Insights on the molecular mechanisms of SINEUP activity

Thesis submitted for the degree of

"Doctor Philosophiae"

SISSA- International School of Advanced Studies

Area of Neuroscience- Curriculum in Joint Molecular Biology

CANDIDATE **SUPERVISOR**

Abraham Tettey Matey **Prof. Stefano Gustincich**

Academic year 2017/2018

DECLARATION

The research works reported in this thesis including *in silico* mapping of cis-acting regulatory elements predicted in SINEUP-EDs were carried out at SISSA (International School of Advanced Studies) in Trieste, between the period of November 2014 and October 2018.

All my reported data on SINEUPs and IRES in this thesis are yet to be published.

TABLE OF CONTENTS

ABSTRACT

Translation initiation during protein synthesis is one of the rate limiting steps in regulating gene expression in eukaryotes. It may occur through a cap- dependent or independent mechanism. Cap-independent translation initiation usually takes place when the canonical cap-dependent process involving the cap-eIF4F complex and/or the ternary complex (TC) are inhibited, mostly during cellular stress. It can be activated by modular RNA elements like internal ribosome entering sites (IRESs) that can act alone or in conjunction with other RNA cis-acting elements such as upstream open reading frames (uORF), terminal oligopyrimidines (TOPs) and N6-methyladenosine (m^6A).

Recently, antisense (AS) long non-coding RNAs (lncRNAs) to mouse and human gene targets were reported to up-regulate translation of those genes, representing a new functional class of lncRNAs termed SINEUPs, for SINE sequences UP-regulating translation. SINEUP modular organization consists of a binding domain (BD) which specifically targets the mRNA of interest by an antisense sequence to its 5'UTR, and an effector domain (ED) constituted by an embedded SINE sequence which confers the biological function of translation activation. However, the underlying molecular mechanism mediated by SINEUPs is unknown.

Here, I show using both SINEUP- and IRES-activity assays that the ED of AS Uchl1, the lncRNA representative member of SINEUPs, promotes cap-independent translation through an internal ribosome entry site that is mediated by the inverted SINEB2 stem loop 1 (SL1) hairpin structure, which probably recruits the ribosome using an IRES-like mechanism similar to the one in Hepatitis Virus C.

Since synthetic SINEUPs can be targeted to potentially any gene of interest by swapping the BDs, they can be used as molecular tools, in protein manufacturing and RNA therapeutics to increase protein synthesis.

Therefore, in this work I optimized a miniaturized version, the synthetic microSINEUP, and showed for the first time that SINEUPs are functional in *Drosophila* cells.

In addition, taking advantage of SINEUP modular organization, I synthesized IRES-containing SINEUPs, called SINEUP-IRESs, that present cellular and viral IRESs acting as ED in SINEUP RNA molecules and therefore proving that IRES sequences can activate translation in *trans* through a BD. By the study of IRES-containing c-myc mRNA, I discovered the ability of natural IRES elements to increase endogenously expressed, targeted proteins synthesis in *trans*.

These data are the first evidence for a potential new molecular mechanism of gene expression control with long-range consequences in health and disease.

ABBREVIATIONS

Apaf-1	Apoptotic protease-activating factor 1
AS	Antisense
AS BACE-1	AS B-secretase-1
BACE-1	B-secretase-1
BD	binding domain
bp	base-pair
cDNA	complementary DNA
ceRNA	Competing endogenous RNA
CDS	coding sequence
cMyc	avian myelocytomatosis virus oncogene cellular homolog
CrPV	Cricket paralysis virus
DNA	deoxy-ribonucleic acid
DMD	Duchenne muscular dystrophy
dsRNA	double strand RNA
ED	effector domain
eIF	eukaryotic initiation factors
eEF	eukaryotic elongation factors
ELG1	Enhanced Level of Genomic-instability 1
EMCV	encephalomyocarditis virus
ENCODE	Encyclopedia of DNA elements
eRNA	enhancer RNA
FANTOM	Functional Annotation of the Mammalian Genome
FBS	fetal bovine serum
FCS	fetal calf serum
GCN	general control of amino acid biosynthesis protein
Gtx	murine homeobox gene
HCV	hepatitis B virus
HRP	Horseradish peroxidase
IRES	internal ribosome entry sites
ITAF	IRES trans-acting factors

LINE	long interspersed elements
lncRNA	Long non-coding RNA
LTR	Long terminal repeat
miRNA	microRNA
mRNA	Messenger RNA
m^6A	N6-methyl adenosine
m^7G	N7-methyl guanosine
ncRNA	non-coding RNA
nt	nucleotide
ORF	open reading frame
Polio	Poliomyelitis virus
Pol II	polymerase II
Pol III	polymerase III
qRT-PCR	quantitative real time PCR
TOP	terminal oligopyrimidine tracts
RBP	RNA binding proteins
RNA pol	RNA polymerase
RNP	ribonucleoprotein
rRNA	ribosomal RNA
RT-PCR	reverse transcriptase polymerase chain reaction
TC	ternary complex
TE	transposable elements
S	Sense
SINE	Short interspersed elements
SINEUP	SINE UP-regulation
siRNA	Short-interfering RNA
SL1	stem loop 1
snoRNA	Small nucleolar RNA
snRNA	Small nuclear RNA
ssRNA	Single strand RNA
TE	transposable elements

tRNA	Transfer RNA
TOPs	terminal oligopyrimidines
Uchl1	Ubiquitin carboxyl-terminal hydrolase 1
uORF	upstream open reading frames
UTR	Untranslated regions
WT	Wild-type
WB	Western blot

1.1 INTRODUCTION

Advancement in genomic and transcriptomic technologies have led to the realization of pervasive transcription in the mouse (The FANTOM Consortium, Carninci et al., 2005; Hayashizaki and Carninci, 2006) and human genome (The ENCODE consortium, Djebali, et al., 2012) ***(Figure 1)***, paving ways for the extensive characterization of the mammalian transcriptome. Using genomic sequencing combined with cap analysis of gene expression-CAGE protocol (Shiraki et al., 2003), the FANTOM consortium, Carninci et al., 2005 reported that over 63% of the mouse genome produces transcripts, while 74.7% was reported in the human, using RNA-seq and profiling technologies (The ENCODE consortium, Djebali, et al., 2012). However, not so much on the functionality and regulations of gene expression could be comprehended. Extensive use of RNA profiling and functional studies technologies in the ENCODE projects (The ENCODE Project Consortium, Cheng et al., 2014; Derrien et al., 2012; Yue et al., 2014) have however provided further information on the biochemical function and new insights into the mechanisms of mammalian genes transcription, differences and similarities in RNA expression and evolutionary conserved genomic regulations between the mouse and human genome. The ENCODE project Derrien et al., 2012 show that about 80.4% of the human genome participate in at least one biochemical function, including RNA- and/or chromatin-associated events. Subsequent studies revealed that the state and structure of chromatin, controlling genomic replication, are relatively stable and conserved in evolution between both genomes, but the cis-regulatory landscape has diverged between mouse and human (Yue, et al., 2014). Also, Cheng et al., 2014 reported that transcription from orthologous promoters in mouse and human usually give rise to similar transcriptional activity, which seems to be the case for the similarities in the coding transcriptome. However, there is almost 50% differences in cis-regulatory transcription binding sites such as enhancers occupying different locations, which could be one of the reasons for the differences in overall transcriptional activity between the two genomes (Stergachis et al., 2014). Moreover, most of the evolutionary conserved retrotransposons containing DNA elements that function in transcriptional activities seems to be species specific, suggesting further regulatory differences between both genomes (Yue, et al., 2014), that could give rise to differences in the non-coding transcriptome, and hence differences in the complexity of gene regulatory networks driving different cellular and developmental functions.

Interestingly, these works have shown that less than 2% of mammalian genomes encode for proteins while the rest are non-coding, and these include small and long non-coding RNAs (lncRNAs) (Carninci et al., 2005; Djebali et al., 2012; ENCODE Project Consortium, Birney et al., 2007; Kapranov et al.,

2007a/b). Emerging evidence indicates that the complexity of higher organisms, which correlates with an increase in the size of the non-coding transcriptome, arises from the increase in the number and complexity of regulatory pathways that control gene expression, and that it is the variation in the non-coding transcription and sequences that causes inter- and intra-specific diversity (Levine and Tjian, 2003), which implicates an evolutionary conserve role for both protein-encoding and non-coding genomic regions within higher eukaryotes (Barrett et al., 2012). Therefore the unidirectional flow of genomic information from gene via mRNA to protein as associated with the central "dogma" of Molecular Biology is incomplete (Mattick, 2001) and it only represents a tiny portion of the repertoire of genomic regulations.

The non-coding transcriptome is made up of both small non-coding RNAs and lncRNAs. The lncRNAs are all ncRNAs with transcript length greater than 200 nucleotides (nt), but lack functional open reading frames (ORFs) to be encoded for proteins (Dinger et al., 2008). Instead, they have diverse RNA regulatory functions and are actively involved in biological processes, including DNA damage response, cell proliferation, stemness, differentiation and dosage compensation and whole organismal development (Sánchez and Huarte, 2013). A significant number of lncRNAs are antisense (AS) lncRNAs to sense (S) mRNAs, capable of forming sense-antisense (S/AS) complementary base-pairs during interactions. The FANTOM 3 project identified that over 70% of all the mapped mouse genomic transcriptional units are associated with AS lncRNAs, and perturbation of these AS lncRNAs alters the expression of the sense gene, which suggest functional regulations between the two transcript pairs (Katayama et al., 2005). The genomic regulations by S/AS pairs may be more dependent on the kind of pairing interactions, with currently known divergent, convergent and full-length S/AS-pairing configurations (Chen et al., 2004; Kapranov et al., 2005; Werner, 2013; Wight and Werner, 2013). AS lncRNAs have been shown to negatively control gene expression at different levels, including transcriptional gene silencing and post-transcriptional mRNA stability regulations (Pelechano and Steinmetz, 2013; Tufarelli et al., 2003; Werner, 2013). However, positive transcriptional upregulation of sense mRNAs has for instance been reported for the AS β-secretase-1 (AS BACE-1) gene, which is known to be pathophysiologically associated with Alzheimer's disease (Faghihi et al., 2008), and more recently, a novel functional class of AS lncRNAs (Carrieri et al., 2012), subsequently named SINEUPs, increase their targeted sense genes protein synthesis post-transcriptionally by S/AS interactions (Patrucco et al., 2015; Schein et al., 2016; Zucchelli et al., 2015a). Overall, the advancements in genomic and transcriptomic technologies in the past decades and collaborations among scientist of different backgrounds have contributed to the

generation of large functional genomic datasets, leading to the enhancement of the understanding of the functional organization of the mammalian genomes and transcriptomes. This knowledge could then potentially be expanded towards the exploration of the use of the non-coding transcriptome in biotechnology and molecular medicine.

1.2 Non-coding transcriptome: small ncRNAs and lncRNAs

It is now common knowledge that there is pervasive transcription in the mammalian genome (Carninci et al., 2005; Katayama et al., 2005; Birney et al., 2007; Kapranov et al., 2007, Djebali, et al., 2012); more specifically, about 75% of the human genome is transcribed (Djebali, et al., 2012), but less than 2% of the transcriptome are protein-encoding, resulting in enormous non-coding RNAs belonging to the small ncRNA and lncRNA families *(**Figure** 1*). ncRNAs have functional diversities, ranging from structural to regulatory functions. Highlighted are some examples of general functional roles mediated by some key ncRNAs in both the small ncRNA and lncRNA families.

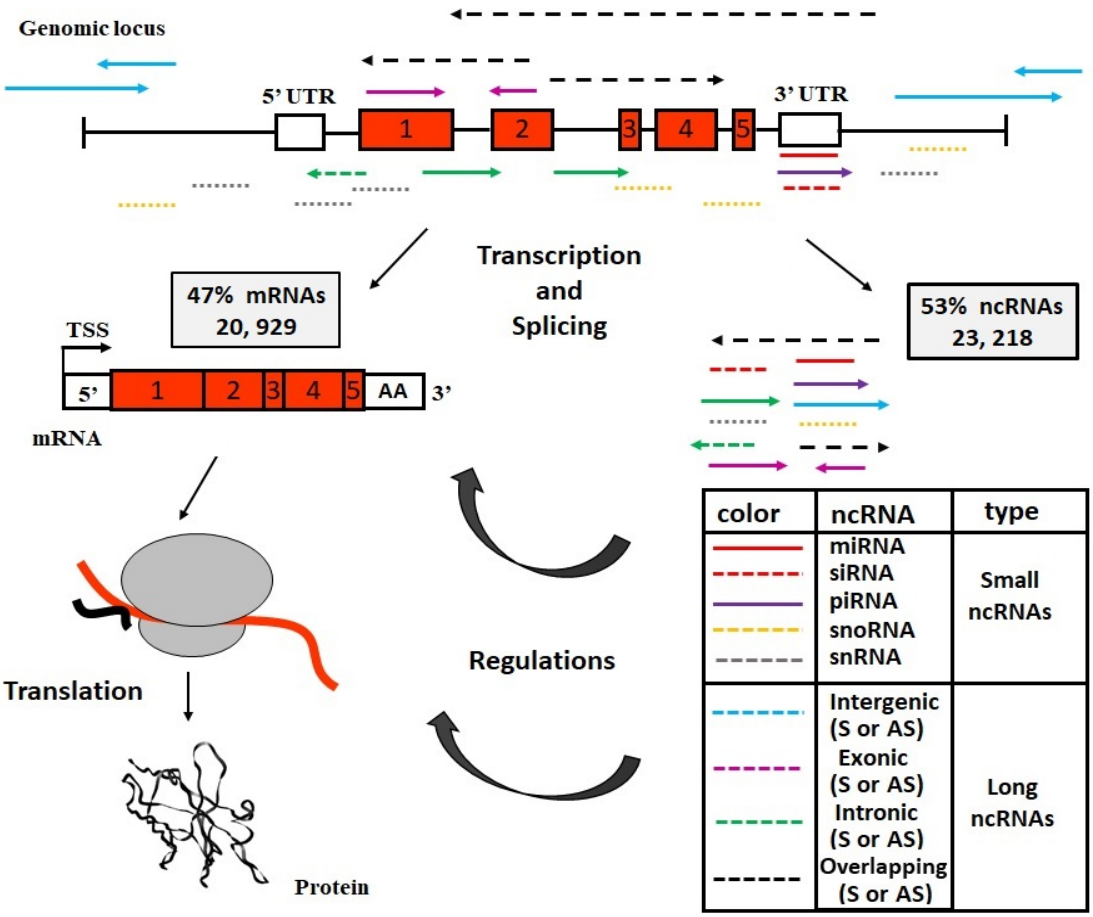


Figure 1 **Pervasive expression of non-coding RNAs in the mammalian genome**. Scheme showing a typical genomic locus with pervasive transcription at all genomic regions relative to a typical gene (upper panel). Exons are represented by red boxes separated by the introns (depicted as black lines). The light boxes are representation of 5' and 3' untranslated regions (UTRs). There is extensive transcription of both small ncRNAs and long-non coding RNAs (lncRNAs) at the intergenic, intronic, exonic, upstream and downstream of the gene locus as categories in the table (lower panel, right). Over 62.5% of the genome is transcribe (The FANTOM consortium, Carninci et al., 2005). From FANTOM 3 dataset, 47% of the transcribed RNAs in mouse form proteins while 53% are non-coding RNAs (Carninci et al., 2005; Hoon et al., 2015). As at now the LNCipedia v5.2 database has reported 127,802 annotated lncRNAs from 56,946 genes within the human genome (Volders et al., 2015, https://lncipedia.org/). Transcribed sequences may be spliced or processed to form the various ncRNAs and mRNAs. Bidirectionality of promoters may account for transcription in all direction, sense and antisense directions. Alternative splicing and RNA processing may give rise to different ncRNAs. The lncRNAs comprise of intergenic, exonic, intronic and overlapping sequences in both sense and antisense orientations (table, lower right panel); they are the majority with diverse functions in regulating cellular processes; involving proliferation, differentiation and development. Small ncRNAs such as microRNA (miRNA), small nuclear RNAs (siRNA) and piwi interacting RNA (piRNA) targets mRNAs, usually at the 3' UTR to post-transcriptionally inhibit gene expression, while the small nuclear RNA (snRNA) forms part of a complexes that function in splicing of RNAs, and the small nucleolar RNAs (snoRNA) is involve in the chemical modification of ribosomal RNAs and t-RNAs.

1.2.1 The small non-coding RNAs (ncRNAs)

The small ncRNAs are usually short non-coding RNA sequences with many classified different types; miRNAs (microRNAs), siRNAs (small interfering RNAs), piRNAs (Piwi interacting RNAs), snoRNAs (small nucleolar RNAs), snRNAs (small nuclear RNAs), tRNAs (transfer RNAs) with different functions. miRNAs and siRNAs are relatively small, about 19-28 nucleotides (nt) in length and are often involved in RNA degradation through RNA interference (RNAi) and gene silencing pathways (Bartel, 2004; Kim, 2005). The piRNAs are 29-30nt in length that specifically interacts with piwi-proteins involved in germ cells maintenance regulations (Barrett et al., 2012; Girard et al., 2006) through post-transcriptional gene silencing of retrotransposons (Siomi et al., 2011). The snRNAs are about 150nt in length, and form the major component of the spliceosome involved in splicing regulation (Wassarman and Steitz, 1992b), which evidently play roles in the global processing and regulation of eukaryotic RNAs (Barret et al., 2012). The snoRNAs are about 60-250nt in length and are classified into three types according to the conserved sequence motifs ("boxes") (Scott and Ono, 2011), and are often involved in the chemical modifications such as pseudouridination and methylation of ribosomal RNAs, tRNAs and snRNAs (Dragon et al., 2006). The tRNAs are about 75-95nt in length (Goodenbour and Pan, 2006), and are involved in decoding the mRNA, by transferring codon-anticodon matching amino acids to the mRNA bound ribosomal complex for protein synthesis.

On the contrary to direct regulation of cellular processes, other ncRNAs are structural components contributing to the architecture of multicomplex biological components. The ribosome is one typical example of a complex biological machine that comprise of both small ncRNAs and lncRNAs that act in coordination together with ribosomal proteins to mediate protein synthesis. The eukaryotic ribosomal RNAs (rRNA) sequences and structures seem to be evolutionary conserved among various organisms. The human rRNA is approximately 7216nt (NCBI Biosystem Database). It comprises of the eukaryotic large subunit (LSU) rRNAs consisting of 5S: 121 nt, 5.8S: 156nt and 28S: 5070nt while the small subunit (SSU) rRNAs is composed of 18S: 1869nts; altogether forms approximately 60% of the ribosomal material by weight and participate in recruiting the translation machinery-mRNA complexes, through interactions with mRNAs and RNA binding proteins (RBPs) during protein synthesis. Although small ncRNAs are relatively short in sequence as compared to lncRNAs, they all seem to have evolved with unique modes of regulations through recruitment of effectors to selectively bind nucleic acids to control gene expression on a global scale.

1.2.2 The long non-coding RNAs (lncRNAs)

According to LNCipedia v5.2 database, there are at this reported moment 127,802 annotated lncRNAs from 56,946 genes within the human genome (Volders et al., 2015, https://lncipedia.org/). However, other annotations of human lncRNAs in GENCODE, FANTOM and NONCODE among others fall short of this high number of annotation in the LNCipedia v5.2 database. For instance the GENCODE v27 has just 15778, FANTOM CAT v1 has 27919, NONCODE v5 has 96308, among other annotated browsers as reported (Uszczynska-Ratajczak et al., 2018). This rise in numbers is partly due to the development of cap analysis of gene expression (CAGE), genomic microarray, chromatin immunoprecipitation combined DNA sequencing (ChIP-seq), RNA-seq and detection of histone modification technologies as used extensively by the FANTOM and ENCODE consortia (Uszczynska-Ratajczak et al., 2018). A large portion of lncRNAs are expressed, which means they form part of the active transcriptome as annotated, but most of them have no known functions. Therefore, there is yet more work to be done for their functional identification, validation and characterization. Analysis from the GENCODE consortium, among others, have so far indicated that lncRNAs are transcribed and processed similarly to protein-encoding transcripts. However, they are expressed at lower levels than protein-coding genes, and their expression is also more tissue-specific and are localized within the nucleus as compared to the coding counterparts (Derrien et al., 2012). This expression properties seem to suggest that although, lncRNAs exhibit diversity in primary sequences, they seemed to have evolved to regulate key, but similar, biological activities. The lncRNAs have been shown to function at almost all cellular processes with core effect on controlling gene expression by forming parts of complex structures like the nuclear structure, spliceosome and ribosome. They can remodel chromatin states influencing replication and transcription, regulation of splicing, cellular trafficking, mRNA stability and translational processes (Garitano-Trojaola et al., 2013; Marín-Béjar and Huarte, 2015).

1.3 Classification of lncRNAs

LncRNAs are highly diverse, making them possess different primary sequence anatomies, secondary structures, sizes, orientations, subcellular localization and function. This makes it very difficult to classify them. However, they could be categorized according to the aforementioned generic properties such as their genomic location and organization, regulatory properties and modes of action that could be shared

among a subset of them. As such, some relevant classifications would be highlighted below according to reported data.

1.3.1 Classification of lncRNA by genomic organization

The GENCODE v7 catalog has classified human lncRNAs into two main categories according to the genomic location as intergenic and generic lncRNAs with considered sequence length >200bp and no protein encoding properties (Derrien et al., 2012). The intergenic lncRNAs are those obtained from the so called "junk DNA"; they are lncRNAs with detectable chromatin signatures but do not intersect with any protein-coding genes. They are the majority (Derrien et al., 2012; Khalil et al., 2009). The remaining genes have the transcription intersected with or are transcribed from protein-coding genes and they are classified as generic lncRNAs ***(Figure 2A)***, including exonic, intronic and overlapping lncRNAs that can be transcribed in sense (S) and antisense (AS) orientations to the protein-coding genes (Derrien et al., 2012). The exonic types cover protein-coding genes but are non-coding, while the intronic ones are spliced introns and the overlapping ones are lncRNAs consisting of both exonic and intronic sequences that are not spliced. The generic classification subtypes that are exonic, intronic and overlapping consist of both sense and antisense sequences derived from each subtype classified according to the orientation with the protein-coding gene ***(Figure 2A)*** (Derrien et al., 2012).

