promoter methylation and is a better prognostic factor for glioma. PloS one *6*, e23332.
Ohms, S., Lee, S.-H., and Rangasamy, D. (2014). LINE-1 retrotransposons and let-7 miRNA: partners in the pathogenesis of cancer? Frontiers in genetics *5*.
Oricchio, E., Sciamanna, I., Beraldi, R., Tolstonog, G., Schumann, G., and Spadafora, C. (2007). Distinct roles for LINE-1 and HERV-K retroelements in cell proliferation, differentiation and tumor progression. Oncogene *26*, 4226-4233.
Pattamadilok, J., Huapai, N., Rattanatanyong, P., Vasurattana, A., Triratanachat, S., Tresukosol, D., and Mutirangura, A. (2008). LINE□1 hypomethylation level as a potential prognostic factor for epithelial ovarian cancer. International Journal of Gynecological Cancer *18*, 711-717.
Pezic, D., Manakov, S.A., Sachidanandam, R., and Aravin, A.A. (2014). piRNA pathway targets active LINE1 elements to establish the repressive H3K9me3 mark in germ cells. Genes & development *28*, 1410-1428.
Rangasamy, D., Lenka, N., Ohms, S., Dahlstrom, J., Blackburn, A., and Board, P. (2015). Activation of LINE-1 retrotransposon increases the risk of epithelial-mesenchymal transition and metastasis in epithelial cancer. Current molecular medicine *15*, 588-597.
Reyes-Reyes, E.M., Aispuro, I.,Tavera-Garcia, M.A.,Field, M., Moore, S., Ramos, I., and Ramos, K.S. (2017). LINE-1 couples EMT programming with acquisition of oncogenic phenotypes in human bronchial epithelial cells. Oncotarget *8*, 103828-103842.
Reyes-Reyes, E.M., Ramos, I.N., Tavera-Garcia, M.A., and Ramos, K.S. (2016). The aryl hydrocarbon receptor agonist benzo (a) pyrene reactivates LINE-1 in HepG2 cells through canonical TGF-β1 signaling: implications in hepatocellular carcinogenesis. American journal of cancer research *6*, 1066.
Rodić, N., and Burns, K.H. (2013). Long interspersed element–1 (LINE-1): passenger or driver in human neoplasms? PLoS genetics *9*, e1003402.
Rodić, N., Sharma, R., Sharma, R., Zampella, J., Dai, L., Taylor, M.S., Hruban, R.H., Iacobuzio-Donahue, C.A., Maitra, A., and Torbenson, M.S. (2014). Long interspersed element-1 protein expression is a hallmark of many human cancers. The American journal of pathology *184*, 1280-1286.
Roman-Gomez, J., Jimenez-Velasco, A., Agirre, X., Prosper, F., Heiniger, A., and Torres, A. (2005). Lack of CpG island methylator phenotype defines a clinical subtype of T-cell

acute lymphoblastic leukemia associated with good prognosis. Journal of Clinical Oncology *23*, 7043-7049.
Ruthenburg, A.J., Allis, C.D., and Wysocka, J. (2007). Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark. Molecular cell *25*, 15-30.
Sahakyan, A.B., Murat, P., Mayer, C., and Balasubramanian, S. (2017). G-quadruplex structures within the 3 [prime] UTR of LINE-1 elements stimulate retrotransposition. Nature Structural & Molecular Biology *24*, 243-247.
Samuelov, L., Fuchs□Telem, D., Sarig, O., and Sprecher, E. (2011). An exceptional mutational event leading to Chanarin–Dorfman syndrome in a large consanguineous family. British Journal of Dermatology *164*, 1390-1392.
Santourlidis, S., Florl, A., Ackermann, R., Wirtz, H.C., and Schulz, W.A. (1999). High frequency of alterations in DNA methylation in adenocarcinoma of the prostate. The Prostate *39*, 166-174.
Savas, S., Tuzmen, S., and Ozcelik, H. (2006). Human SNPs resulting in premature stop codons and protein truncation. Human genomics *2*, 274.
Schwahn, U., Lenzner, S., Dong, J., Feil, S., Hinzmann, B., van Duijnhoven, G., Kirschner, R., Hemberger, M., Bergen, A.A., and Rosenberg, T. (1998). Positional cloning of the gene for X-linked retinitis pigmentosa 2. Nature genetics *19*, 327-332.
Schwarz, D.S., Tomari, Y., and Zamore, P.D. (2004). The RNA-induced silencing complex is a Mg 2+-dependent endonuclease. Current Biology *14*, 787-791.
Sciamanna, I., Landriscina, M., Pittoggi, C., Quirino, M., Mearelli, C., Beraldi, R., Mattei, E., Serafino, A., Cassano, A., and Sinibaldi-Vallebona, P. (2005). Inhibition of endogenous reverse transcriptase antagonizes human tumor growth. Oncogene *24*, 3923-3931.
Shukla, R., Upton, K.R., Muñoz-Lopez, M., Gerhardt, D.J., Fisher, M.E., Nguyen, T., Brennan, P.M., Baillie, J.K., Collino, A., and Ghisletti, S. (2013). Endogenous retrotransposition activates oncogenic pathways in hepatocellular carcinoma. Cell *153*, 101-111.
Sigurdsson, M.I., Smith, A.V., Bjornsson, H.T., and Jonsson, J.J. (2012). The distribution of a germline methylation marker suggests a regional mechanism of LINE-1 silencing by the piRNA-PIWI system. BMC genetics *13*, 31.
Sinibaldi-Vallebona, P., Matteucci, C., and Spadafora, C. (2011). Retrotransposon-encoded reverse transcriptase in the genesis, progression and cellular plasticity of human cancer. Cancers *3*, 1141-1157.
Smit, A.F. (1999). Interspersed repeats and other mementos of transposable elements in mammalian genomes. Current opinion in genetics & development *9*, 657-663.
Soifer, H.S., and Rossi, J.J. (2006). Small interfering RNAs to the rescue: blocking L1 retrotransposition. Nature structural & molecular biology *13*, 758-759.
Spadafora, C. (2015). A LINE□1–encoded reverse transcriptase–dependent regulatory mechanism is active in embryogenesis and tumorigenesis. Annals of the New York Academy of Sciences *1341*, 164-171.
Speek, M. (2001). Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes. Molecular and cellular biology *21*, 1973-1985.
Spengler, R.M., Oakley, C.K., and Davidson, B.L. (2013). Functional microRNAs and target sites are created by lineage-specific transposition. Human molecular genetics *23*, 1783-1793.