However, the classifications based on genomic localization relative to the protein-coding genes could only be useful in identifying new lncRNAs but not their functional characterization. Therefore, there should be the need to follow up with functional characterizations.

1.3.2 Classification of lncRNAs by function in *cis* or *trans*

Aside the identification of lncRNAs according to their genomic localization, there is a classification based on the ability of lncRNAs to work locally in *cis* to regulate its original or a nearby genomic locus, where it was originally transcribed, or to act in a distal location in *trans* ***(Figure 2B, right).*** The *cis*-acting lncRNAs are considered to regulate chromatin remodeling and transcriptional events (Yan et al., 2017). Thus, they are expressed and processed as functioning lncRNAs and then act to regulate neighbor gene expression as enhancer RNA for promoters or through direct binding and recruiting chromatin-associated proteins that could remodel chromatin structure within the vicinity regions (Chu et al., 2011).

This kind of local regulation could occur by at least 3 potential mechanisms (Kopp and Mendell, 2018): (1) regulation of nearby genes by recruiting regulatory factors to the locus; (2) its expression causes chromatin remodeling that confers gene regulation of nearby genes, independent of the sequence of the RNA gene transcript and (3) the presence of unique regulatory elements within the promoter or in the gene body of the lncRNA. This is a typical effect mediated by XIST on X chromosome inactivation, where transcription of the XIST gene from one X chromosome is processed, binds and deposit repressive markers on the chromatin, causing transcriptional silencing of almost all gene expression on such inactive X chromosome (Kopp and Mendell, 2018; Wutz, 2011). Other instance of transcriptional dependent regulation is the ability of lncRNAs to act in *cis* as enhancer or repressor RNA to influence the transcription of other nearby genes. For instance, the lncRNA HOXA upstream noncoding transcript, Haunt 1 DNA locus, contains a potential HOXA enhancers element to activate Haunt 1 transcription. The transcribed Haunt 1 then paradoxically forms a negative feedback loop to inhibit high-level transcriptional activation of the HOXA gene expression, leading to promotion of embryonic stem cell (ESC) differentiation (Yan et al., 2017).

There is also post-transcriptional cis-regulation by AS lncRNAs at the levels of increasing mRNA stability and in translation to upregulate targeted sense genes protein synthesis. One typical post-transcriptional cis-regulation could be seen by the AS β-secretase-1 (BACE1) gene, which is transcribed from chromosome 11, on the opposite strand of BACE1 gene locus, process, pair and then enhance BACE1 mRNA stability post-transcriptionally and this upregulation is known to be pathophysiologically associated with Alzheimer's disease (Faghihi et al., 2008). Other post-transcriptional cis-regulation could be observed with AS lncRNA families called SINEUPs (***Figure 2B, left***, Carrieri et al., 2012; Patrucco et al., 2015; Zucchelli et al. 2015a/b; Schein et al., 2016), which activate translation of their sense mRNAs by acting as RNA chaperones to activate translation of the sense mRNA under normal and stressful conditions, including mTORC1 inhibition by rapamycin (Carrieri et al., 2012).

Whereas lncRNA *cis*-regulators exhibit local or nearby gene regulations, the lncRNA trans-regulators act on a genome or transcriptome wide scale as RNA scaffolds, decoys and guides, through specific interactions involving RNA:DNA:DNA (Garitano-Trojaola et al., 2013) or RNA:RNA and protein complexes formations. These interactions are typical in epigenetic regulation of distantly located genes by cooperating with chromatin modifiers, resulting in transcriptional gene activation or repression (Marín-Béjar and Huarte, 2015). A typical example is mediated by HOX transcript antisense RNA (HOTAIR), which is expressed from a HOXC gene cluster on chromosome 12, processed to bind and

target chromatin-modifying complexes consisting of polycomb repressor complex 2 (PRC2), and LSD1/CoREST/REST complexes to repress HOXD gene locus, located on a different chromosome 7 (Gupta et al., 2010; Tsai et al., 2010). HOTAIR is for instance known to induce genome-wide re-targeting of PRC2 occupancy, leading to H3 lysine 27 (H3K27) trimethylation (Tsai et al., 2010) and inducing cancer progression (Gupta et al., 2010).

Overall, lncRNAs exhibit functional regulatory domains, capable to recruit partners to be able to exhibit varieties of functions by acting in *cis* or *trans* or both at a time to mediate a cellular task. This is probably promoted by the organization of lncRNAs into structural domains that act as signal. decoy, guide and scaffold archetypes to act in *cis* or *trans* in regulating cellular processes.

1.3.3 Classification lncRNAs by general molecular mechanisms

Since most lncRNAs harbor regulatory elements, they could probably form reliable secondary structures to be able to exhibit function in *cis* and *trans* depending on the recruited partners. This is why there seem to be similarities in the modes of action by lncRNAs, despite differences in the primary sequences. In fact, different reports suggest that secondary and tertiary structures are instead conserved among different types of lncRNAs, suggesting structural-functional relationship among lncRNAs (Li et al., 2016; Novikova et al., 2013a, 2013b). Therefore, most lncRNAs exhibited similar molecular mechanism of action, which would be based on specialized structural elements they could form when expressed. Thus, lncRNAs with similar structural domains may exhibit similar mode of action. This has probably led to the suggestion to classify lncRNAs according to their molecular mode of function as signal, decoy, guide and scaffold archetypes (Wang and Chang, 2011a), which will briefly be elaborated with known examples.

1.3.3.1 Signal lncRNA archetypes

The signal archetype ***(Figure 2C, i)*** suggests that some lncRNAs expression and recruitment are as a result of diverse external and internal stimuli influencing their transcription and localization. Thus, some lncRNAs are either expressed in response to internal or external signaling agent or are part of the signal relay leading to a biological process. For instance, there are different patterns of lncRNAs expression in response to DNA damage repair, where their expression is activated and recruited to either activate

transcription of key genes or recruit other ncRNAs or proteins necessary for the damage repair. As an example, a damage-induced lncRNAs (dilncRNAs) expression are activated by RNA pol II upon DNA damage, and as a response recruit small ncRNAs by RNA-RNA pairing, followed by p53 binding protein 1 (53BP1) recruitment to mediate DNA repair (Michelini et al., 2017).

1.3.3.2 Decoy lncRNA archetypes

The decoy lncRNAs archetype (***Figure 2C, ii)*** regulates gene expression through competitive inhibition of proteins function. Thus decoys bind and sequester regulatory proteins involve in several processes including chromatin remodeling, transcription and splicing or they could even act as pseudogenes or competitive endogenous RNAs (ceRNAs) (Salmena et al., 2011) to competitively protect mRNAs from small RNA interference, enhancing mRNA stability. As an example, the non-coding repressor of NFAT (NRON) lncRNA expression is capable of binding and sequestering NFAT protein, regulating it nuclear trafficking and inhibiting it function in cell lines (Willingham et al., 2005).

1.3.3.3 Scaffold lncRNA archetypes

The scaffold lncRNA archetypes ***(Figure 2, iv)*** are capable of regulating several cellular processes by possessing different RNA binding protein domains (RBP) to recruit different effectors; both RNAs and proteins in achieving a biological effect. As mentioned above, DNA double strand breaks act as a signal to activate RNA pol II expression of dilncRNAs, which then act as a RNA scaffold to recruit small non-coding RNAs, and altogether bind 53BP1 to the site for DNA damage repair (Michelini et al., 2017). In other case scenario is the role play by HOTAIR ***(Figure 2B, right)*** to recruit both PRC2 and LSD1/CoREST/REST complexes to repress HOXD gene locus, through chromatin remodeling effects.

1.3.3.4 Guide lncRNA architypes

Finally, the guide lncRNAs archetype (**Figure 2**C, iii) can bind to proteins like RBPs and other RNAs and then potentially guide them to specific sites like a gene locus to activate gene expression. Here again, HOTAIR ***(Figure 2B, right)*** can function as a guide lncRNA to recruit chromatin modifying complexes PRC2 and LSD1 to HOX D gene locus, leading to gene repression.

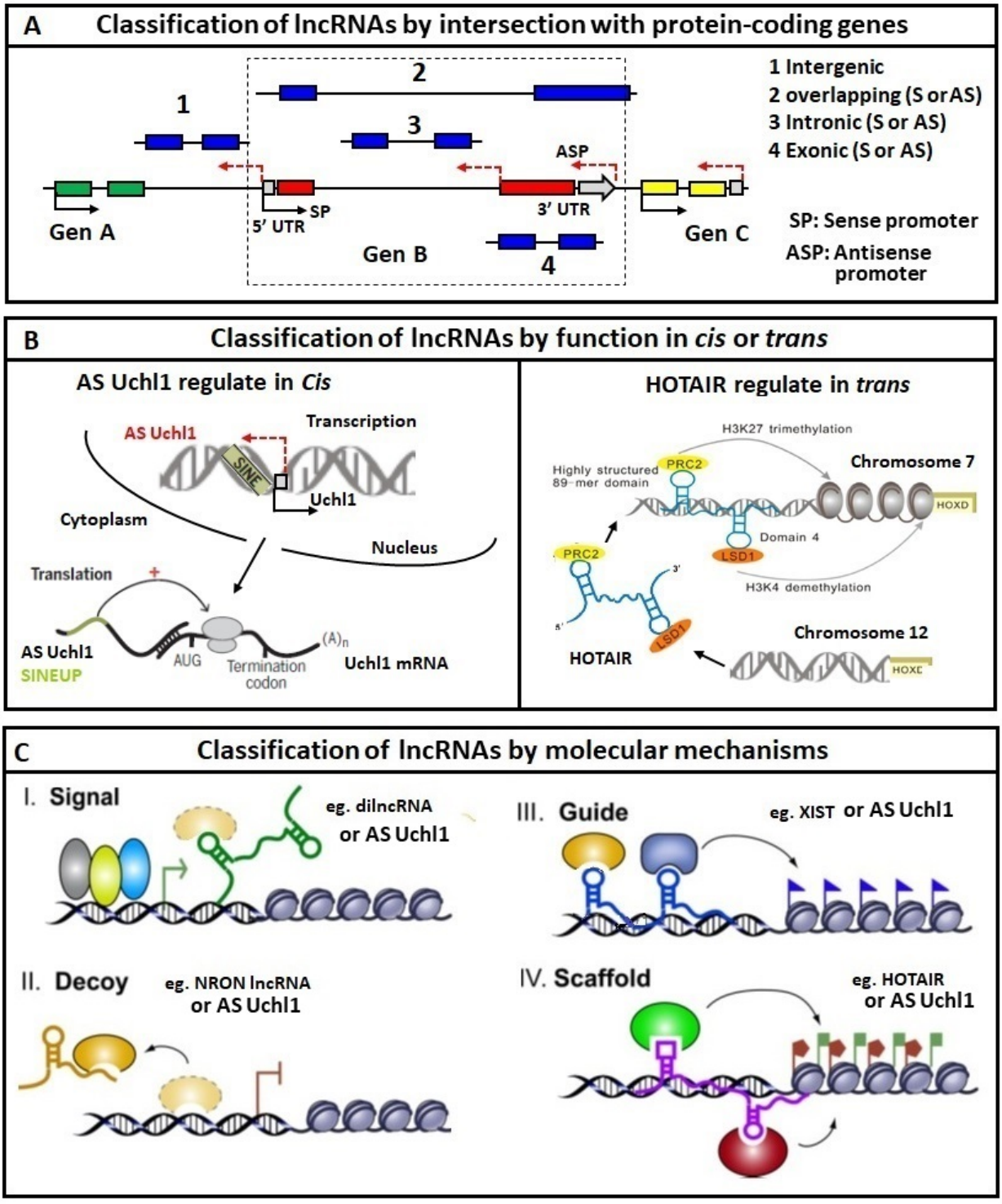


Figure 2. Genomic organization and functional classification of lncRNAs. (A) classification of lncRNAs by intersection with protein coding genes. Scheme of genomic loci comprising of gene A, B and C with coding exons in green, red and yellow, respectively, and various class of lncRNAs: intergenic (1), overlapping (2, either in sense-S or antisense-AS), intronic (3, S or AS) and exonic (4, S or AS) with exons in blue colored, and all are positioned relative to the nearest genes. Direction of transcription are indicated by the sense promoters (SP, in bold black arrows) and antisense promoters (ASP, in dotted red arrows). (B) Functional classification of lncRNAs according to their function in

cis (left) or *trans* (right). Depicted is a scheme showing transcription and translation activation function in *cis* mediated by SINEUP of AS Uchl1 that targets and increase Uchl1 protein synthesis in the cytosol (B, left), whereas depicted on the left is scheme of HOTAIR (modified from Li et al., 2016), which is transcribed from HOX C locus on chromosome 12, processed and then recruits PRC2 and LSD1/CoREST/REST complexes to function in *trans* to epigenetically modify chromatin structure in the HOX D locus on chromosome 7. (C) Classification of lncRNAs by molecular archetypes (modified from Wang and Chang, 2011) with typical examples (eg.) in each architype; signaling mediators (eg. Damage-induced (di)lncRNAs mediates DNA damage repair by recruiting microRNAs and p53 binding protein 1, 53BP1), protein or nucleic acid complexes decoys (eg. NROF, non-coding repressor of NFAT 1 transcription factor binds to NROF when expressed, reducing NFAT 1 mediated transcriptional activation, protein and RNA guides (eg. XIST guides PRC2 complex and YY1 protein to cause epigenetic gene silencing) and scaffolds (eg. HOTAIR is scaffold for PRC2 and LSD1 protein complexes for epigenetic programming of chromatin structure). However, AS Uchl1 seem to exhibit characteristics of all 4 archetype; where it is expressed and activated to mediate mTORC1 inhibition by rapamycin induced stress signaling, shuttle to the cytoplasm where it can act as lncRNA decoy to recruit away translational complexes and as a scaffold to present and guide such complexes to the translation of Uchl1 protein synthesis (B, left).

However, it seems generally that these molecular modes of lncRNAs functions are not mutually exclusive, making some lncRNAs like HOTAIR ***(Figure 2B, right)*** capable of exhibiting more than one molecular archetype upon expression in time and space. This seems possible because, most lncRNAs sequence are organized into functional domain units (**Figure 2**B, right), and each domain is capable of functioning upon activation by cellular cues.

Overall, the genomic architecture of lncRNAs ***(Figure 2A)*** is important in influencing their expression and functionality. Thus, lncRNAs have canonical gene structures (Carninci et al., 2005; Derrien et al., 2012), influencing their expression, processing and mobilization. Although, lncRNAs are diverse in sequence, they exhibit functional domain units and structures, enabling recruitment of effectors; proteins and other RNAs to mediate specific functions by acting in *cis* ***(Figure 2B, left)*** or *trans* (***Figure 2B, right)*** or both at a time to promote a specific cellular process such as proliferation, differentiation and development. lncRNAs molecular mechanisms are probably promoted by their primary sequences, functional domain units or entire secondary and tertiary structures acting as signal transductors or agents, decoys, guides and scaffold archetypes ***(Figure 2C)*** in regulating key cellular processes. Therefore, understanding lncRNAs genomic organization, conserved primary sequences and motifs, and structural domains may aid in their functional and molecular characterization in health and diseases.

1.4 Repetitive elements in the mammalian genome

In humans, canonical genes are estimated to cover about 30% of the genome, but less than 2% codes for proteins (Matylla-Kulinska et al., 2014), while repetitive elements are estimated to account for over two-thirds of the genome (de Koning et al., 2011). The repetitive elements in a genome can be widely interspersed or be located next to each other, forming interspersed and tandem repeats, respectively. Some regions are highly composed of repeat elements, while other areas are depleted, and the distribution of repeats like transposable elements (TEs) are lower in gene-rich regions than intergenic and gene-poor areas (Walter et al., 2016). The tandem repeats which comprise of satellite and variable number of tandem repeats (VNTRs) cover about 22-25% of the genome, while the interspersed repeats which consist of retrotransposons and DNA transposons altogether cover about 42-45% of the genome ***(Table 1 and Table 2)***, Casa and Gabellini, 2012). The satellite and VNTR DNAs ***(Table 2)*** span up to hundreds of kilobases within the constitutive heterochromatin, which differ from the rest of repeats by nucleotide content, allowing for their easy separation by density gradient centrifugation as satellite bands (Matylla-Kulinska et al., 2014)

The mammalian retrotransposons are composed of 8% of long terminal repeats (LTR) and about 37% of non-LTR families. The non-LTR family consist of long interspersed nuclear elements (LINEs), which are the majority and are about 20% of genomic repeats and short intersperse nuclear elements (SINEs) and SVA, which are usually presented in high copies but are the minority, comprising of about 13% in the genome. The SVA are transposons formed by the fusion of a SINE-R and an Alu, and are separated by VNTRs. Thus an SVA consist of SINE-R, VNTR and Alu TEs in their respective order and flanked by target-site duplication (TSD) sequences (Ostertag et al., 2003). While LINEs are autonomous TEs, SINEs and SVA are non-autonomous TEs and require the activity of LINEs for their mobilization in *trans*. Among this RNA transposons, SVAs have low nucleotide divergence, suggesting that they have evolved within the genome in recent times, making them possess a small number of mutations (Ostertag et al, 2003). The DNA transposons such as MER 1 and 2, Mariner and Merlin, accounted for only 2-3% of the intersperse repeats in the human genome (***Table 1,*** Casa and Gabellini, 2012). Although most eukaryotic genomes are full of repeats, some DNA transposons are also found in prokaryotes as simple insertion sequences or as part of complex sequences; their subclasses are distinct based on the number of DNA strands that are cut during transposition, which usually occurs during chromosomal replication (Wicker et al., 2007).

Repeats can function in maintenance of genomic replication and integrity, chromosomal structural organization, and transcriptional regulation (***Table 1,*** Casa and Gabellini, 2012). Genomic repeat elements role in evolution has long been known (Makałowski, 2000), since these elements caused genetic mosaicism leading to phenotypic variability among species. They are known to be the main source of epigenetic variability within genomes (Ekram et al., 2012) and some TEs like LTRs contribute to genomic gigantism in some eukaryotes (Sun et al., 2012). Their retrotransposition could lead to interactions with surrounding sequences and nearby genes, and overall, influencing genes expression. They seem to regulate processes at the genomic level through chromatin remodeling and epigenetics regulation of transcription, and also at the post-transcriptional levels regulating splicing, mRNA stability and translation (Carrieri et al., 2012; Ekram et al., 2012; Gong and Maquat, 2011; Kramerov and Vassetzky, 2011a; Ponicsan et al., 2010). Therefore, TEs effects are under selective pressure and often repressed through epigenetic controls such as DNA methylation and histone modification, and other silencing pathways like RNA interference (Casa and Gabellini, 2012; Ekram et al., 2012). These TEs are mobile and their insertion in the genome is one source of genomic evolution, contributing to genomic size and organismal complexity through regulation of chromatin structure and gene expression at all levels; transcription, splicing, mRNA stability and translation (Carrieri et al., 2012; Ekram et al., 2012; Gong and Maquat, 2011; Kramerov and Vassetzky, 2011a; Ponicsan et al., 2010)

Table 1 Features of represented intersperse DNA repeat types in the human genome (source: Gabellini and Casa, 2012).

Repeat type				Estimated number of copies	Average length	Mobility	Estimated % genome coverage	
Interspersed	Retrotransposons	LTR	LTR (Long terminal repeat) or ERV (Endogenous retroviruses) (MaLR, ERV, ERV1, ERV-K, ERV-L, etc.)	200,000	6–11 kb	Autonomous retrotransposition (retroviral-like)	8%	42%
		Non-LTR	LINE (Long interspersed element) (L1, L2, CR1, etc.)	500,000	6 kb	Autonomous retrotransposition	20%	
			SINE (Short interspersed element) (Alu, MIR, etc.)	1,000,000	0.3 kb	L-1 dependent Retrotransposition	13%	
			SVA SINE-R/VNTR/Alu	2700	2–5 kb	L-1 dependent Retrotransposition		
	DNA transposons	DNA transposons (MER1, MER2, Mariner, Merlin, etc.)		300,000	1–3 kb	inert	2–3%	

Table 2 Features of represented tandem DNA repeat types in the human genome (source: Gabellini and Casa, 2012)

Repeat type			Unit length	Array length	Estimated % genome coverage
Tandem	Satellite	Alpha-Satellite	171 bp	3–5 Mb	22–25%
		Satellite II (HsatII)	23–26 bp or multiple	10–70 kb	
		Satellite III (GAATGn- simple sequence)	5 bp or multiples up to 70 bp	7.5–100 kb	
		Beta-Satellite	68 bp	2–14.5 kb	
		Gamma-Satellite	220 bp	10–200 kb	
	VNTR (Variable number of tandem repeats)	Microsatellite (Short tandem repeat)	1–13 bp	Hundreds bp	
		Minisatellite (including telomeric repeats)	6–100 bp	1–15 kb or more	
		Macrosatellite	2–12 kb or more	Tens up to hundreds bp	

1.4.1 Mammalian interspersed transposable element (TE) families: Class I and II TEs

TEs were originally classified into two main classes according to their mechanism of retrotransposition and mobilization intermediates: RNA or retrotransposons, class I and DNA transposons, class II (Wicker et al., 2007). The class II TE encodes a transposase to mediate "cut and paste" mode of DNA repeat sequences mobilization into a genomic locus, while the class I TEs mobilizes a repeat sequence via retro-transcription of a repetitive RNA sequence and insertion back into a new genomic site via "copy and paste" mode of mobilization. In the class I families, no member cleave and transfer DNA strands at the donor site, instead an RNA intermediate is transcribed and then reversed transcribed into a DNA before insertion, and each complete replication cycle produces a single new copy (Wicker et al., 2007). The class II elements, DNA transposons, are also found in prokaryotes, and they are not mobilized through RNA intermediates (Wicker et al., 2007). The retrotransposons consist of two subclasses, LTR and non-LTR that represent up to 42% endogenous TEs human genome (***Table 1,*** Casa and Gabellini, 2012).

1.4.1.1 Long terminal repeats (LTRs)

The endogenous LTR elements possess identical direct LTR sequences flanking the coding sequences (CDS) at the 5' and 3' region of the genes. They are found in reverse transcriptase (RT), gag and envelope protein genes of some LTR elements like retroviruses (Kramerov and Vassetzky, 2011). They can range

in size from few bases, about 100bp up to about 11 kbp, as represented ***(Table 1)***. There are two main subtypes, the *copia*-like and *gypsy*-like LTR-transposons and are present in all eukaryotic genomes (Kramerov and Vassetzky, 2011). Endogenous LTR are like retroviruses in terms of genomic structures and mode of retrotransposition, but do not encode functioning envelope genes to aid in extracellular entry into host cells similarly to canonical retroviruses. Therefore, inactivation or deletion of functioning envelope genes that allow extracellular mobility in retroviruses can transform them into endogenous LTR in germ line genomes, which are henceforth only capable of undergoing mobilization and transmitted vertically in germ lines for propagation (Wicker et al., 2007). LTRs TEs in general cannot be replicated autonomously or be mobilized in trans by other functioning TEs and they lack functioning ORFs, and their sequences can be found alone without internal coding sequences (Walter, 2015). Endogenous LTR constitute about 8% of all repeats in humans (***Table*** 1; Casa and Gabellini, 2012) and 10% in mouse genome, but are the majority of all repeats in most plant genome, making up to 75% and 90% of all repeats in the maize and wheat genome, respectively (McCarthy and McDonald, 2004).

1.4.1.2 Non-long terminal repeats (Non-LTRs)

The non-LTR are the majority of all repeats in the mammalian genome, making up to 38% of all repeats in humans (***Table 1,*** Casa and Gabellini, 2012) and widespread in most eukaryotic genomes. They consist of two main subtypes, LINEs and SINEs. All forms of retrotransposition are mediated by LINEs, since they possess two open reading frames (ORFs), protein-encoding for functioning reverse transcriptase and endonuclease. Instead, SINEs are nonautonomous retroelement and depends on the LINE machinery to retrotranspose in the genome. Therefore, LINEs are responsible for mobilization and retrotransposition of non-autonomous SINEs, SVAs and some processed pseudogene substrates in the mammalian genome (Casa and Gabellini, 2012).

1.4.1.3 Long intersperse nuclear elements (LINEs)

The LINE TEs are up to several kilobases in length, lack LTRs and are found in all eukaryotes (Wicker et al., 2007). There are several subgroups; LINE (L) 1, L2 and L3 comprising 20% of the human genome (***Table 1***, Casa and Gabellini, 2012), The L1 is the predominant LINE representing 17% of the human DNA. There are about 80-100 L1 copies in humans, but over 99.9% of all LINEs are “dead” or inactive (Doucet et al., 2010), due to accumulated mutations such as 5’ truncations and other mutations (Denli et al., 2015) that arise during mobilization. All LINEs are autonomous retrotransposons and encode for at

least one protein, mostly ORF2 for mediating retrotransposition. Human protein-encoding L1 is approximately 6kb in length. It begins at the 5' UTR, a bidirectional sense-antisense RNA pol II promoter containing 910bp sequence, followed by a non-overlapping bicistronic ORF 1 and 2 sequences and a 3'UTR poly-A rich sequence (Doucet et al., 2010; Denli et al., 2015). Due to the bidirectional pol II promoter activity, there is an active antisense ORF, called ORF0 in the 5'UTR that encodes for an active protein that is known to influence L1 mobilization in a genome (Denli et al., 2015). The ORF1 encodes for an RNA-binding protein (RBP) that is crucial for L1-mediated retrotransposition with essential RNA chaperone activity (Li et al., 2006), and the ORF2 encodes for a large protein with reverse transcriptase (RT) and endonuclease activities, and a C-terminus that altogether allows for L1 autonomous mobilization within the genome (Denli et al., 2015). The mouse L1, which is about -6.5kb is similar to that of human L1, however, it has IRES elements upstream of both ORF1 and 2 sequences, which mediates their translation into ORF 1 and 2 proteins (Li et al., 2006). Instead, in humans, the LI ORF2 is considered to be translated by termination/reinitiation mechanism, and both ORF1 and the inter-ORF spacer sequences are dispensable for ORF2 translation (Alisch et al., 2006).

Expressed ORF1 and 2 proteins form ribonucleoprotein particle (RNP) complexes in *cis* with a transcribed L1 mRNA, and are required for all L1 retrotransposition events, including insertion of the reverse transcribed DNA into a new genomic locus (Doucet et al., 2010). However, new LI DNA insertions usually create severe 5' truncated ends and short tandem site duplications (TSD), which are usually non-functional and incapable of self -mobilization. L1 activity can be diverted and caused retrotransposition in *trans* of various non-specific processed genic mRNAs, including other TEs like SINEs and SVAs, creating most of the time "dead-on-arrival" pseudogenes lacking introns and promoters, (Esnault et al., 2000). Overall, the retrotransposition events of LI definitely influence transcriptome diversity and contributes to gene regulations (Denli et al., 2015) at different levels; from chromatin remodeling and transcription to translation. The AS transcription mediated by the ORF0 can positively influence it sense L1 and nearby genes expression through recruitment of transcription machinery to cause chromatin structure remodeling or via formation of ncRNAs (Denli et al., 2015).

1.4.1.4 Short intersperse nuclear elements (SINEs)

SINEs are non-autonomous TEs, therefore, they rely on autonomous functional LINEs for their retrotransposition events in *trans*. They have distinct evolutionary origin, thus unlike LINEs, they

originate from accidental retrotransposition of RNA pol III transcripts such as tRNA, signal recognition particle 7SL RNA and 5S rRNA, making them possess an internal pol III promoter at the 5' end that influence their expression (Wicker et al., 2007; Kramerov and Vassetzky, 2011). They are generally known to be very small in size, about 80-700bp (Kramerov and Vassetzky, 2011; Vassetzky and Kramerov et al., 2012; Walter, 2015), but widely distributed in higher copies, about 10^4- 10^6 copies or at least 100 copies in the mammalian genome (Kramerov and Vassetzky, 2005; Vassetzky and Kramerov et al., 2011). They represent about 13% of all known repeats in the human genome (***Table 1***; Casa and Gabellini, 2012). Any genome can contain several TE SINE families, but are usually non-protein encoding (Kramerov and Vassetzky, 2011). They are ubiquitous and are spread at all regions including intergenic, gene-rich and -poor regions (Ponicsan et al., 2010). Their expression, especially mouse B1 and B2 SINEs are increased in response to varieties of cellular stresses (Li et al., 1999) and DNA-damaging agents (Rudin and Thompson, 2001).

SINEs, like other TEs, can act as genomic parasites, creating insertion mutagens at the genomic level, influencing negatively nearby genes transcription. However, SINEs presence in the genome can also act as source of pol III promoters, enhancers, silencers and insulators influencing nearby gene expression positively as well as a source of ncRNAs sequences (Kramerov and Vassetzky, 2011). Transcribed SINE RNAs have long been suggested to act as regulators of gene expression by modifying genomic architecture and mediating alternative splicing (Ponicsan et al., 2010), mRNA stability (Gong and Maquat, 2011) and translation (Carrieri et al., 2012; Zucchelli et al., 2015a). They have also been implicated to be sources of miRNAs that play roles in complex-gene regulatory networks through RNAi, inhibiting genes expression (Gong and Maquat, 2011).

1.5 SINE evolution: origin, structure and classification

SINEs sequences are within 80-700bp long (***Figure 3***B). However, functional SINEs are typically about 150-200bp in length (Kramerov and Vassetzky, 2005 and 2011). The SINE sequence structure is composed of at least two unidentical parts: the 5' head terminal, main body and the 3' tail terminal regions, and the head is usually used to deduce the typical pol III-mediated RNA source, including tRNA, 7SL RNA and 5S rRNA (***Figure 3B***, Kramerov and Vassetzky, 2005 and 2011), which by itself maintains pol III promoter activities in the corresponding SINE sequence. The body defines the subfamily types and is usually composed of LINE TEs from which they descended from, and the 3'tail is composed of

short mono-, di-, tri-, and penta-nucleotide repeat sequences, including AT-rich or poly-A tail short sequences (***Figure 3B***; Kramerov and Vassetzky, 2011). While the head is usually present and noticeable, some SINEs body or tail may be short or absent (Kramerov and Vassetzky, 2005). The body of most SINEs has 30-100bp region similar to the 3'-terminal of a LINE TE type and, as such, the typical LINE RT may be involved in the retrotransposition events of the corresponding SINE TE (Matveev and Okada, 2009). However, SINEs unarguably originate from pseudogenes of cellular pol III RNAs with the tRNA source being the most abundant (Kramerov and Vassetzky, 2005 and 2011). The 7S RNA-derived SINE family are so far identified only in rodents, while the 5S rRNA SINE family ancestors are found in some fishes and few mammals like the bats (Gogolevsky et al., 2009; Nishihara et al., 2006).

Many SINE families possess atypical SINE structures, thus while some SINEs are just monomeric, containing only one pseudogene source, many of them are chimeric and more complex, consisting of dimeric or multimeric SINEs that arise from either the same or different pseudogene RNAs (Gogolevsky et al., 2009; Kramerov and Vassetzky, 201, Nishihara et al., 2006). For instance, the primate Alu consist of two similarly derived 7SL RNA parts, while most rodent SINEs are derived from the fusion of 7SL RNA- and rRNA-derived TEs (Churakov et al., 2010; Kramerov and Vassetzky, 2011a) or rRNA/tRNA SINEs (Gogolevsky et al., 2009). Dimeric SINEs usually outnumbered their monomeric ancestors, and even most monomeric SINEs like B1 SINEs ***(Figure 3B)*** are not unique, they have about -30nt internal duplications in their body parts (Kramerov and Vassetzky, 2011a, 2011b). Furthermore, SINEs-like rodent ID ***(Figure 3B)*** have reduced tRNA-derived body and poly-A rich tail structures which are unusual in most SINE elements (Kramerov and Vassetzky, 2005). As a general rule for SINEs classification, SINEs consisting of exclusively head and tail sequences are called simple SINEs, whereas dimeric or all multimeric SINEs are called complex SINEs ***(Figure 3A)***; the SINE sequences must be identical by at least 60% to either tRNA and 5S rRNA or to 7SL RNA species by at least 60nt overlap, and must also possess pol III promoter A and B boxes (Vassetzky and Kramerov, 2013).

New SINEs arise from *de novo* processed pseudogene RNAs which are produced by pol III-mediated transcription and by the activity of LINEs RT, giving rise to several SINE inserted copies in the eukaryotic genome, consisting of different monomeric and multimeric or short and unusual SINE structures. Apart from the pol III mediated transcription of tRNA, 7S RNA and 5S rRNA genes for generating SINEs, SINEs may also arise from other unknown pol III mediated RNAs, including processed cellular pseudogenes (Kramerov and Vassetzky, 2011a, 2011b).

Overall, SINE evolution includes their emergence, progressive optimization and counteraction to the cellular arsenal against mobile genetic elements, making them undergo substantial change like duplications or chimera formation with other pseudogenes to produce new SINE families with divergent sequences. On the other hand, small changes such as point mutations and deletion/substitutions can give rise to subfamilies that can coexist or replace each other; (Kramerov and Vassetzky, 2011a, 2011b). The absence of universal SINE structures means different pathways were deployed in their structural and functional evolution.

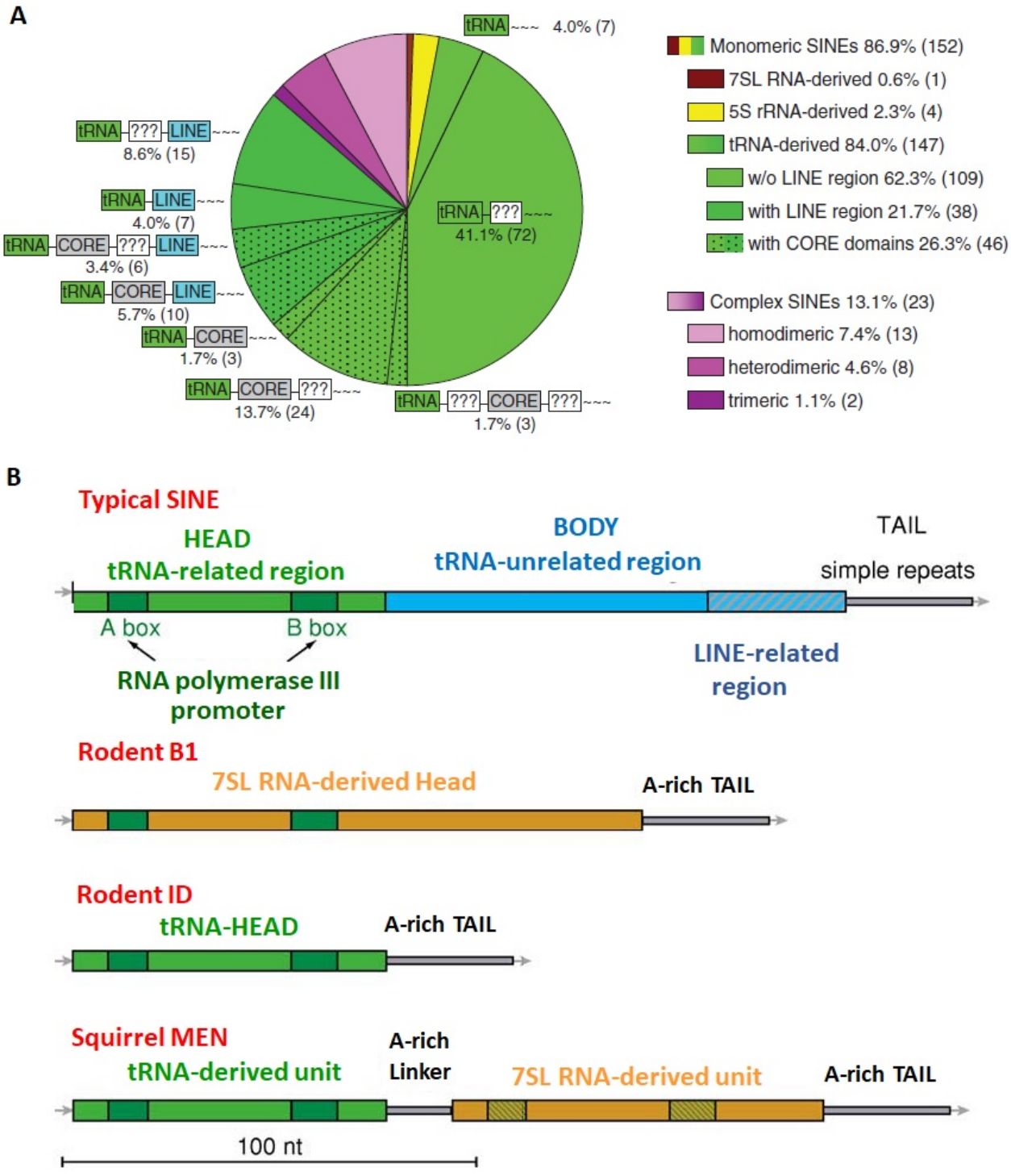


Figure 3 SINEs have evolved with different structures and complexities. (A) Occurrence of different kinds of SINE TEs (modified from Vassetzky and Kramerov, 2013). Majority of SINEs, (41% of 175) are simple and derived from tRNA pseudogenes (green, dark and light green represent tRNA derived SINEs with and without LINE-derived regions, respectively). Some SINEs have core domains, indicated with dots. The brown and yellow chart portion represent 7SL RNA and 5S rRNA0derived SINEs, respectively. Complex SINEs are represented in shades of purple, light to dark purple indicates increase in complexity and the question marks indicates unknown sequence origin.

(B) SINE structure examples (modified from Kramerov and Vassetzky, 2005). A typical SINE structure has a tRNA derived Head, a body of unknown sequence source and a 3' region comprising of sequences from a partner LINE and simple repeat tail. Simple SINEs such as Rodent B1 and ID are derived from 7SL RNA and tRNA, respectively. Squirrel/Men is a heterodimeric SINE from tRNA/7SL RNA

1.6 TEs as functional domain units in lncRNAs modular structures

Since the large portion of the mouse and human genomes are transcribed (Carninci et al., 2005; Djebali et al., 2012), repetitive sequences can also be transcribed in specific regulated fashion. For instance, the FANTOM consortium, Katamaya et al., 2005 reported that 18% (14,828) of capped-RNAs transcription start at repetitive genomic regions, which are regulated in cell line-specific fashion. This suggests that the majority of transcripts harbors repeat sequences like TEs in driving functional activities. Therefore, the exaptation of TEs in all genomic regions such as promoters, enhancers, chromatin barriers, gene bodies (exons, introns and untranslated regions, UTRs) and intergenic areas may have contributed to the evolution of regulatory networks (De Souza et al., 2013) in driving cellular processes such as proliferation, differentiation and development. Analysis of annotated human genes have revealed that the vast majority of lncRNAs, (7,710, 83.4%) overlap with at least one known TE. This is in sharp contrast with the observation that only 6.2% of protein coding sequences overlap TEs. Importantly, about 41.9% of lncRNA transcript sequences are TE-derived with LINE1 and SINE Alu families the most prevalent ones, accounting for 29% of all annotated sequences, while LINE2 and MIR as well as some DNA TE, hAT-Charlie and TcMar-Tigger are significantly depleted (Kelley and Rinn, 2012). Multi-family TE composition could be observed in lncRNA such as TUG1 and Linc-ROR ***(Figure 4A)*** that respectively function in mediating histone modification and cellular reprogramming (Kelley and Rinn, 2012). In particular, Linc-ROR has seven different TE families in its sequence ***(Figure 4A),*** comprising of endogenous retrovirus 1 (ERV1) LTR and human endogenous retroviral H (HERV-H) elements at the transcription start site (TSS) and are suggested to mediate Linc-ROR function in regulating cell stemness and proliferation processes (Kelley and Rinn, 2012).

Direct Alu and IRAlus elements in lncRNAs can promote STAU1- mediated mRNA decay (SMD) regulation, via duplexing through partial complimentary base-pairing with mRNA 3' UTR Alus (Elbarbary et al., 2016; Gong et al., 2013) (Elbarbary et al., 2016, Gong et al., 2013). SMD is also promoted by 3'UTR sequences of mRNA containing SINE identifier (ID) and SINE Bs (Gong et al., 2013).

The AS lncRNA functional family SINEUPs may represent another example of simple-family TE found in mouse and humans. The mouse SINEUPs comprise of inverted SINEB2 and Alu SINEs that are organized in modular structures ***(Figure 4B)*** in regulating translation during stress (Carrieri eta al., 2012 and Zucchelli et al., 2015a/b), and human SINEUP of R12A-AS1 and ITFG2, respectively, have simple repeats of free left/right Alu monomer (FRAM) and mammalian-wide intersperse repeats 1b (MIR1b) regulating translation activation of targeted protein synthesis through an antisense sequence (Schein et al., 2016). Therefore, there are currently evidences for functional TEs in domain units within lncRNAs such as Linc-ROR, TUG1, SINEUP of AS Uchl1 ***(Figure 4A)*** that modulate coding genes expression.

1.7 lncRNAs as organized modular structures and functions

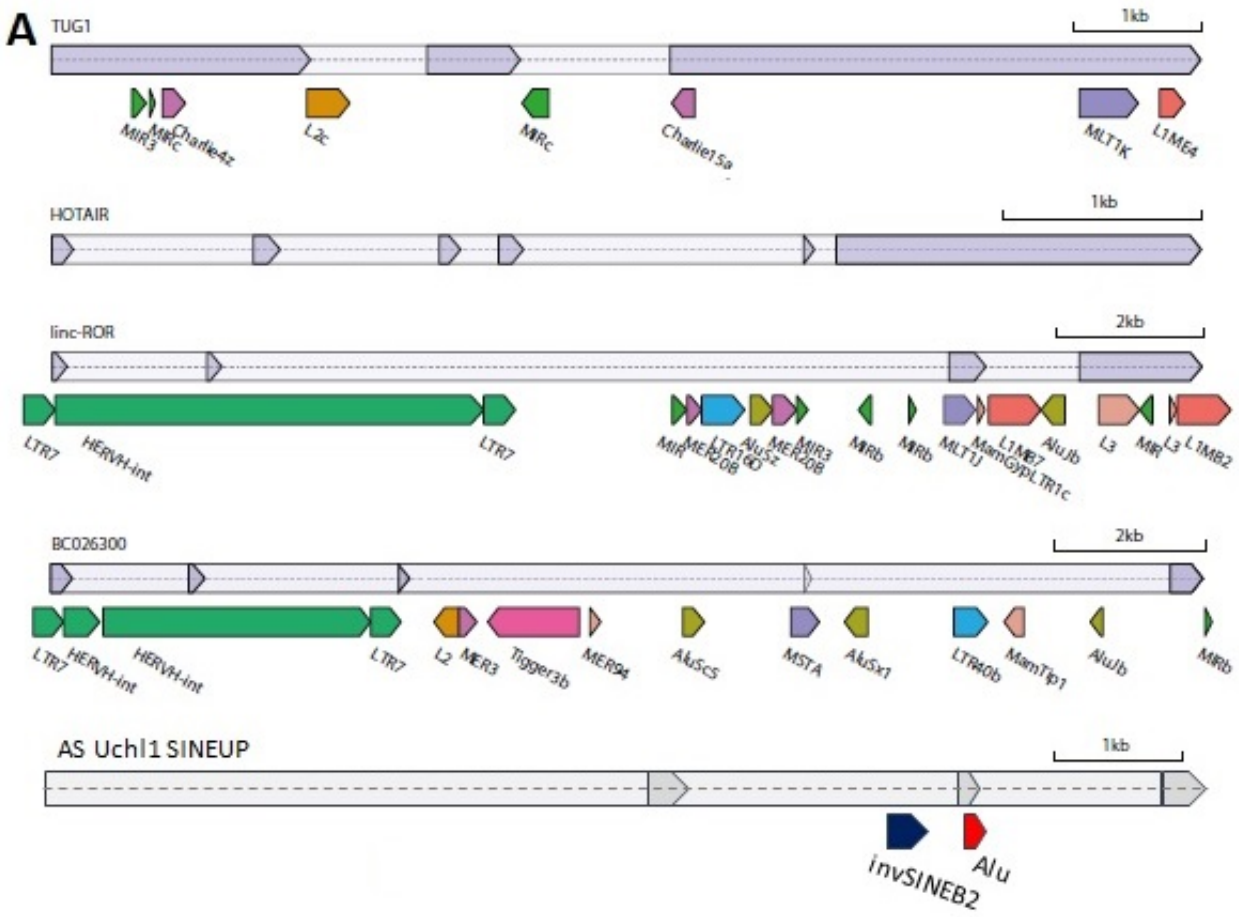
As it was recently proposed by (Guttman and Rinn, 2012), lncRNAs, like proteins, follow a modular organization consisting of discrete domain units that can act in combination to drive functionality by assembling diverse combinations of proteins, RNAs or DNAs. Therefore, this lead to the suggestion that expressed lncRNA sequences containing TEs or exonic TEs may act as repeat insertion RNA domains that are essential for lncRNAs function (Johnson and Guigó, 2014), providing recognition surfaces for interactions with DNA, other RNAs (mRNAs or ncRNAs) and proteins in driving cellular processes. From evolution perspective, domain organization explains how insertion and rearrangement of functional subunits can alter the function of existing genes or creation of novel ones as duplication from existing functional sequences rather than continual de novo evolution (Johnson and Guigó, 2014). The modular organization of lncRNAs is therefore a way by which distinct functionalities can be preserved by distinct discrete sequence units that are separated by flexible linkers, and may provide independent interactions and function (Guttman and Rinn, 2012). The lncRNAs functional domains are likely to act either by sequence-based hybridization to nucleic acid targets or adoption of secondary structural motifs that mediates the interaction with protein or RNA partners and hence offering biological activity (Johnson and Guigó, 2014). In this way, depending on the lncRNA functional activity, both proteins and nucleic acids like mRNAs and other ncRNAs or even DNA sequences can be presented together in a complex through binding interactions to mediate a biological process.