Stribinskis, V., and Ramos, K.S. (2006). Activation of human long interspersed nuclear element 1 retrotransposition by benzo (a) pyrene, an ubiquitous environmental carcinogen. Cancer research *66*, 2616-2620.

Sunami, E., Vu, A.T., Nguyen, S.L., Giuliano, A.E., and Hoon, D.S. (2008). Quantification of LINE1 in circulating DNA as a molecular biomarker of breast cancer. Annals of the New York Academy of Sciences *1137*, 171-174.

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

Teneng, I., Montoya-Durango, D.E., Quertermous, J.L., Lacy, M.E., and Ramos, K.S. (2011). Reactivation of L1 retrotransposon by benzo (a) pyrene involves complex genetic and epigenetic regulation. Epigenetics *6*, 355-367.

Teneng, I., Stribinskis, V., and Ramos, K.S. (2007). Context□specific regulation of LINE□1. Genes to Cells *12*, 1101-1110.

Thammaiah, C.K., and Jayaram, S. (2016). Role of let-7 family microRNA in breast cancer. Non-coding RNA Research *1*, 77-82.

Ting, C.-N., Rosenberg, M., Snow, C., Samuelson, L., and Meisler, M. (1992). Endogenous retroviral sequences are required for tissue-specific expression of a human salivary amylase gene. Genes & Development *6*, 1457-1465.

Wang, B.G., Huang, H.-Y., Chen, Y.-C., Bristow, R.E., Kassauei, K., Cheng, C.-C., Roden, R., Sokoll, L.J., Chan, D.W., and Shih, I.-M. (2003). Increased plasma DNA integrity in cancer patients. Cancer research *63*, 3966-3968.

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

Xu, J., Lamouille, S., and Derynck, R. (2009). TGF-β-induced epithelial to mesenchymal transition. Cell research *19*, 156-172.

Yang, L., Chen, D., Duan, R., Xia, L., Wang, J., Qurashi, A., Jin, P., and Chen, D. (2007). Argonaute 1 regulates the fate of germline stem cells in Drosophila. Development *134*, 4265-4272.

Yang, N., and Kazazian, H.H. (2006). L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells. Nature structural & molecular biology *13*, 763-771.

Yoshida, K., Nakamura, A., Yazaki, M., Ikeda, S.-i., and Takeda, S.i. (1998). Insertional mutation by transposable element, L1, in the DMD gene results in X-linked dilated cardiomyopathy. Human molecular genetics *7*, 1129-1132.

Zhang, F., Wang, J., Xu, J., Zhang, Z., Koppetsch, B.S., Schultz, N., Vreven, T., Meignin, C., Davis, I., and Zamore, P.D. (2012). UAP56 couples piRNA clusters to the perinuclear transposon silencing machinery. Cell *151*, 871-884.

Zhu, Y., Feng, F., Yu, J., Song, B., Hu, M., Gao, X., Wang, Y., and Zhang, Q. (2013). L1-ORF1p, a Smad4 interaction protein, promotes proliferation of HepG2 cells and tumorigenesis in mice. DNA and cell biology *32*, 531-540.