As an example, SINEUPs are organized into modular structures consisting of a Binding Domain (BD) and an Effector Domain (ED) with functionality in translation activation. The BD mediate S/AS bas-pairing between the SINEUP and targeted mRNA, while the ED may exitbit protein-RNA, or RNA-RNA

interactions to recruit specific complexes for functionality. The ED of SINEUP is composed by an invSINEB2.

The lncRNA XIST is another example with modular structures containing TEs. It has two distinct domains (figure 4B); a silencing domain known to bind to PRC2 and a localization domain that binds to YY1 protein to deposit repressive markers on the chromatin in causing transcriptional silencing of genes on the X chromosome (Wutz et al., 2011; Jeon and Lee, 2011). It is known that six out of 10 XIST transcript exons have various tandem repeats that originate from ERV, DNA transposons, SINEs and LINE 1 TEs (Elisaphenko et al., 2008). However, lncRNAs may be modular with domain without known TE elements as in the case of HOTAIR (Kelley and Rinn, 2012). In HOTAIR one domain acts as RNA scaffold to present proteins of the PRC2 and LSD1/CoREST/REST complexes to repress HOXD gene expression, through chromatin remodeling effects (Gupta et al., 2010; Tsai et al., 2010).

Overall, lncRNAs evolution has been shaped by TE insertions organized into domain units, which presents preformed structures and sequences motifs conferring molecular interactions and functionalities to their host lncRNAs.


A
TUG1
1kb
MIR3
MIRc
Charlie4z
L2c
MIRc
Charlie15a
MLT1K
L1ME4
HOTAIR
1kb
linc-ROR
2kb
LTR7
HERVH-int
LTR7
MIR
MER20B
LTR16D
AluSz
MER20B
MIR3
MIRb
MIRb
MLT1J
MamGypLTR1c
L1MB7
AluJb
L3
MIR
L3
L1MB2
BC026300
2kb
LTR7
HERVH-int
HERVH-int
LTR7
L2
MER3
Tigger3b
MER94
AluScS
MSTA
AluSx1
LTR40b
MamTip1
AluJb
MIRb
AS Uchl1 SINEUP
1kb
invSINEB2
Alu

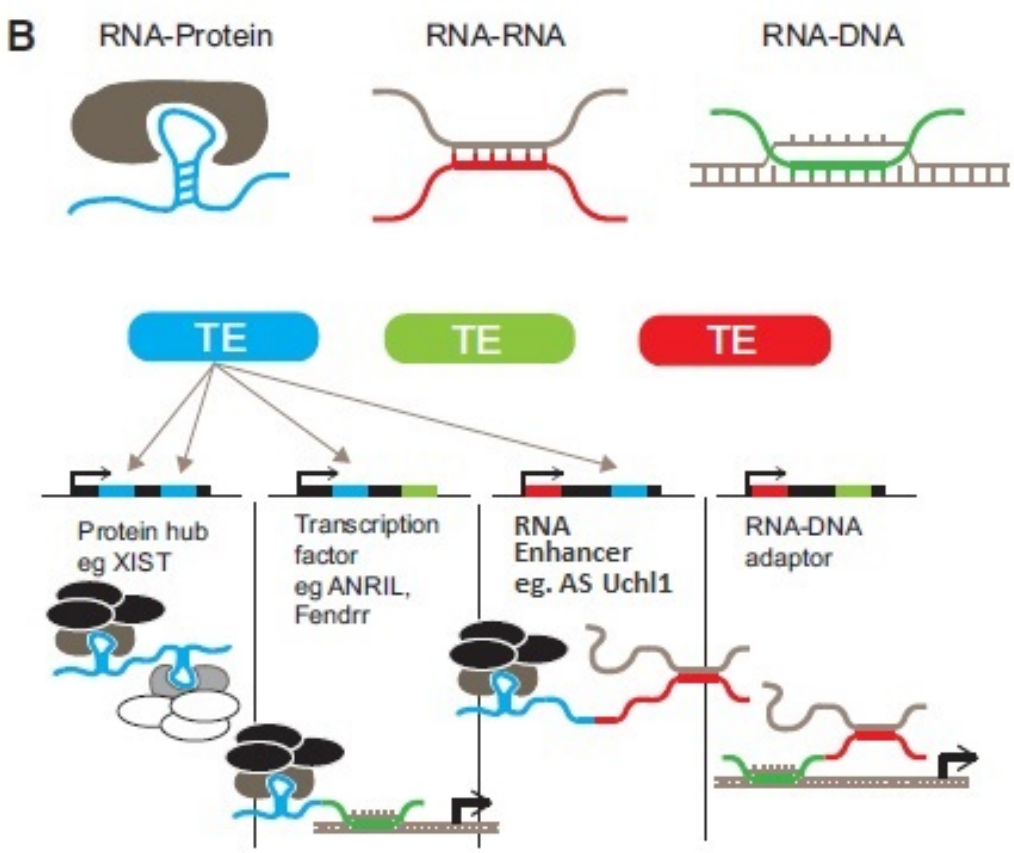


Figure 4 Activities of represented inserted TE domains of lncRNAs. (A) scheme showing some examples of lncRNAs with functional embedded TEs of various types (modified from Kelly and Renin, 2012). TEs are colored by different colors matching different families in each lncRNA. Some lncRNAs such as TUG1 and Linc-ROR are multi-family TE containing lncRNAs while AS Uchl1 SINEUP may represent a simple-family TE, having 2 known TEs, invSINEB2 and Alu organized into modular structures in driving functionality. However, HOTAIR is a typical lncRNA known to have modular domains, but it is devoid of TEs. Linc-ROR and BC026300 are lncRNAs with several TEs including LTRs in the TSS. (B) Functional classification of lncRNAs with inserted TE domain interactions (modified from Johnson and Guigó, 2014). TEs in modular domains acting as RNA scaffolds to recruit proteins, through RNA-protein interaction (depicted in blue) or hybridize with nucleic acids such as RNA or single strand DNA during open chromatin, through reverse or direct complementary Watson-Crick or other form of base-pairing forming RNA-RNA pairs (depicted in red) or RNA-DNA (depicted in green) hybrids (upper panel B). This pairing may lead to further recruitment of molecular interactors leading functional regulation as represented by XIST, ANRIL and AS Uchl1 SINEUP (lower panel B). While some lncRNAs like XIST may exhibit pairing with only proteins or nucleic acids, others like SINEUP of AS Uchhl1 may pair with both molecules in mediating functional activities.

1.8 SINEUPs: Antisense (AS) lncRNAs with SINE elements that UP-regulate translation

The combination of CAGE analysis with paired-end sequence (Wei et al., 2004) technology by the FANTOM consortium has led to the discovery of pervasive transcription in the mouse genome, with over 70% of genes showing some forms of antisense transcription (Katayama et al., 2005). Natural antisense transcripts (NATs) are transcripts of both coding and non-coding RNAs (figure 5A) which are transcribed from the opposite forward DNA strand of a gene locus (Lapidot and Pilpel, 2006). NATs can either be cis-NAT, which are transcribed from the opposite strand of the same gene locus and has perfect overlapping with the sense gene or trans-NAT that are mainly transcribed from different gene locus, and display imperfect complementarity, but are capable of targeting and regulating several sense RNAs (Lapidot and Pilpel, 2006). Although it is not clear to the extent of how both cis-NAT genes are transcribed, it seems that they are transcribed from independent or bidirectional promoter activities within the gene locus. Additionally, cis-NATs may also arise from independent cryptic promoter activity within intronic regions of genes, while trans-NAT may arise from the transcription start sites of adjacent genes or from functioning antisense promoters within intergenic regions (Lin et al., 2015). Depending on the transcription start site on both strands, the generated NAT may exhibit fully- or partial-overlapping S/AS pairs, while depending on the exact location of the overlapping regions in both transcript pairs, three main S/AS pairing classifications arises; 5' head-to-head divergent, 3' tail-to-tail convergent and fully overlapping configurations (figure 5B).

Most genomic regulations by NATs involve processes that either inhibit or activate transcription through events of chromatin remodeling and epigenetic mechanisms (figure 5C5 and 6) in repressing newly formed TEs insertions into the genome (Zener et al., 2017), and of physical blockage of incoming polymerases on the forward sense strand (Wight and Werner, 2013) known as transcriptional interference (figure 5C5). At the molecular levels, pairing may trigger recruitments of complexes to mediate RNA editing and chemical modifications of nucleotides (figure 5C1). RNA masking may promote alternative splicing or acting as a pseudogene RNA to protect the sense mRNA from RNAi (figure 5C1,2 and 3). Usually, fully-overlapping S/AS pairs form dsRNA, which may trigger protein kinase R activities in mediating immune response to repress expressed TEs or viral-like elements from the genome (Ilott et al., 2014). The NATs may also act as an RNA enhancer or chaperone to mediate cis-regulatory activity on the sense mRNA at the level of translation. A paradigmatic example are the SINEUPs (figure 5C4) (Carrieri et al., 2012, Zucchelli et al., 2015a/b, Schein et al., 2016).

NAT AS Uchl1, a lncRNA antisense to the mouse ortologue of the human Uchl1/Park5 gene, increases UchL1 protein synthesis at post-transcriptional level (Carrieri et al., 2012). AS Uchl1 is a head-to-head divergent antisense lncRNA, that partially overlaps with Uchl1 mRNA covering initiating AUG. When overexpressed in murine dopaminergic cell line, AS Uchl1 is able to increase UchL1 protein product without affecting its mRNA levels. Under physiological conditions, AS Uchl1 is retained in the nucleus. Upon rapamycin treatment, AS Uchl1 shuttles from the nucleus to the cytoplasm with an unknown mechanism. Once in the cytoplasm, AS Uchl1 induces Uchl1 mRNA association to heavy polysomes, thus increasing its translation.

AS Uchl1 can be considered the representative member of a new functional class of AS lncRNAs that utilize embedded inverted SINEB2 elements to increase translation of partially overlapping protein-coding genes acting on target mRNAs (Carrieri et al., 2012; Zucchelli et al., 2015b). These natural and synthetic molecules were named SINEUPs since their function requires the activity of an embedded inverted SINEB2 sequence to UP-regulate translation

SINEUPs are organized into two functional domains: i. the overlapping region, called BD, is antisense to the sense mRNA and confers target specificity at the 5'UTR; ii. an embedded inverted SINEB2 sequence, called the ED, which has functional activity (Zucchelli et al., 2015, Schein et al., 2016). This modular organization of SINEUPs suggest that embedded TEs are organized into discreet sequence and structural motifs that confer recognition by a complex consisting of proteins and RNAs; in this case these may be the translation machinery, initiation factors and ribosomal complexes (figure 4B, 5C4 and 6A). The BD confers target specificity through nucleic-acid interactions during or after transcription and may guide and shuttle it to the cytoplasm where the TE-containing ED domain are presented, acting as RNA scaffolds for recruitment of the necessary interactors for mediating biological activity in increasing gene expression at the level of translation.

This family of AS lncRNAs are also known to be conserved in humans where human SINEUP of AS phosphatase 1 regulatory subunit 12A (R12A-AS1) and integrin-alpha FG-GAP repeat containing protein 2 (ITFG2) are also able to confer translation activation on their targeted protein synthesis (Schein et al., 2016). The active units in these human SINEUPs are free left/right Alu monomer (FRAM) and mammalian-wide interspersed repeats 1b (MIR1b) TEs respectively in R12A-AS1 and ITFG2 mRNAs, indicating that TEs have evolved to be essential for lncRNAs function and contributed widely to mammalian genome evolution and functionality in regulating gene expression at the level of translation.

However, the molecular bases by which these elements are able to activate translation is still unclear. It is only known that both natural and synthetic AS lncRNAs are enhancing the association of targeted mRNAs to heavy polysomes (Carrieri et al., 2012; Yao et al., 2015), suggesting that the embedded TEs may contribute preformed sequence and structural motifs, acting as RNA scaffolds to bind and guide the targeted sense mRNAs to the translating machinery.

Indeed, structural studies by chemical footprinting reveals that a conserved terminal SL1 hairpin structure of invSINEB2 is a structural determinant required for AS Uchl1 ability to upregulate translation (Podbevsek et al., 2018), strengthening that embedded TEs functionality in SINEUPs and other lncRNAs relies on their modularization into discrete functional structural units.

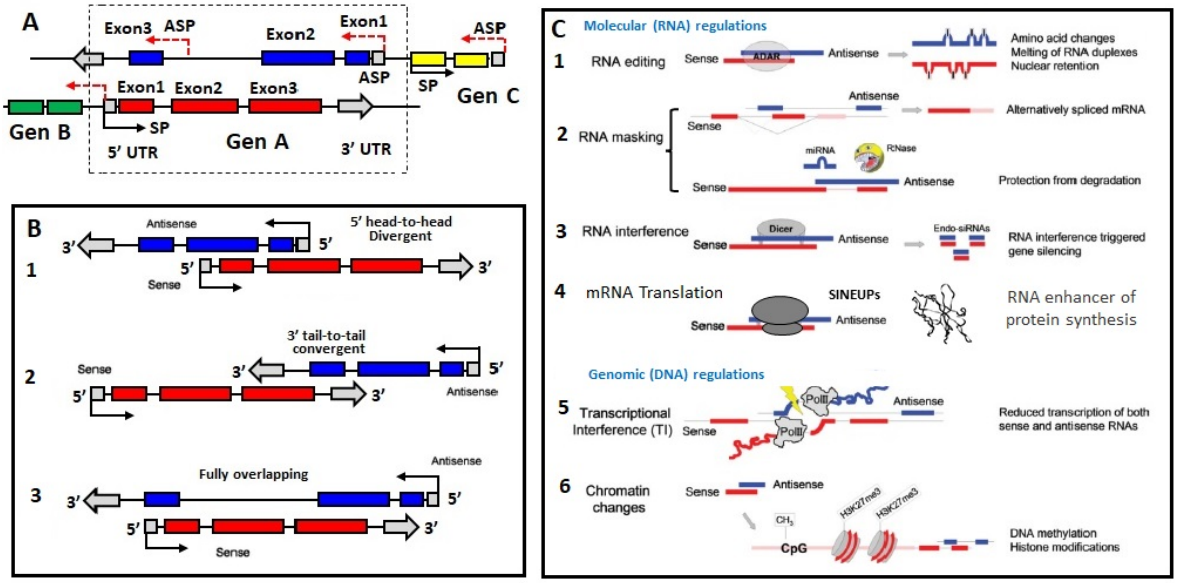


Figure 5 Genomic classification of bidirectional transcription and AS lncRNAs regulations. (A) Characteristics of bidirectional transcription from sense promoters (SP, black solid arrow) and antisense promoters (ASP, red reverse dotted arrows), resulting in sense and antisense transcription, respectively. The ASP can be located at any position within the gene locus, gene A; the 5'UTR (white box), exons (blue and red box, respectively for sense and AS transcripts), introns (straight line) and 3'UTR (bold arrow) and outside the gene locus; intergenic regions (between gene A and B or C) and upstream and downstream of nearby genes B and C, respectively). (B) classification of S/AS pairs. The sense transcripts are in red and antisense in blue. Pairing may be by 5' divergent head-to-head (B1), or 3' convergent tail-to-tail (B2) or fully overlapping (B3), forming double strand RNAs (dsRNAs). (C) schemes of various regulations by S/AS pairing at genomic (DNA) and molecular (RNA) levels (modified from Wight and Werner, 2015). The DNA regulations involves chromatin modifications through DNA and histone methylations (C6), leading to either transcription inhibition or activation of gene expression from the promoter (C6). Usually, di or trimethylation on histones or methylation at CpG residues leads to epigenetic gene silencing. At the DNA level, there could be transcriptional interference (C5). At the RNA levels, dsRNA formation may induce adenosine deaminase (ADARs) mediated RNA editing and modification of the sequence, leading to activation or repression of the gene (C1). The AS lncRNA can also induce alternative splicing post-transcriptionally or act as pseudogene to protect the sense mRNA from miRNAs (C2), enhancing gene expression. Also, dsRNA formation trigger RNAi pathways, leading to dicer mediated cleavage of the dsRNA and hence silencing of the mRNA (C3). Finally, the AS lncRNA may act as a SINEUP, an RNA enhancer of translation and guide the sense mRNA to the ribosome activating translation (C4). Thus, pairing by the S/AS may both be positive and negative, depending on the functional domain units of expressed AS.

1.8.1 Synthetic SINEUPs

SINEUP modular structure can be employed to artificially engineer their BD and design synthetic SINEUPs to specifically enhance translation of virtually any target gene of interest making them the first example of gene-specific inducers of protein synthesis. Synthetic SINEUPs so far have been proven effective with a number of targets, including GFP (Carrieri et al., 2012), FLAG-tagged proteins (Zucchelli et al., 2015b), secreted recombinant antibodies and cytokines (Patrucco et al., 2015), thus showing SINEUP technology is scalable, once provided the required target recognition with BDs.

Most importantly, synthetic SINEUPs can act on endogenous mRNAs in vitro and in vivo, as demonstrated by specific SINEUPs designed to target genes associated to neurodegeneration (PARK7/DJ-1 for PD) (Zucchelli et al., 2015b) and brain developmental disorders (Indrieri et al., 2016). Over competing technologies aimed to increase quantities of a protein of interest, SINEUPs present three major advantages: 1) Induce a 2-to-5-fold up-regulation thus limiting side effects due to exaggerated overexpression in gene therapy approach; 2) Act on endogenous mRNAs in situ, restricting translation enhancement to the time and space of endogenous gene expression, 3) Do not trigger hereditable genome editing. Similar to other nucleic acid-based therapies for brain disorders, the final outcome of SINEUPs-based therapy will depend on reaching their target mRNAs in the brain.

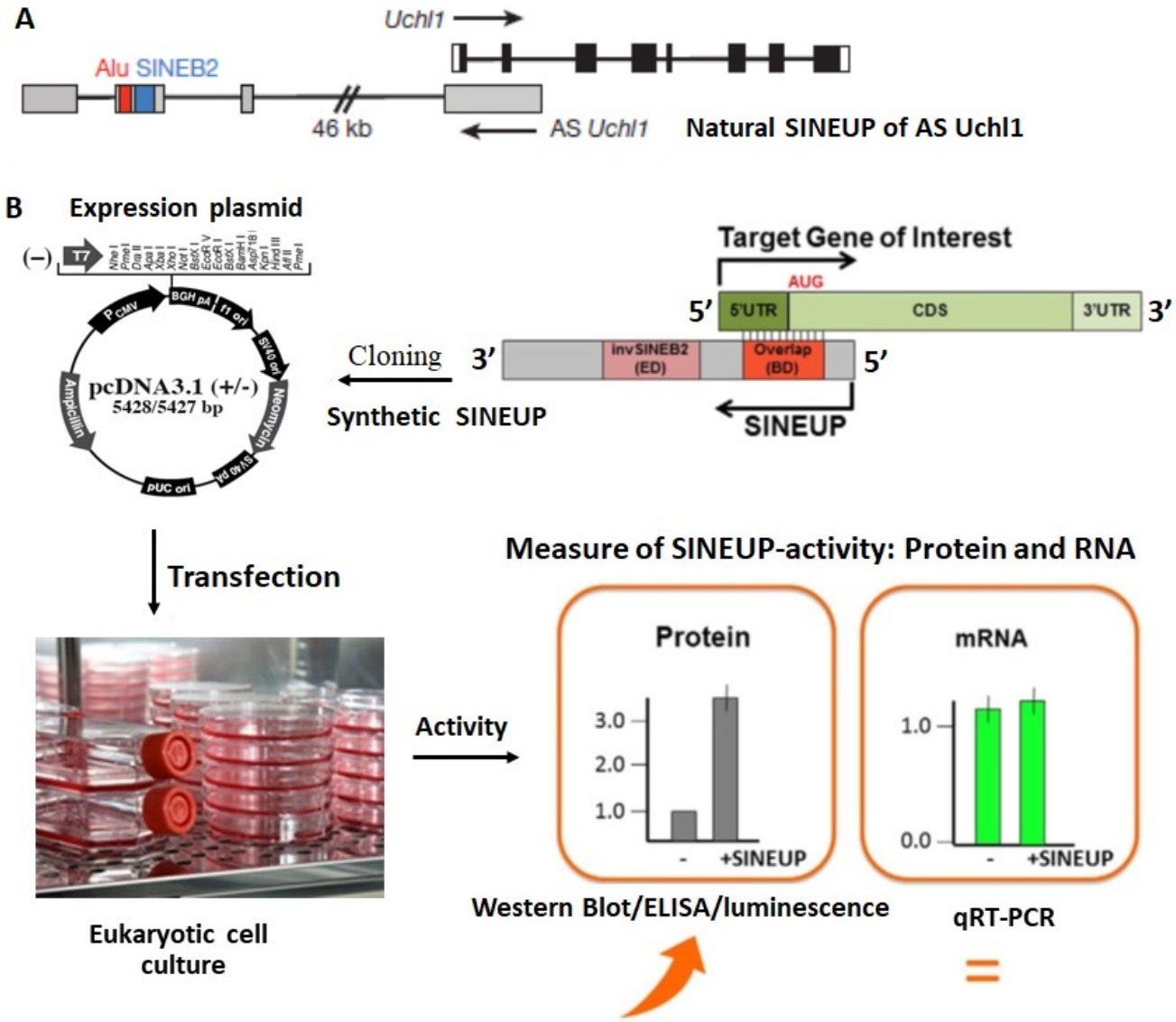


Figure 6. SINEUP modular structure and measure of SINEUP activity.(A) Scheme of S/AS pair, showing SINEUP of AS Uchl1 genomic organization (modified from Carrieri et al., 2012). Uchl1 exons are in black and 3' and 5'UTR are in white while targeted NAT, AS Uchl1 exons are in grey, invSINEB2 and Alu TEs are in blue and red, respectively. Introns are indicated as lines. S/AS pairing is 5'-head-to-head divergent. (B) SINEUP modular organization and measure of synthetic SINEUP activity (upper panel, right, modified from Zucchelli et al., 2016). SINEUPs activate translation and increase protein synthesis, using 2 main functional domains: The binding domain (BD) which provides target specificity and the effector domain (ED) which activates translation of the targeted mRNA. Following 5' to 3' modular structure of natural SINEUPs (A), synthetic SINEUPs can be created using S/AS pairing rules to any gene target at the 5'UTR around the AUG, by taking the ED of natural SINEUPs (example. AS Uchl1), separated by a short linker and adding about 72nt synthetic BD sequence that is antisense to the sense mRNA at the 5'UTR. BDs of 44nt are known to work (Zucchelli et al., 2015b). The generated synthetic SINEUP is then clone into eukaryotic expression vectors that is functional in the cell line of choice; SINEUP function in different mammalian cell lines has been established. Synthetic SINEUP construct is then used to transfect cells for the measure of SINEUP activity; by comparing protein expression with Western blot (WB) or ELISA or luminescence and

mRNA by qRT-PCR and levels compared relative to empty vector transfection controls. Activity is a measure of increase in protein levels ($p < 0.05$) without appreciable change in mRNA levels ($p > 0.05$).

1.9 Molecular Mechanisms of translational regulation of eukaryotic gene expression

1.9.1 Components of translational control

Eukaryotic gene expression is a tightly and complex multi-step process, involving several factors and different control levels. In this context, translational control allows for more rapid change of the cellular concentration of the encoded protein (Hinnesbusch and Sonenberg, 2009), which is critical for maintaining cellular and organismal homeostasis (Hershey et al., 2012). Although, most RNAs are known to have short half-lives, majority of eukaryotic mRNAs have quite long half-lives (Raghavan et al., 2002; Schwanhäusser et al. 2011); therefore for such genes, it is only at the level of translational control and protein stability that their encoded protein levels can be regulated (Hershey et al., 2012). Translational control is also important to eliminate foreign invaders like viral and microbial proteins synthesis, influencing their replication and spreading in infected cells; as such, it is the main control check point to regulate protein synthesis in oocytes and reticulocytes which lacks active transcription (Hershey et al., 2012), and very useful to conserved cellular energy, since protein synthesis consumes about 40% of total cellular energy. Protein abundance in cells can therefore be predicted mainly by the rate of mRNA translation (Schwanhäusser et al., 2011).

Translation is controlled by the main components of the translating machinery; ribosome and translating factors and features of the mRNA itself through several mechanisms, including synthesis and recycling of ribosome, translation factors and tRNAs (Sonenberg and Hinnesbusch, 2009; Jung et al., 2014), and secondary structures and cis-acting elements activation within the mRNA. As such, a given mechanism might affect translation of just a single, multiple or subset of mRNAs at a time. Specific regulations are often associated with features of the mRNA 5' and 3' UTR sequence and structures, whereas global regulations are often based on the activation or inhibition of one or more components of the translational machinery (Hershey et al., 2012). The cellular physiology influences the regulation of translating factors and their regulators though phosphorylation signaling cascades (Jung et al., 2014), leading to most eIFs inactivation and inhibition of 43S pre-initiation ribosomal complex (PIC) assembly on the eukaryotic mRNA 5'm^7G-cap-eIF4F complex to form the 48S initiation complex (48IC, figure 7). Therefore, the

cap-eIF4F and 43S PIC independent complexes formation are the major targets of global translational regulation of protein synthesis. The cis-acting elements in most mRNAs enable such mRNAs to be well-translated even under intensive global shutdown of translation through regulations of complexes formation; therefore, some mRNAs are capable of escaping global activation or inhibition of translation, leading to their proteins synthesis and influence in regulating and modulating cell physiology or fate.

1.9.2 Regulatory mechanisms in translation initiation

Translational regulation of any gene expression occurs mechanistically at 3 main stages, initiation, elongation and termination. The translation initiation step is considered the rate limiting, highly regulated and complex phase of the translation cycle (Hinnesbusch and Sonenberg, 2009; Hinnesbusch and Lorsch, 2012). It aims to assemble the 80S ribosome and position the initiator methionyl tRNA (met-tRNAi) at the initiation AUG (AUGi) of the mRNA (Aitken and Lorsch, 2012). In canonical eukaryotic translation, the 5' m^7G cap-structure of the mRNA, bound by the eIF4F-complex (figure 7), plays a critical role in translation initiation, as regulated by the eIF4E, the cap-binding protein component of the complex (figure 8A). The other rate limiting step of translation initiation is the recruitment of the mRNA to the 43S pre-initiation complex (PIC): the 43S ribosome (40S, eIFs 1, 1A, 3 and 5A and 5B), bound in the P site by the ternary complex (TC: eIF2, met-tRNAi and guanosine-5'-triphosphate, GTP) (Sonenberg and Hinnesbusch, 2009; Hinnesbusch and Lorsch, 2012, Hinnesbusch, 2017). After scanning 5'UTRs or the direct loading of the PIC to the initiation AUG start codon (AUGi) in a m^7G-cap-dependent (figure 7) or -independent manner, the 60S ribosomal unit is joined and the and the 80S bound ribosome-mRNA complex is assembled for translation (figure 7). In the final stage, GTP hydrolysis by eIF5B allows for 80S initiation complex formation and release of eIFs (Chen et al., 2016). Therefore, eIF4E within the eIF4F complex and the eIF2 of the TC are subjected to various forms of regulations (figure 8), and their activities modulation play critical roles in translation initiation.

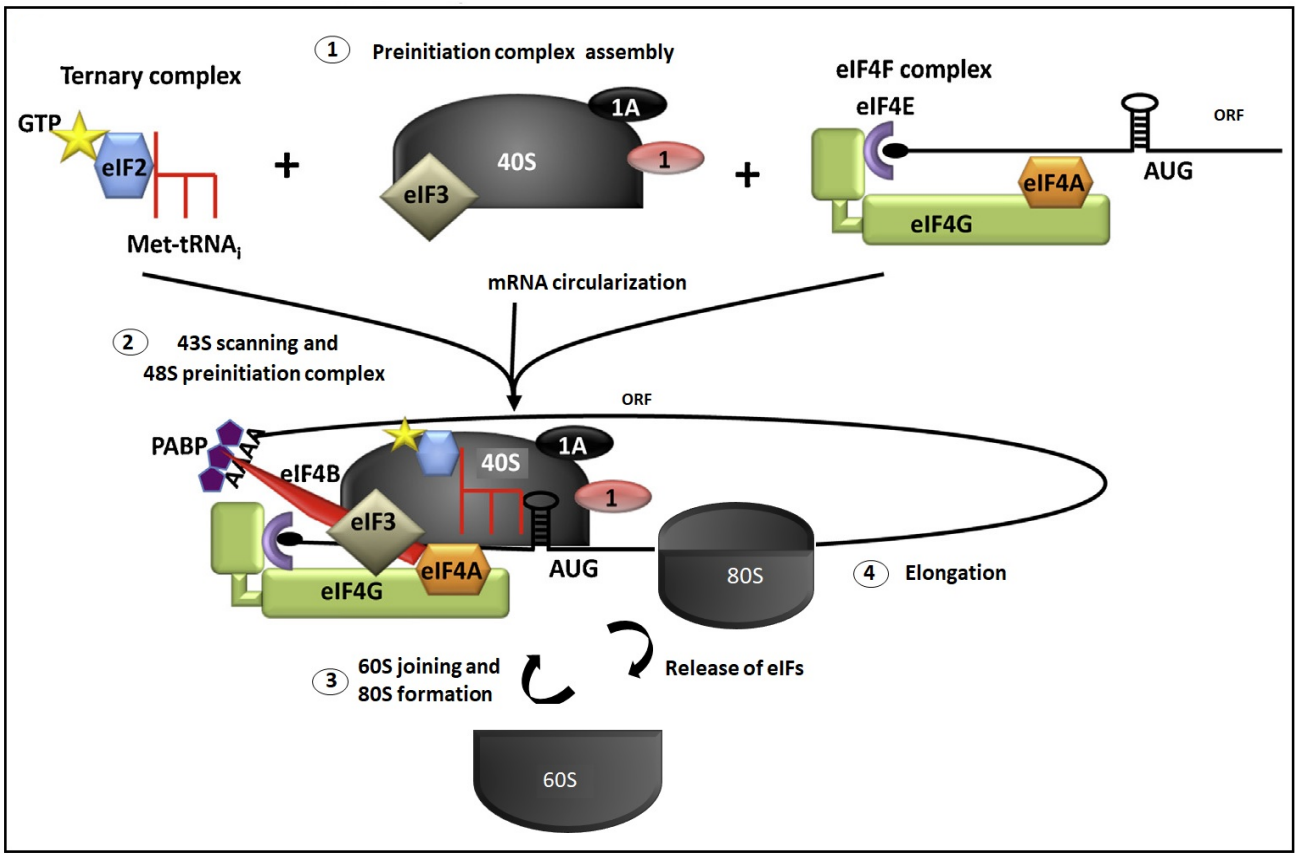


Figure 7. The m^{7}G cap-dependent translation. The 5' m^{7}G cap and the 3' poly-A are considered to mediate cap-dependent translation by connecting together through the eIF4F-complex. The small ribosome subunit is activated by preforming a 43S preinitiation complex (43S PIC), consisting of the 40S ribosome, ternary complex (eIF2, initiator methionyl-tRNA, met-tRNAi and guanosine-5'-triphosphate, GTP), eIF1 and eIF3, which has 5'end scanning on the mRNA. The m^{7}G cap-eIF4F complex is consisted of the cap-binding protein, eIF4E which is connected by eIF4G, the scaffold protein that also binds the helicase protein, eIF4A and the 3'end poly-A tail by the poly-A binding protein (PABP), circularizing the mRNA. The joining of the cap-eIF4F complex by the preform 43S PIC leads to the formation of the 48S initiation complex (48S IC) which scans the circularized mRNA to the initiation AUGi start site. The 60S subunit can then join the 48S complex to form 80S IC as the eIFs are released, entering into the elongation phase of translation. The scheme is modified from Somers et al., 2013. eIF; eukaryotic initiation complex, PIC; pre-initiation complex, S; subunit

1.9.2.1 Regulation of cap-dependent translation initiation, via eIF4F and 43S complexes assembly

All eukaryotic nuclear-pol II transcribed mRNAs have 5' m^{7}G cap-structure (Jung et al., 2014). As such, translation initiation involves binding of the 5'm^{7}G cap-structure by the heteromeric cap-binding complex, eIF4F composed by: i. the cap-binding eIF4E protein; ii. the protein scaffold eIF4G that links eIF4E and eIF4A (ATP-dependent DEAD-box RNA helicase) to the cap and poly(A)-binding proteins (PABPs) at the 3' UTR (Sonenberg and Hinnesbusch, 2009; Komar and Hatzoglou, 2010; Hinnesbusch

and Lorsch, 2012; Hinnesbusch, 2017); and iii.the ATP-dependent DEAD-box RNA helicase eIF4A that unwinds any other 5'-proximal secondary structures to facilitate the binding of the 43S pre-initiation complex (Jung et al., 2014, Merrick, 2015). The binding of eIF4G to cap and poly-A tail through PABP results in the circularization of the mRNA (figure 7) for promoting translation initiation (Sonenberg and Hinnesbusch, 2009; Komar and Hatzoglou). The binding of m^7G cap by eIF4F-complex through eIF4E is the main rate limiting step for cap-mediated translation and is regulated by the phosphorylation status of eIF4E-binding proteins (4E-BPs). Hypophosphorylated 4E-BPs tightly binds eIF4F-complex assembly and inhibiting the binding site of eIF4G (figure 8A) (Gingras et al., 1999; Lachance et al., 2002). Phosphorylation of 4E-BPs is under the control of mammalian target of rapamycin (mTOR) signaling pathways (figure 8A). mTORC1 kinase phosphorylates 4E-BPs (figure 8A, right), which then results in their release from eFF4E, promoting eIF4F-complex assembly and cap-dependent translation initiation (Gingras et al., 1999; Lachance et al., 2002). Many extrinsic processes, that lead to activation of cell-surface receptors such as growth factors, hormones and neurotrophic factors, generate signaling cascades that activate mTORC1(figure 8A), influencing protein synthesis (Jung et al., 2014). On the contrary, drugs like rapamycin and related analogs inhibit cap-dependent protein synthesis though mTORC1 inhibition, activation of 4E-BPs and subsequent inhibition of eIF4F-cap complex formation (figure 8A).

The main factor in the TC is eIF2, a heterotrimer protein composing of α, β and γ subunits that is associated with the energy source, GTP. The TC, together with the 43S, forms the 43S PIC that is recruited by the cap-bound eIF4-complex to assemble the 48S initiation complex (figure 7). This scans the 5'UTR in 5' to 3' direction to locate the AUGi, hydrolysis of GTP to guanosine-5'-diphosphate (GDP) and subsequent joining by the 60S subunit to form the 80S ribosomal initiation complex (Jung et al., 2014). The eIF2α subunit eIF2 is a regulatory target through phosphorylation at serine 51 (Ser51) by dsRNA activated protein kinase R (PKR), PKR-endoplasmic reticulum (ER) like kinase (PERK), hemin-regulated inhibitor kinase (HRI) and general control nonderepressible2 (GCN2) (figure 8B) (Krishnamoorthy et al., 2001; Chen 2014). It causes the increase in binding affinity for its guanine nucleotide exchange factor (GEF), eIF2B, hampering it catalytic subunits (γ and ε), leading to low levels of TC and therefore inhibition of 48S IC formation and subsequent decrease of global protein synthesis (Krishnamoorthy et al., 2001). This inhibition is usually activated as cellular response to viral infection, iron deficiency, protein aggregation, nutrients starvation (figure 8B), and other forms of environmental stresses (Chen 2014; Jung et al., 2014).

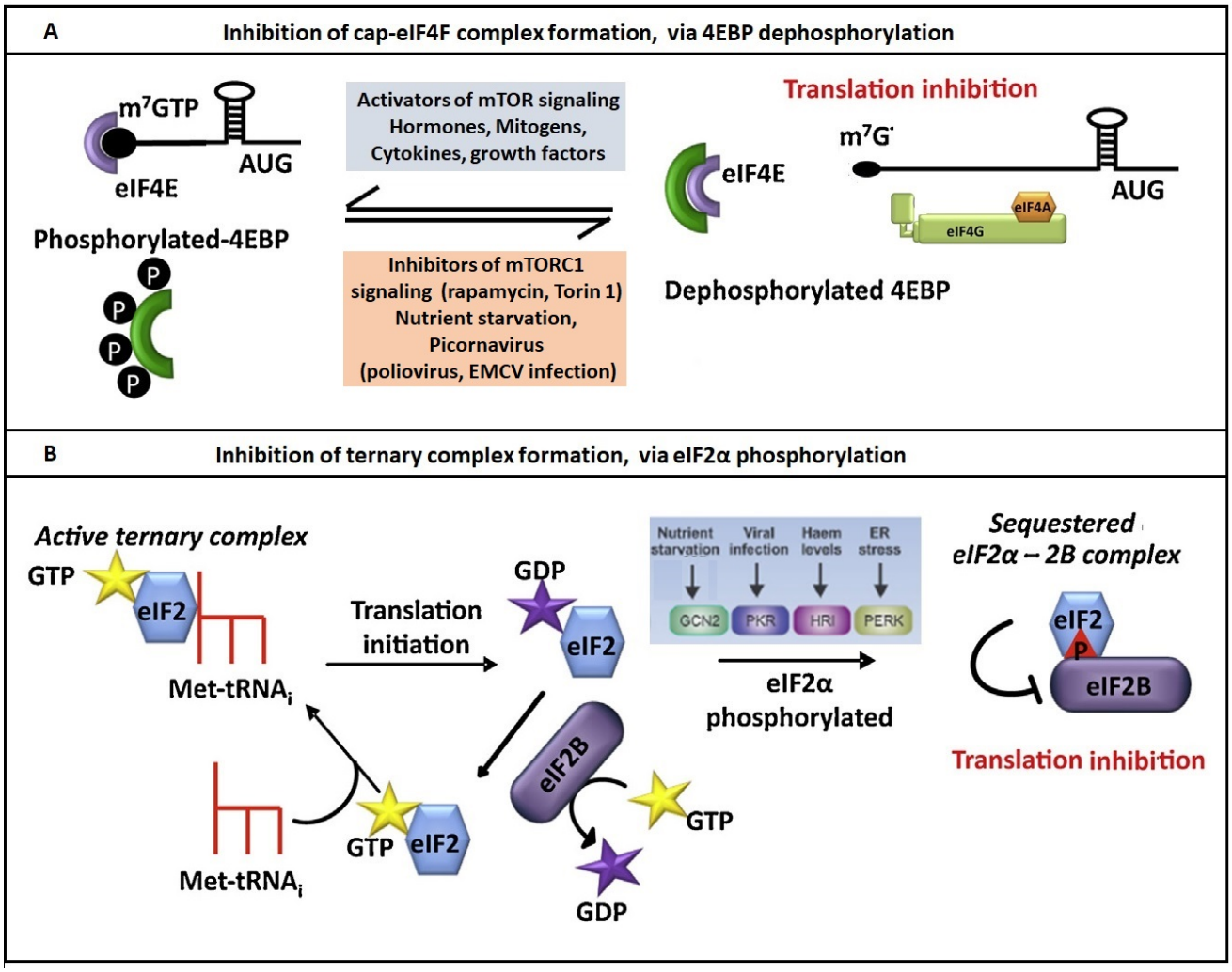


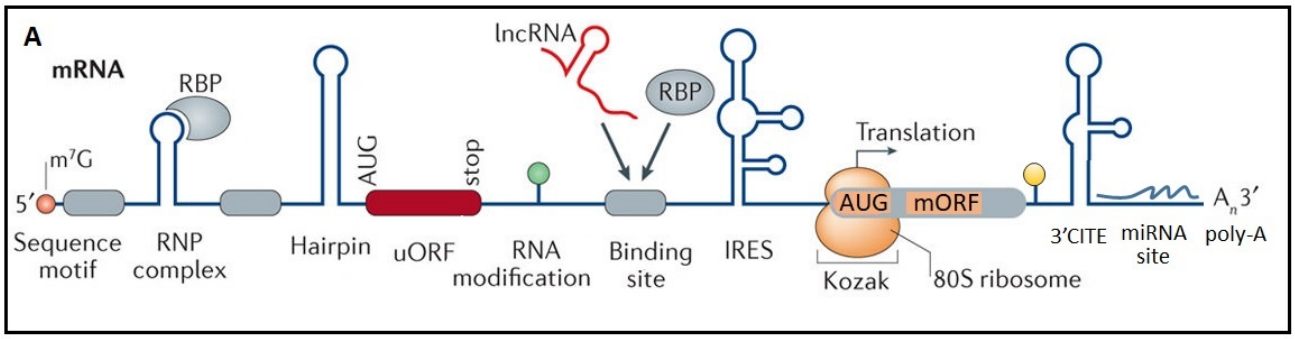
Figure 8. Signaling pathways inhibit eIF4F-complex and ternary complex (TC) formation, inhibiting global protein synthesis. Both eIF4F-complex and the ternary complex (TC; eIF2, initiator methionyl-tRNA and guanosine-5'-triphosphate, GTP) are required to form the 48S preinitiation complex (48S PIC) with the mRNA m^7G-cap. (A) For the eIF4F complex formation, the cap binding protein eIF4E is the main target of regulations by eIF4E binding proteins (4E-BP) phosphorylation and dephosphorylation mechanisms. The mTORC1 and other signaling pathways phosphorylate 4E-BP, deactivating it from binding to eIF4E allowing cap-eIF4F-complex formation and 48S PIC formation. Activators of mTORCI signaling agents like mitogens, hormone, growth factors phosphorylate 4E-BP, inhibiting it and allowing formation of cap-eIF4F complex formation. mTORC1 inhibition by viruses, nutrient starvation and rapamycin/torin1 activates 4E-BP, repressing cap-eIF4 complex formation and hence translation inhibition. (B) Nutrient starvation together, double strand viruses infection, iron metabolism can activate key kinases; general control nonderepressible 2 (GCN2), protein kinase R (PKR), PKR-endoplasmic reticulum (ER) like kinase (PERK), hemin-regulated inhibitor kinase (HRI), leading to phosphorylation of eIF2α and inhibition of TC formation. Both inhibitions prevent cap-dependent translation. Schemes are modified from Somers et al., 2013.

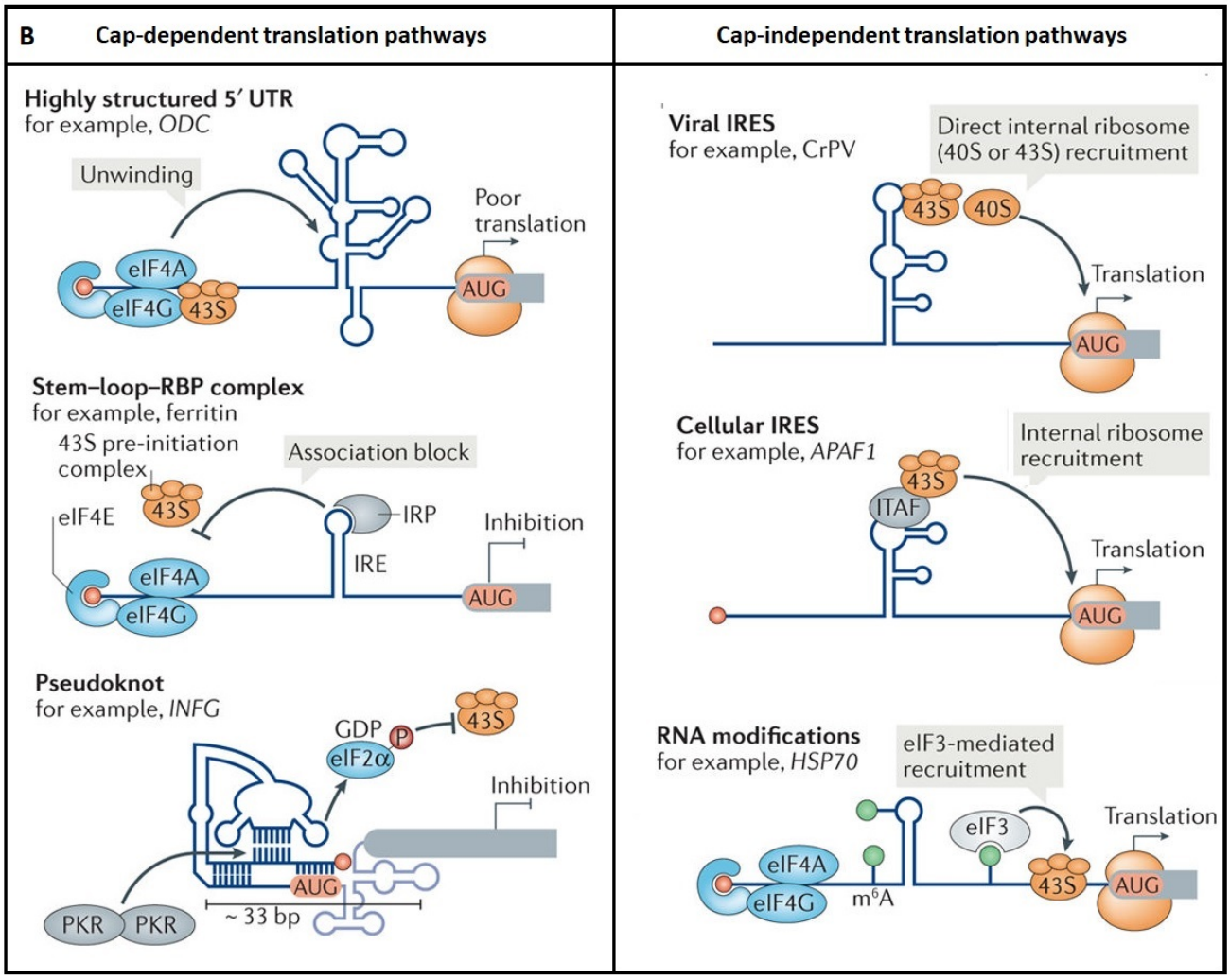
1.9.2.2 Cap-independent translation initiations and regulation by cis-acting elements

Various stress signals such as viral infection, iron deficiency, protein aggregation, nutrients starvation, double strand DNA breaks, heat shock, oxidative stress converge on initiation factors leading to inhibition of canonical global cap-dependent translation (Spriggs et al., 2010). However, a subset of nuclear-transcribed mRNAs and viral RNAs can be selectively translated by alternative cap-independent (CIT) pathways through the recruitment of some initiation complexes to the AUGi without the involvement of the 5' cap structure in their protein synthesis (Thompson, 2012). Two main mechanisms can be involved: i. by internal ribosome entry sites (IRESs); ii. cap-independent translation enhancers (CITEs) (Miller et al.,2007; Thompson 2012). Although, they may occur anywhere within the mRNA, IRES elements typically forms stable structures at the 5'UTR, while CITEs are typically found at the 3' UTR, but they can function at the 5' UTR as well (Miller et al., 2007; Thompson, 2012) and are very common in plant viruses (Miller et al., 2007). Therefore, some mRNAs contain cis-acting elements like IRESs and CITEs that confer specialized translation advantage for their protein synthesis under unfavorable cellular conditions as in stress.

1.9.3 Untranslated regions (UTRs) and cis-acting translational regulatory elements

Any transcribe mRNA in all domain of life retain some non-coding sequences upstream the start codon and downstream the stop codon, respectively at the 5' and 3' UTRs of the transcript. The primary purpose of the mRNA is to encode for a protein, however it harbor within the introns and UTRs much information to regulate it (Mignone et al., 2002; Xue and Barna, 2015). The UTRs harbor cis-acting elements such as 7-methyl guanosine (m^7G)-cap, poly-A tail, terminal oligopyrimidine tracts (TOPs), N6-methyl adenosine (m^6A), upstream open reading frames (uORFs) and internal ribosome entry sites (IRESs) that are mainly involved in regulating translation. The cis-acting elements interact and communicate with relevant trans-acting factors, which are primarily RBPs, but also ncRNAs and metabolites (Keene and Lager, 2005; Keene, 2007) to regulate generalized and specialized processes pertaining to gene expression.


A
mRNA
RBP
lncRNA
RBP
m7G
5′
AUG
stop
Translation
AUG
mORF
An 3′
Sequence motif
RNP complex
Hairpin
uORF
RNA modification
Binding site
IRES
Kozak
80S ribosome
3′CITE
miRNA site
poly-A


B
Cap-dependent translation pathways
Cap-independent translation pathways
Highly structured 5′ UTR
for example, ODC
Unwinding
Poor translation
eIF4A
eIF4G
43S
AUG
Stem–loop–RBP complex
for example, ferritin
43S pre-initiation complex
Association block
eIF4E
43S
IRP
IRE
Inhibition
eIF4A
eIF4G
AUG
Pseudoknot
for example, INFG
GDP
P
eIF2α
43S
Inhibition
AUG
~ 33 bp
PKR
PKR
Viral IRES
for example, CrPV
Direct internal ribosome (40S or 43S) recruitment
43S
40S
Translation
AUG
Cellular IRES
for example, APAF1
Internal ribosome recruitment
43S
ITAF
Translation
AUG
RNA modifications
for example, HSP70
eIF3-mediated recruitment
eIF3
Translation
eIF4A
eIF4G
m6A
43S
AUG

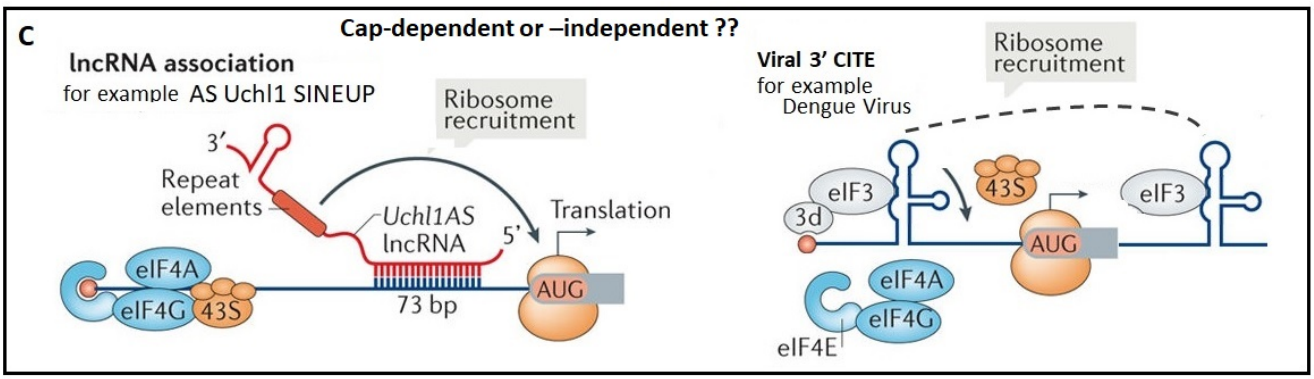


Figure 9. mRNAs harbor different cis-acting regulatory sequence and structural motifs to modulate cap-dependent and -independent translations. Scheme showing a typical mRNA with different motifs at the 5'UTR, coding sequence (CDS) and 3'UTR. Cap-dependent translation usually involves the 5'm^7G cap-structure and 3'UTR pol-A tail elements. Highly structured elements like pseudoknots, upstream open residing frames (uORFs), hairpins and 3'UTR microRNA binding sites represses translation, while cis-acting sequences like internal ribosome entry sites (IRES), terminal oligopyrimidine (TOPs), N6-methyl adenosine (m^6A) modification, and cap-independent translation enhancers (CITEs) promotes translation. (B) Samples of the regulatory elements activating or inhibiting alterantive translations; highly structured 5'UTR as in ornithine decarboxylase mRNA, Iron response element (IRE) in ferritin, pseudoknot structures in interferon gamma (IFNG) are known to inhibit cap-dependent translation (B, left), while IRES such as cricket paralysis virus (CrPV) and apoptotic peptidase activating factor (Apaf1), 5' UTR CITE like m^6A modification activate cap-dependent translation (B, right). (C) binding site sequences can attract both proteins and lncRNAs like SINEUP of antisense ubiquitin carboxyl-terminal hydrolase 1 (AS Uchl1, C, right) and other 3'UTR CITE as in the dengue virus 3'UTR (C, right) to promote complex forms of translation, because it is not yet known fully if these processes are cap-dependent or independent translation. Schemes modified from Leppek et al., 2017.

1.9.3.1 Cis-acting elements in 3'UTR and translation control

Sequences at the 3'UTRs, situated downstream the stop codon of the CDS, are involved in several regulatory processes including transcript cleavage, stability, polyadenylation, mRNA localization and translation and critical in determining the fate of the mRNA (Barret et al., 2013) in time and space. The 3'UTR is usually large and it sequences are more relaxed compared to other regions of the RNA, with potentiality of newly evolving regulatory elements (Barret et al., 2013). Alternative 3'UTRs influence mRNA sub-cellular localization, stability and translation providing transcriptome diversity (Mayr et al., 2016). 3'UTR sequences can also be cleaved acting as regulatory RNAs (Meyr, 2017). The regulatory cis-elements specific to the 3'UTR include i. miRNA response elements (MRE); ii. AU-rich elements (AREs); iii. poly-A tails and cytoplasmic poly-A elements (CPE) and iv. 3' CITEs (figure 9A). Here I will discuss CPEs and CITES.

The poly-A tail arises from the nuclear and cytoplasmic poly-A signal activities that lead to addition of adenine stretches to 3'UTR of mRNAs during transcription in the nucleus and post-transcriptionally in the cytoplasm. These sequences are recognized and bound by PABPs that influence transcripts export, stability, decay and translation (Barret et al., 2013). In particular, the poly-A may act as scaffold to promote the function of PABPs. In translation, the link of the 3'UTR tail to the 5'UTR through eIF4G scaffold protein found in the eIF4F-cap complex, influences cap-dependent translation. The cytoplasmic addition of poly-A plays a fundamental role in regulating translation (Villalba et al., 2011).

The 3' CITEs have similar functions to 5' UTR IRESs to promote CIT translation in eukaryotic cells. They are common in plants viruses, and represent diverse classes according to their requirement of recruitment of few or none of eIFs. They are known to form structures that promote long distance interaction with the 5'UTR through kissing-stem loop structures (Miller et al., 2007, Truniger et al., 2017). There are seven different classes which are structurally variable, and just like IRES, their structures are difficult to predicted but could only be identified functionally (Truniger et al., 2017). However, it is still unclear if CITEs represent separate structural and functional entities or variations of IRES (Thompson 2012). Just like m^7G cap and poly-A tail elements, IRESs and CITEs act as *cis*-acting regulatory elements capable of interacting with the ribosome directly with or without the involvement of all eIFs to mediate translation initiation at the AUGi. This ability is due to their highly organized structural and sequence motifs which are formed hitherto in response to key stress signals that inhibit cap-dependent translation initiation. Unlike IRES, all known indications showed that CITEs needs 5' end scanning to mediate ribosome position at the AUGi. Thus, CITEs function seems quite different from IRES, since they do not promote direct internal entry. However, their function supports emerging views that enriched 3'UTR IRES elements are also functional in mediating CIT, where they promote translation from the start codon by mRNA looping (Weingarten-Gabbay and Segal, 2016).

Overall, the 3'UTRs has evolved with diverse cis-acting elements (MRE, ARE, poly-A, CITEs/IRES etc) to recruit trans-acting proteins and play specialized local and cooperative roles in the regulation of gene expression, influencing phenotypic diversity in eukaryotes.

1.9.3.2 Cis-acting elements in the 5'UTR and translation control

5'UTRs are upstream the ORF of any transcribed transcript. Apart from leaderless mRNAs which basically have no 5'UTR sequences, most human mRNAs have a 5'UTR from 18 up to 2800nt long and

about 210nt on average length (Mignone et al., 2002) and in all eukaryotes the 5'UTR median lengths ranges from 53-218nts (Leppek et al., 2017). Variable 5' UTR sizes of the same gene are commonly produces as a result of transcription from multiple promoters and alternative splicing events (Zhang et al., 2004). These variations in 5'UTR length and structures suggest a great potential of regulatory signaling for translation. For instance, most mitochondrion transcripts have shorter 5'UTR and are known to be translated under energy stress (Hadar et al., 2015). Other regulatory relevant cis- acting elements specific to the 5'UTR that influence mostly cytoplasmic mRNA translation includes m^7G-cap, uORF, IRES, TOPs, m^6A motifs (Hadar et al., 2015)

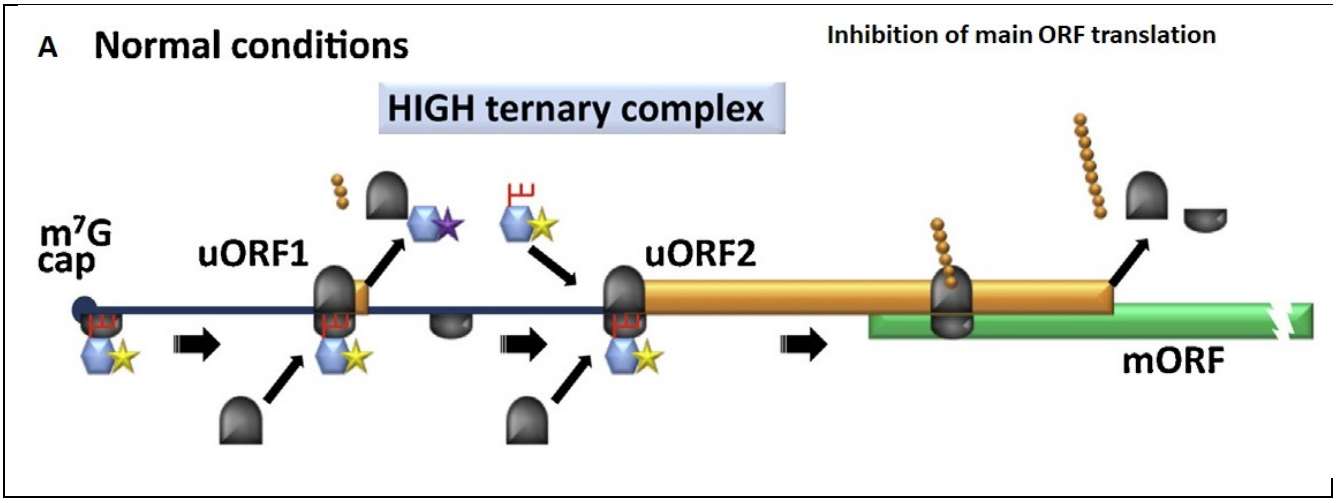
1.9.3.2.1 5' end m^7G-cap structure and trans-acting factors role in translation

The m^7G-cap (figure 9) is a regulator of cap-dependent translation initiation found on pol II transcripts. It is 7-methyl guanosine added to the 5' end by a reverse 5' to 5' triphosphate linkage to the first nucleotide of the RNA as a co-transcriptional process by RNA pol II in the nucleus (Banerjee et al., 1980; Moteki and Manley, 2000; Ramanathan et al., 2016). Some set of transcripts can be capped in the cytoplasm (Otsuka et al., 2009), indicating that the cap mediate further layer of regulations beyond what are currently known. In addition to mediating recruitment of the eIF4F complex during cap-dependent translation, the m^7G-cap also function to oppose alternative translation initiation, and it protect transcripts from 5' to 3' endonuclease activity, recruit proteins for splicing, polyadenylation, nuclear export (Ramanathan et al., 2016).

1.9.3.2.2 Upstream open reading frames (uORFs) role in translation

An uORF occur when there is random start and stop codon preceding a main start codon in a transcript. They could be considered as short translating sequences (figure 10). Genome-wide studies have recently revealed a wide spread of over 50% uORFs presence in the 5'UTRs of human transcripts (Ingolia et al., 2009), decreasing mRNA stability levels by 30% and protein levels from 30%-80% through translation regulation (Calvo et al., 2009; Barret et al., 2012). Translation of the uORF can basically influence how the downstream main ORF is translated. It is suggested this will depend on the number, length, and the context from the main ORF; stop codon distance or overlapping from the AUGi of the main ORF (Somers et al., 2013). The presence of the upstream (u)AUGi influences how the 43S complex scan the 5'UTR

either through a process of leaky scanning or termination-reinitiation mechanisms; thus the scanning 43S stops upon encountering the downstream stop codon before reinitiation from the main AUGi (Zhou et al., 2018). The ability of the uORF to dictate how a ribosomal complex behaves on the mRNA 5'UTR during normal and stressful conditions suggest that the uORFs have evolved in the 5'UTR of RNAs to regulate translation in response to different stresses. This could be seen in an adaptive response to mammalian nutritional stress by arginine/lysine transporter (cat-1) mRNA by which an IRES structure is locked up and become activated upon translation of an uORF (figure 10C) (Yamen et al., 2003). Also, a recent work has shown that reinitiation of translation on ATF4 main AUGi involves uORF regulation by m^6A methylation/demethylation during amino acid starvation stress (Zhou et al., 2018). Therefore, uORFs number, position and distance from the main ORF can have enhancing or inhibiting effects in regulating the response to the TC and 43S complex availability in different stresses. Importantly, the translation of the preceding uORF may thus regulate translation reinitiation rates from the main ORF.



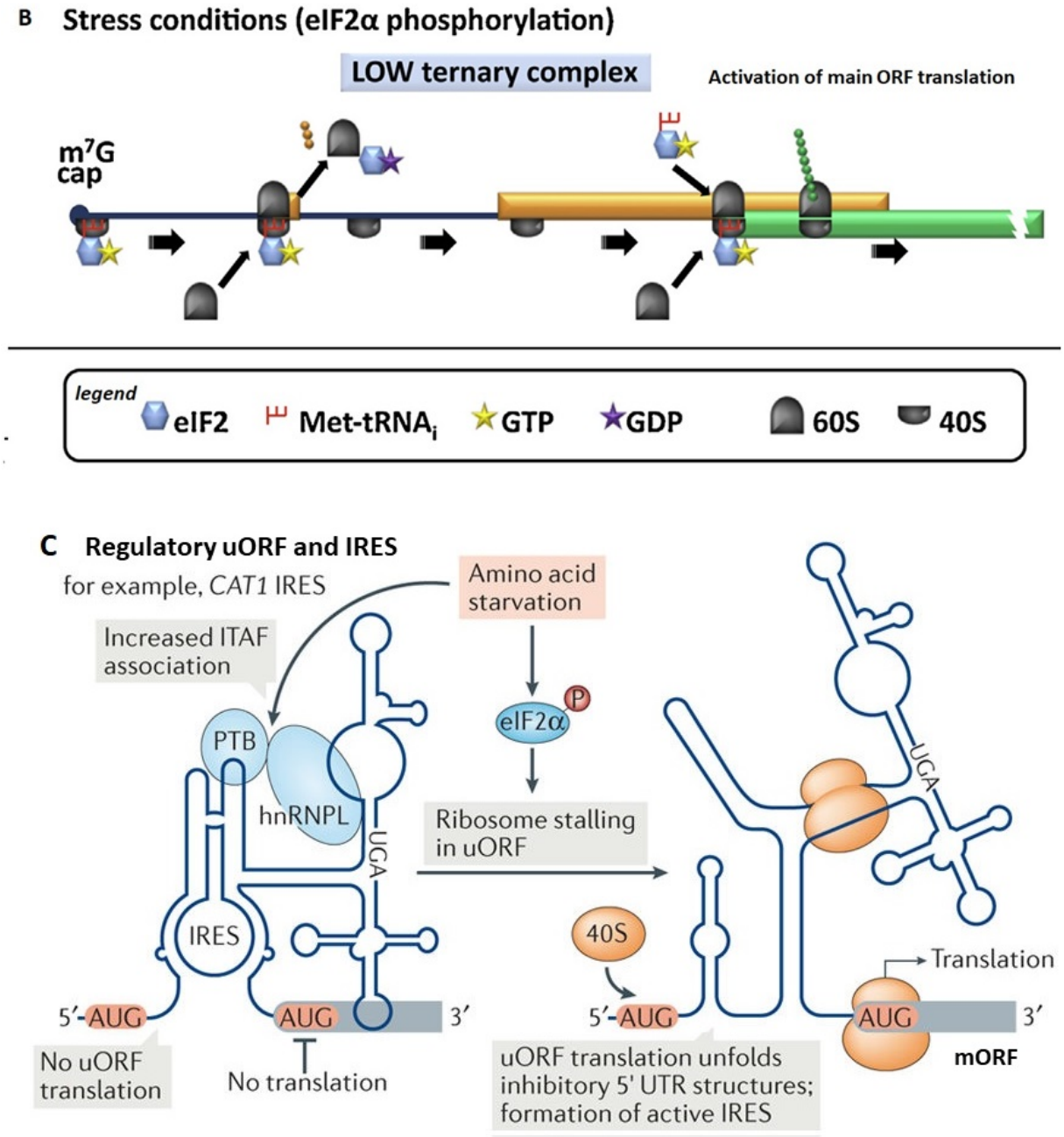


Figure 10. Upstream open read frames (uORFs) regulate main ORF translation under different cellular conditions. (A and B) scheme showing translation regulation in normal (A) and stress (B) conditions involving ternary complex (TC; eIF2, met-tRNAi and GTP regulations) (modified from Somers et al., 2013). In (A) reinitiation of translation occurs on uORF2 leading to repression of translation on the main ORF2. But in (B), upon some stress, there is less abundancy of the TC, so from uORF2, the ribosome continue leaky scanning through uORF2 until some TC could join the ribosome for initiation of translation from the main ORF AUGi and hence activation of translation of the protein. (B) Regulation of an IRES mediated translation by uORF. Scheme of the arginine-lysine transporter of amino acid transporter 1 (CAT 1) mRNA 5'UTR structural motif containing an uORF and IRES (modified from Leppek et al., 2017). In amino acid stress, the uORF translation is induced leading to the activation of IRES for the main ORF protein synthesis.

1.9.3.2.3 N6-methyl adenosine (m^6A) modification and it *trans*-acting factors role in translation

In RNA metabolism, post-transcriptional modification and editing of nucleotides may lead to changes in the sequence context of the mRNA, influencing structure-function relationship. m^6A, a reversible N6-methyl Adenine modification, is known to be the most common and widespread internal modification in eukaryotic mRNAs (Meyer and Jaffrey, 2014), and although it may occur at the 5'UTR and CDS, it is enriched near stop codons and 3'UTRs (Meyer et al., 2012; Liu et al., 2014). This modification involves recognition of the RRACH (R: A, G or U; and H: non guanine base) core consensus sequence (Zhou et al., 2016) by the enzyme catalyzing complex of METTL3, METTL4 and WTAP (Liu et al., 2014), which is reversible by FTO (Jia et al., 2012) and ALKBH5 (Zheng et al., 2013) (figure 11A). It can be found in other ncRNAs; tRNA and rRNA, and viral RNAs (Meyer et al., 2012), suggesting potential roles in several regulations including translation. m^6A roles in 3'UTR has been known to regulate mRNA translation and degradation, promoting increase in protein synthesis through interaction with YTHDF1 in physiological conditions (Wang et al., 2015), and METTL3 in cancer (Lin et al., 2016) (figure 11C). In the 5'UTR, m^6A promotes cap-independent translation in heat shock stress by direct interaction with eIF3 through 5' end scanning to initiate translation (Meyer et al., 2015, Zhou et al., 2015) (figure 11C). Furthermore, it has been shown that 5'UTR m^6A is the ultimate choice for eIF4F-cap complex-independent translation for non-terminal oligopyrimidine (TOP) tract mRNAs protein synthesis in response to stress (Coot et al., 2017). In summary, m^6A-modified 5'UTR and 3'UTR sequences use alternative pathways to recruit eIF3 and few other eIFs including some of the writers of m^6A, METTL3 and other trans-acting factors; YTHDF1/2 (figure 11C) to activate 5'end scanning cap-independent translation in mediating heat shock stress and it is the dominant mode of translation activation upon global induced stress in non-TOP mRNAs.

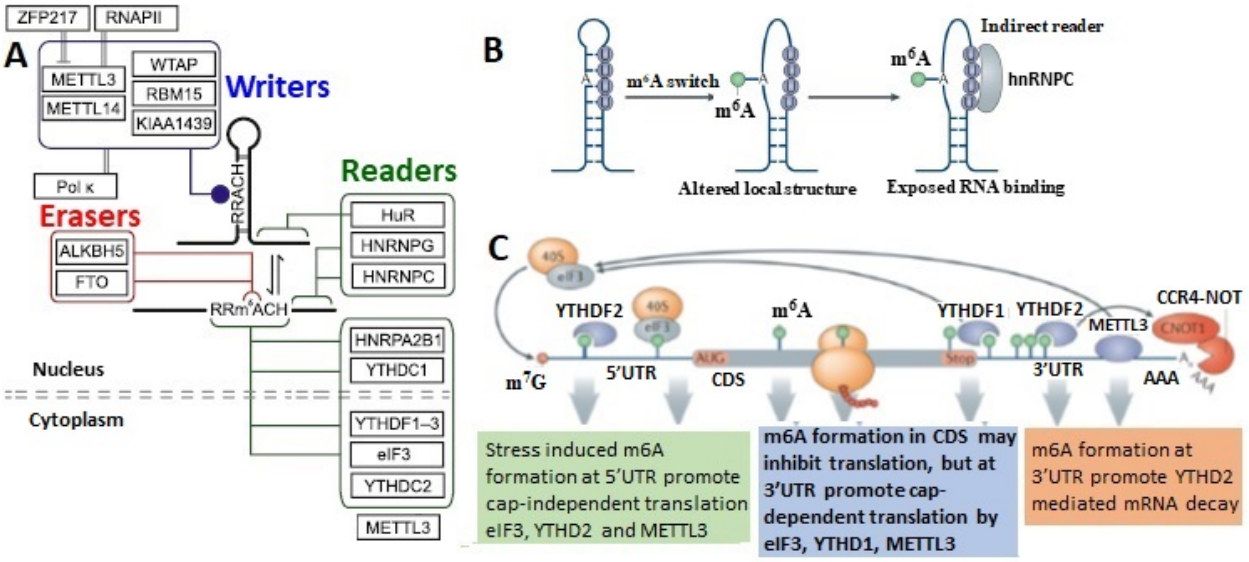


Figure 11. The regulatory roles mediated by N6-methyl adenosine (m^6A) modification on mRNA translation. (A) Scheme of the m^6A regulatory pathways, showing the known "writers" which establish the m^6A modification in RNAs, the "eraser" and "reader" proteins and their localized functions. The methylase complex (writers, blue) recognize and modify underlined adenosine in the RRACH (R; G/A and H; G/A/C) motif, which is prone to inhibition by ZNF217 protein. The demethylase (erasers, red); fat mass and obesity-associated protein (FTO) and alkB homolog 5 (ALKBH5) can reverse the m^6A modification while the readers (green) can bind to sequence or structural motifs induced as a result of the modification. The m^6A modification can occur and alter the mRNA structure from anywhere, opening it for RNA binding proteins as shown in (B) Upon the modification, structural changes can occur, allowing non direct readers like heterogenous nuclear ribonucleoprotein C (hnRNPC) to bind to poly-U tracts indirectly on the m^6A itself to regulate the mRNA stability, splicing, export or translation. (C) Modification can often occur within the coding sequence (CDS) and 3'UTR, but also at the 5'UTR under certain conditions like stress. Different readers read it according to the newly form structure or directly on the m^6A modification. The CDS modifications often lead to inhibition of translation. The 3'UTR m^6A modification can also activate cap-dependent translation when read by YTH domain-containing family protein 1 (YTHDF1) or methyltransferase-like 3 protein (METTL3), which doubled as a writer and reader. The YTHDF2 instead cause instability and mRNA decay. Under heat shock stress, 5'UTR m^6A modification could be activate, where eIF3 can be recruited to bind directly to just one m^6A to promote cap-independent translation of modified mRNAs. B and C; modified from Leppek et al., 2017.

1.9.3.2.4 Terminal oligopyrimidine (TOP) tract motifs and it *trans*-acting factors role in translation

TOPs are cis-acting sequences, starting with cytidine which is followed by 4 up to 15 pyrimidine nucleotides following the m^7G cap at the 5'UTR of TOP containing RNAs. Mostly translation initiation and elongation factors, PABPs, protein kinase C receptor, polypyrimidine tract-binding protein 1 (PTPB1 or HNRNP1) (figure 12) and 79 out of 80 ribosomal proteins contains 5' TOP motifs in their mRNAs

(Patursky-Polischuk et al., 2009; Pichon et al., 2012; Harvey et al., 2017), thus 5' TOP motifs regulate translation of the protein synthesis machinery (figure 12). The large number of proteins which mRNAs contain TOP motifs, indicates its importance in protein synthesis control (Pichon et al., 2012). Their translation seems to be under the control of mTOR signalling pathways, requiring activation by growth factors like insulin (Patursky-Polischuk et al., 2009).

In amino acid starvation, 5'TOP mRNAs are selectively inhibited through activation of kinases like GCN2. Inactivation of mTORC1 leads to their release from polysome fraction, accumulation in stress granules, leading to reduction in protein synthesis (Damgaard and Lykke-Anderson, 2011). Therefore, 5'TOP mRNAs signalling functions to promote translation under normal conditions, utilizing *trans*-acting factors and eIF4F components while repressing it during cell cycle arrest and nutrient starvation. However, not much is known about the trans-acting factors employed by 5'TOP mRNAs. One of the known proteins includes pyrimidine tract binding protein (PTB), which function in a complex with other RBPs, like hnRNP K and E1 in translation activation (Sawicka et al., 2008) an La antigen (Pellizzoni et al., 1996) and La-related protein 1, LARP1 (Lahr et al., 2016). In translation repression, 5'TOP mRNAs utilise TIA1 and TIAR stress-granule associated proteins for function (Damgaard and Lykke-Anderson, 2011).

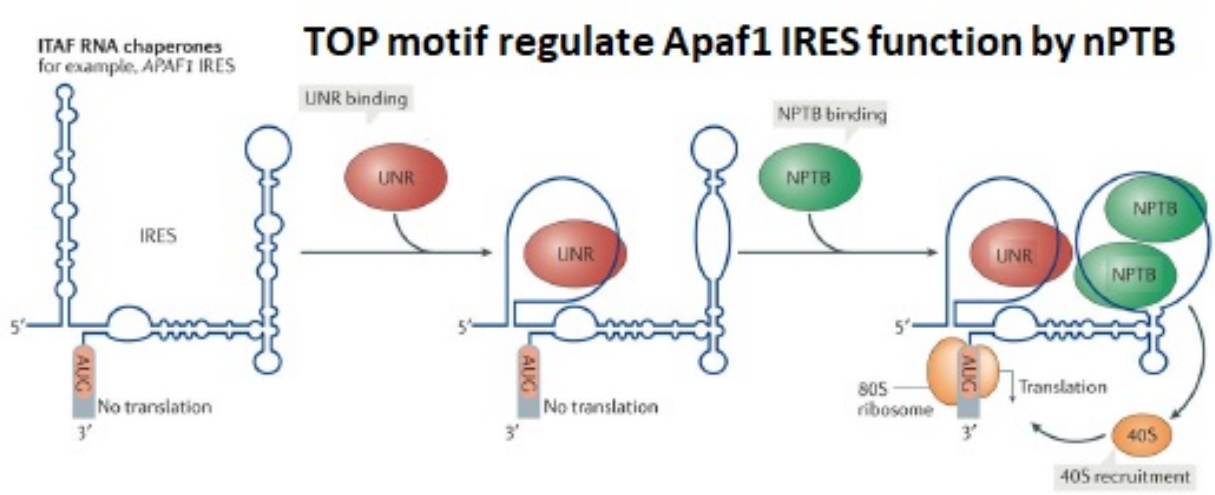


Figure 12. TOP motifs combine with other cis-acting elements to regulate cap-independent translation initiation. Scheme showing the roles play by IRES trans-acting factors (ITAFs) in modulating cis-acting elements function in ribosome recruitment during translation (modified from Leppek et al., 2017)**.** The ITAFs; upstream of N-Ras protein (UNR) and neuronal polypyrimidine-tract binding protein (nPTB) respectively bind sequentially to their respective cis-acting sequences within the Apaf1 IRES structure, which create a ribosome landing space fore recruiting the initiation complex for translation. The nPTB in particular binds to TOP motifs to activate IRES mediated translation of Apaf1 in stress.

1.9.3.2.5 Internal Ribosome Entry Site (IRES) elements and it *trans*-acting factors role in translation

IRES are cis-acting structural and sequence motifs in mRNAs that promote direct internal entry of a translational competent ribosome close to or at the initiation site for translation without involving 5'cap-structure in the ribosome recruitment. Depending on the class of IRES, there may be a little or no need for 5'end scanning of complex to the initiation site (Hellen and Sarnow, 2001). Thus, IRES allow internal ribosome recruitment to both capped and uncapped transcripts to promote translation mostly when the 5' end recognition or cap-dependent pathway is compromised as in stress, and during other cellular processes, like cell-cycle arrest or apoptosis, to ensure continual production of proteins for cellular activities (Hellen and Sarnow, 2001; Komar and Hatzoglou, 2011; Barret et al., 2012). As such, IRESs may utilize some cap-binding eIFs, cofactors or auxiliary proteins to initiate translation (Hellen and Sarnow, 2001; Hellen, 2009; Komar and Hatzoglou, 2011). These are trans-acting RBPs known as IRES trans-acting factors (ITAFs) and are thought to cause structural and conformational changes in the IRES

(King et al., 2010), enabling the recruitment of the translation machinery onto the initiation site of the mRNA.

IRES-mediated function was originally discovered in picornaviruses in the early 1980s and then later, the first cellular IRES was discovered in BiP mRNA (Hellen and Sarnow, 2001). Currently, there are several types of IRES from both viral and cellular origins. It is now known from systemic studies that there are about 10% and 20% cap-independent like elements in human and viral 5'UTRs respectively (Weingarten-Gabbay et al., 2016). IRES are not only associated to the 5'UTR, but can be found in all locations, including the ORFs (Gabauer and Hentze, 2016) and 3'UTR of human transcripts.

IRESs function using several different mechanisms. Some IRESs like cricket paralysis virus IRES (CrPV) and hepatitis C virus (HCV) can interact directly with the translation competent 40S ribosome (figure 13B and C) (Hellen, 2009) or, as for cellular IRESs like Gxt and IGFIR, they can directly base-pair with the 18S rRNA, a Shine-Dalgarno-like mode of translation initiation. Instead, Hepatitis A virus (HAV) and L-Myc, a cellular IRES, require eIF4G and E to recruit the ribosome for translation (IK et al, 2001; Spriggs et al., 2009). These different modes of recruitments suggest that some IRES just need few stretch of sequence like the 9nt sequence in Gxt IRES to function while others, like EMCV (figure 13B) require the multiple, highly structured regions. A high-throughput mutagenesis experiment revealed that IRESs can be group into "global" and "local" IRESs (Weingarten-Gabbay et al., 2016). The global IRES requires extended structural motifs to be active while the local IRES are defined by short sequence motifs like the 9nt sequence in Gxt to function by displaying complementarity with the 18S rRNA including helix 23 and 26 sites of the 40S ribosome.

1.9.3.2.5.1 Classification of IRES-mediated mechanisms

While the molecular mechanisms of how viral IRESs promote cap-independent translation are started to be well-defined in specific cases, much less is known about cellular IRES. At least in part, this is due to the observation that cellular IRESs possess m^7A cap-structure and therefore seem to switch between both cap-dependent and IRES-dependent modes in activating translation, even under normal cellular conditions. This is therefore, a main challenge in elucidating the mode of regulations mediated by cellular IRESs. Instead, most viral IRESs do not have the m^7A cap-structure and rely solely on other RNA elements like the IRES and other cis-acting sequences to mediate efficiently internal ribosome entry at the initiation site for translation. They use different modes of recruitment and hijack the host machinery

in the process, since most of them, except for the giant viruses, do not encode for the translation machinery components (Schulz et al., 2017), and thus they need to utilize the host components in a quick and efficient way to their advantage. IRESs are therefore categories into four classes (figure 13), according to their primary sequence, length and modular structures including all secondary and tertiary structures and overall mode of action for recruiting the translation machinery and initiation (Mailiot and Martin, et al., 2017). These classification are based on functional and structural studies, performed with SAXS, NMR, X-ray crystallography and Cryo-EM, but so far complete structures of complexes are only available for class III and IV IRESs, HCV and CrPV IRESs, respectively (Mailiot and Martin, et al., 2017).

1.9.3.2.5.1.1 Class I IRESs: Polio Virus (PV) IRES mode of activity

These IRES types (figure 13A) are members of *Picornaviridae,* and this includes enterovirus A71, coxsackie virus B3 (CVB3), human rhino virus 2 (HRV), and polio virus (PV) that has been studied extensively (Lin et al., 2009; Mailiot and Martin, et al., 2017). The PV-IRES containing region at the 5'UTR is 743nt long with modular domain units, DI-VI; the DI is C-rich and folds into cloverleaf structure known to be involved in viral replication, making the minimal window for IRES to be in the 100-743 region. DIV is C-rich and has stable hairpin structures that also recruits ITAFs; DV binds to partially protolyzed eIF4G scaffold protein (Gradi et al., 1995) that retains the eIF3/4A binding-sites but not that of eIF4E; a PTB ITAF region is found between DV and DVI; the minimal IRES entry region is located within DII-VI (figure, Mailiot and Martin, et al., 2017). All eIFs, except eIF4E, are required for PV IRES-mediated initiation with much reliance on eIF4G. It contains binding motifs for ITAFs such as poly-C binding protein 2 (PCBP2) and a PTB which act in the PV IRES entry window. In this context, PTB is known to modulate eIF4G binding to optimize the initiation complex internal ribosome entry for PV IRES activity (figure 13A) (Kafasla et al., 2010).

1.9.3.2.5.1.2 Class/type 1I IRESs: EMCV IRES mode of activity

These include cardioviruses, paraechovirus, aphtovirus such as well-known foot and mouth disease virus (FMDV) and encephalomyocarditis virus (EMCV). The well-known 439nt EMCV IRES is organized into DI-V (figure 13B). DIV is organized into a Y-shape structure and it represents the eIF4G binding

site. Just like the PV, EMCV utilizes almost all eIFs except the eIF4E and the N-terminal domain of eIF4G to activate the 40S ribosome to the entry site (Hellen, 2009; Lin et al., 2009). It requires PTB and other ITAFs for activity. PTB binds at the same site as eIF4B and other ITAFs (figure, Mailiot and Martin, et al., 2017). During EMCV infection, dephosphorylated and activated 4E-BPs are bound to and sequester eIF4E, inhibiting eIF4F complex formation and shutting off global protein synthesis in infected cells (Gingras et al., 1996).

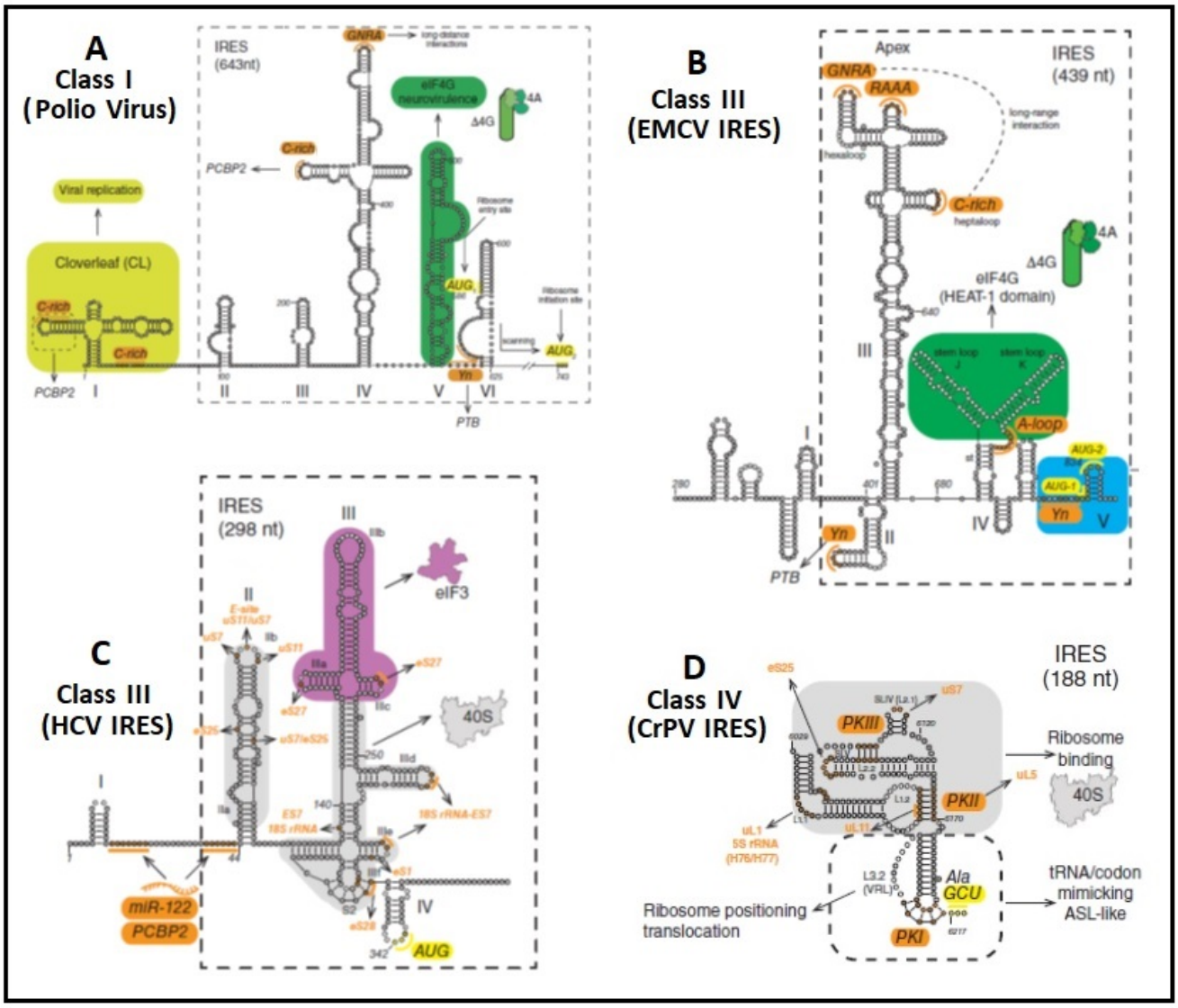
1.9.3.2.5.1.3 Class/type 1II IRESs, HCV-like IRES activity

The class III IREs are typical in the *Flaviviridae* family-like hepatitis C virus (HCV) and swine fever virus (CSFV). Additional members are the simian virus 2 (SV2), simian picornavirus type 9 (SPV9), bovine viral diarrhea virus (BVDV) and porcine teschovirus (PTV) (Hellene, 2009; Lin et al., 2009; Mailiot and Martin, et al., 2017). The HCV IRES (figure 13C) has been well studied due to it mode of action and its role in virulence. The HCV-IRES is organized into DI-IV domains. The DIII consist of subdomains IIIa-f, intervening helixes and a connecting complex pseudoknot structure, all link to DIV, a hairpin structure that is linked by the AUGi to the ORF (Hellene, 2009). However, DII-IV can fold independently, and are conserved in exception of DIV among all known HCV-like virus (Khawaja et al., 2015), and are all responsible for IRES-mediated activity. However, DIIIabcd recruits the 40S ribosome and eIF3 in forming the 43S initiation complex and are the determinant for IRES activity (figure) (Kieft et al., 2001). However, it is also assumed that the IRES can bind to translation competent 48S ribosome, but become trapped in mutant HCV-IRES (Ji et al, 2004). The eIF3 is also known to play a role in stabilizing direct binding of the met-tRNAi to the 40S at the p-site and 80S after the 60S subunit joins, following the loss of eIF1 and hydrolysis of eIF2 GTP to GDP (Hellen, 2009). Alternative pathways are known to coexist with the canonical pathway which are activated, especially in the time of eIF2 inactivation in stress, requiring eIF3/eIF5B, eIF2A or eIF2D and eIF1A, where compared to what is previously known, eIF1A now play role to stabilize the met-tRNAi and inspect the interaction at the p-site in the absence of eIF2 upon stress (figure 13E, right) (Dmitriev et al., 2010; Jaafar et al., 2016). Subsequent dynamic conformation changes occur after the 80S initiation complex has been formed, where the whole DII or DIIb is required to be positioned at the E site to enable the release of other initiation factors like eiF5B or 1A through the contact by ribosomal proteins like rpS5 within the decoding groove (Filbin et al., 2013). The HCV 5'UTR DI is also subjected to regulation by miR-122, a site where ITAF PCBP2 is also known to interact, and it may promote the translocation phase of the

ribosome or aid in the RNA replication (Masaki et al., 2015). Overall, HCV-like IRESs used it DIIId to directly recruit a translation competent initiation complex ribosome, relying mainly on eIF3 recruitment by DIIIabc and few other initiation factors; eIF1, eIF1A, eIF2 and eIF5B. The master recruiter, 40S associated eIF3 can interact independently with eIF1, eIF1A, eIF2 and eIF5B under varying cellular conditions to mediate cap-independent translation of virus proteins.

1.9.3.2.5.1.4 Class/type 1V IRESs, CrPV-like IRES activity

The intergenic region IRES (IGR) of Cricket paralysis virus (CrPV)and Taura shrimp virus (TSV), Israeli acute paralysis (IAPV), Platia Stali intestine virus (PSIV) belong to this class (Hellen, 2009; Mailliot and Martin, et al., 2017). This type of IRES are found between ORF1 and ORF2 sequences that encodes for structural and nonstructural virus encoded proteins, respectively. The IGR are typically organized into domain, DI-III (figure 13D); each connected by a pseudoknot structure (PK) and DI and II that are organized into a single structure for (Wilson et al., 2000, Jan et al., 2002, Pfingsten et al., 2007, Costantino et al., 2008; Hellen 2009). The IGR IRES of CrPV recruits the 40S ribosome directly without requiring any eIFs, positioning PKI at the A site (figure 13E, left) of the ribosome in a manner that blocks tRNA binding (Wilson et al., 2000, Jan et al., 2002, Pfingsten et al., 2007, Costantino et al., 2008; Hellen 2009; Petrov et al., 2016). The 60S subunits joins in the absence of eiF5B to form the 80S initiation complex with the DII positioned at the P-site of the ribosome, similar to codon-anticodon by a canonical tRNA. Additionally, IGR IRES can bind to the initiation competent 80S ribosome (Pestova et al., 2004; Petrov et al., 2016). There is a subsequent pseudotranslocation mediated by eukaryotic elongation factor 2 (eEF2) causing the movement of the PKI to the P-site (figure 13E, left) for the delivery of the aminoacyl-tRNA to the 80S competent elongation state at the A site (Kerr and Jan, 2016; Petrov et al., 2016). Overall, the CrPV can recruit both translation competent 40S and 80S without the involvement of eIFs to directly start elongation of translation.


A
Class I
(Polio Virus)
IRES
(643nt)
GNRA
eIF4G
neurovirulence
4A
Δ4G
C-rich
PCBP2
Viral replication
Cloverleaf (CL)
C-rich
C-rich
PCBP2
I
II
III
IV
V
VI
Yn
PTB
AUG
B
Class III
(EMCV IRES)
Apex
IRES
(439 nt)
GNRA
RAAA
long-range
interaction
hexaloop
C-rich
heptaloop
4A
Δ4G
eIF4G
(HEAT-1 domain)
III
stem loop
J
stem loop
K
A-loop
AUG-2
AUG-1
Yn
V
IV
II
I
Yn
PTB
C
Class III
(HCV IRES)
IRES
(298 nt)
III
IIIb
eIF3
II
E-site
uS7/uS7
uS7
IIb
uS11
IIIa
eS27
eS27
IIIc
40S
eS25
uS7/eS25
250
IIId
140
ES7
18S rRNA
18S rRNA-ES7
IIIe
I
IIa
eS1
IV
eS28
342
AUG
miR-122
PCBP2
D
Class IV
(CrPV IRES)
IRES
(188 nt)
eS25
uS7
SL/V (L2.1)
PKIII
6029
6120
SV
L2.2
L1.2
PKII
6170
uL5
Ribosome
binding
40S
L1.1
uL1
5S rRNA
(H76/H77)
uL11
L3.2
(VRL)
Ala
GCU
6217
PKI
Ribosome positioning
translocation
tRNA/codon
mimicking
ASL-like

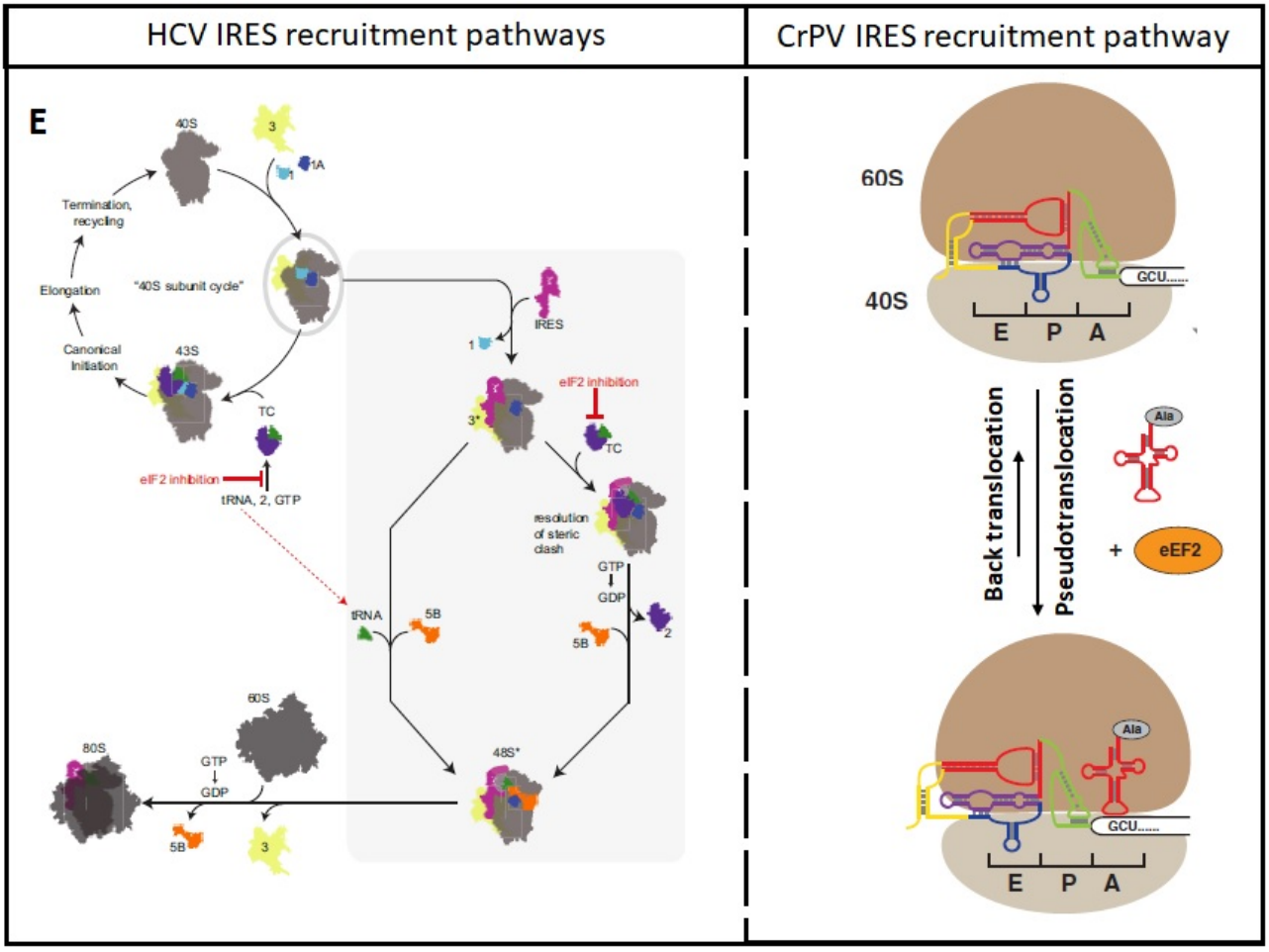


Figure 13. Structural features and mechanistic pathways of all four types of IRES. Respective secondary structures of the 5' UTR of Poliovirus (PV; class or type I IRES, A) encephalomyocarditis virus (EMCV; class II IRES, B), hepatitis C virus (HCV; class III IRES, C) and cricket paralysis virus (CrPV; class IV IRES, D), showing known features including known interacting eIFs (shown in green and violet) and ITAFs (shown in orange) and the 40S ribosome internal entry region within the structure (modified from Mailliot and Martin, 2017). (A) PV and (B) EMCV use similar mechanism in translational initiation using eIF4F complexes except the N-terminal of eIF4G scaffold protein and the cap-binding protein eIF4E and ITAFs including PCBP2 and PTB. Instead, HCV(C) and CrPV (D) IRESs directly recruit the 40S ribosome without the eIF4F complex, but HCV requires the activity of eIF3. (E) Current known mechanisms for HCV (right, modified from Jaafar et al., 2016) and CrPV-like (left, modified from Kerr and Jan, 2016) IRESs mediated cap-independent translation initiation**.** HCV-IRES relies on eIF3 to recruit the ternary complex (TC), but upon inhibition of the TC due to eIF2α phosphorylation in stress, it combine eIF3 with eIF1A and 5B to recruit the met-tRNAi. The CrPV model of mechanism is depicted showing the E, P, and A sites (E, left), the CrPV pseudoknot (PKI, in green) of domain III occupies the A site. For subsequent translocation to begin elongation, an amino-acyl-tRNA delivered by elongation factor 2 (eF2) assist in pseudotranslocation to the P-site, which can be reversed without activated tRNA delivery (shown by the forward and backward black arrows, long and short respectively).

2. AIMS OF THE STUDY

Gene expression in eukaryotes is a complex and tightly regulated process, involving several factors and different levels of regulations. Post-transcriptional regulation at the level of translation determines the fate of protein-coding transcripts (mRNAs), influencing proteins synthesis and therefore genes' activity. Most translational regulation occurs at the level of initiation, which often depends on the cellular conditions, and the availability of trans-acting RBPs and the ribosome, which are recruited by cis-acting elements that are located at RNAs 5' and 3' untranslated regions (UTRs). These usually determine which ribosomal complexes and translation factors are recruited to the mRNA, influencing the translation of the coding sequence (CDS) into protein in a cap-dependent and independent manner. Cap-dependent translation occurs under normal cellular conditions and relies on the 5'UTR m^7G -cap and the 3'UTR poly-A tail sequences to recruit the initiation complex for the CDS translation into protein. Upon the m^7G-cap inhibition during cellular stress, other cis-acting elements like TOPs, m^6A, uORFs and IRESs are uniquely involved in the recruitment of the translation initiation complex onto the 5'UTR to promote cap-independent translation.

In the laboratory of Prof. Gustincich, a family of antisense lncRNAs, named SINEUPs, was recently discovered for their ability to activate translation of target sense mRNAs (Carrieri et al., 2012, Zucchelli et al., 2015b). SINEUPs modular organization consist of the overlapping region called the BD that targets the 5'UTR of the sense mRNA through S/AS pairing and the ED, a downstream embedded invSINEB2 TE sequence, which confers translational activity. By taking advantage of this organization, synthetic SINEUPs could be design with novel BDs targeting in principle any mRNA of interest, making them versatile tools for several applications including molecular biology experiments, protein manufacturing and RNA therapeutics for treating diseases such as haploinsufficiencies.

In this thesis, I aim to:

1. study synthetic SINEUPs to reveal sequence features that could optimize SINEUP activity for therapeutic applications;
2. better understand the molecular mechanism of translational enhancement mediated by SINEUP ED. To this purpose I propose and test the SINEUP-IRES hypothesis that SINEUPs use IRES-like mechanisms to upregulate translation of target mRNAs.

3. MATERIALS AND METHODS

3.1 Oligonucleotides

The complete list of oligonucleotide/primer pairs used for all cloning, normal PCR and quantitative real-time PCR (RT-qPCR) experiments are included in table 1.

3.2 Plasmid Constructs

Synthetic SINEUP target-specific BDs were designed, in antisense orientation, around the ATG of protein-coding sequences with a -40/+4 (assigned short-S) and -40/+32 (assigned long-L) anatomies to endogenously expressed DJ1 and overexpressed peGFP CDS and then combined with Δ5'-AS Uchl1 sequence to create synthetic SINEUP-DJ1 and -GFP as previously described (Zucchelli et al., 2015). Δ5'-AS Uchl1 sequence, which was also called SINEUP-backbone sequence (Zucchelli et al., 2015) lacks the overlapping or BD region to Uchl1 mRNA (gene), but retains AS Uchl1 ED which consist of embedded inverted SINEB2 element, Alu element and 3' tail sequences.

Plasmid pcDNA 3.1 (-) expressing AS GFP construct, hereafter called synthetic SINEUP-GFP(L), where "L" signifies -40/+32 (Long) nucleotides BD targeting GFP mRNA was previously described (Carrieri et al., 2012, Zucchelli et al., 2015). Synthetic SINEUP-GFP(L) was also cloned into another plasmid, pCS2+ as previously described (Indrieri et al., 2016) and used in our studies. Synthetic SINEUP-DJ1(S) construct in pcDNA 3.1 (-), where "S" signifies -40/+4 (short) nucleotides BD targeting endogenous expressed DJ1 mRNA was also described previously (Zucchelli et al., 2015). Also, Synthetic miniSINEUP001-GFP construct targeting GFP mRNA is the miniSINEUP that was previously described (Zucchelli et al., 2015). The miniSINEUP003-GFP and miniSINEUP004-GFP constructs were all variants of miniSINEUP-GFP that were respectively gene synthesized by adding 30 poly-A tails to miniSINEUP001-GFP and a 38 MER sequence of miniSINEUP001-GFP into XbaI/HindIII RE sites in pcDNA3.1- plasmid. MiniSINEUP-Scr018 was generated with a scramble BD not targeting any known gene and cloned into XbaI/HindIII RE sites in pcDNA3.1- plasmid.

For the pCS2+ plasmid expressing SINEUP constructs; Synthetic SINEUP-GFP(S) was cut and cloned between XbaI/HindIII restriction enzyme (RE) sites in pCS2+ plasmid backbone, while Synthetic SINEUP-DJ1(S) was gene synthesized using Δ5'-AS Uchl1 sequence as SINEUP-backbone (Carrieri et

al., 2012, Zucchelli et al., 2015) and cloned into XhoI/XbaI RE sites in pCS2+ plasmid. The SINEUP-backbone sequence, herein called SINEUP-ED was cut and cloned from pCDNA3.1- into pCS2+, using EcoRI/HindiIII RE sites. SINEUP-DJ1_BDΔED, a synthetic SINEUP-DJ1 mutant consisting of DJ(S), BD, Alu and 3'tail sequences of AS Uchl1 but lacking invSINEB2 as main ED was two-step-wise PCR amplified from AS Uchl1 ΔinvSINEB2 sequence in pCDNA3.1- (Carrieri et al., 2012) as SINEUP-DNA template, using two different forward primers while maintaining the reverse primer. The first PCR step used a normal forward primer that annealed to the 5' end of the SINEUP-DNA template to amplified AS Uchl1 ΔinvSINEB2 PCR-product. The second PCR step used a forward primer containing the 44 nucleotide DJ1 (S) BD preceding XhoI RE and reverse primer preceding XbaI RE to produce AS Uchl1 ΔinvSINEB2 with DJ(S) BD called SINEUP-DJ1_BDΔED. The PCR product was then digested and annealed into XhoI/XbaI RE sites in pCS2+ expressing plasmid.

Viral and cellular IRES elements functioning as SINEUPs in SINEUP activity assay were performed by generating synthetic SINEUP-IRES constructs, using SINEUPs modular architecture in sense/antisense (S/AS) pairing rules and function (Carrieri et al., 2012). Target specific SINEUP-IRES to DJ(S) were gene synthesized using SINEUP-DJ1_BDΔED sequence as backbone and cloned into XhoI/XbaI RE sites in pCS2+ carrier plasmid as mentioned before. The various individual viral (HCV, Polio, EMCV and CRPV) and Cellular (c-myc, Apaf1, ELG1, DMD and Gtx) IRES sequences replace invSINEB2 as SINEUP main ED in synthetic SINEUP-IRES constructs. SINEUP_IRES-DJ1 clones have DJ1(S) BD, the respective IRES sequences in direct and inverted orientation, Alu and 3' tail sequences of AS Uchl1 SINEUP.

HCV-IRES mutations effect on targeted DJ1 protein synthesis was also investigated, using synthetic SINEUP-DJ1 constructs harbouring each HCV-IRES mutant as SINEUP main ED in pCS2+ plasmid. Synthetic SINEUP-HCV-IRES mutants M1-M12 were all generated by gene synthesis as was done for the HCV-IRES WT, using SINEUP-DJ1_BDΔED backbone and cloned between XhoI/XbaI RE sites in pCS2+ plasmid.

MiniSINEUP and miniSINEUP-IRES sequences targeting DJ1 mRNA were clone using respective forward and reverse primers as listed in supplementary table 1 to PCR amplify invSINEB2 and each respective IRES (c-myc, HCV and Polio-direct/inverted) EDs, and then cloned into pCS2+ and peGFP-DUAL plasmid backbones. For the PCR amplification, a single forward primer preceding XhoI RE site targeting DJ1(S) BD was used against respective reverse primers spanning individual IRES and

invSINEB2 sequences, all preceded by HindIII RE for target amplification. The respective full-length SINEUP_IRES-DJ1 and SINEUP-DJ(S) constructs were used PCR-DNA templates. The respective amplicons were then cloned into XhoI/HindIII RE sites in both pCS2+ and peGFP-DUAL plasmid backbones to obtain miniSINEUP-DJ1 and miniSINEUP-IRESs-DJ1 constructs. The dual peGFP-miniSINEUP-DJ1 and peGFP-miniSINEUP_IRES-DJ1 constructs expresses both miniSINEUP-DJ1 or miniSINEUP_IRES-DJ1 RNA and non-targeted GFP mRNA, but are all driven under a pol II H1 and a pol III CMV promoters respectively. Expressed GFP in this case was only used as a transfection control check.

For plasmid driven expression of miniSINEUP_HCV-IRES targeting GFP mRNA, a novel dual peGFP-miniSINEUP-HCV-IRES plasmid construct was generated. The HCV-IRES was PCR amplified, cut and paste into EcoRI and HindIII sites, using a peGFP-miniSINEUP-ΔinvSINEB2 plasmid backbone to form a dual peGFP-miniSINEUP-HCV-GFP plasmid construct. The novel dual peGFP-miniSINEUP-GFP(L) construct has GFP cDNA expressed from CMV, a pol II promoter that is in opposite orientation to a pol III H1 promoter, expressing miniSINEUP-GFP(L). It was obtained from the laboratory of Diego Cotello (University of Piemonte Orientale, Novara).

The ability of c-myc IRES to have SINEUP activity in *trans* was investigated using c-myc full length and mutants cDNAs. c-myc full-length and mutants (c-myc-DeltaC, c-myc 5' UTR and c-myc-IRES) cDNA constructs were all gene synthesis and cloned into EcoRI/HindIII RE sites in pcDNA expressing plasmid.

SINEUPs functioning as IRES in IRES-activity studies were performed by cloning SINEUP EDs, the invSINEB2 of various natural SINEUPs and the control (direct SINEB2) sequences between renilla luciferase (Rluc) and firefly luciferase (Fluc) into EcoRI/XhoI RE sites in empty pRUF plasmid. Specific primers habouring EcoRI/XhoI RE sites were annealed to invSINEB2 only or invSINEB2 with spacer sequences in AS UCHL1 DNA templates and PCR amplified, digested and ligated into the EcoRI/XhoI RE digested empty pRUF plasmid backbone. Similarly, target specific primers were used to amplified MIR1b, human FRAM and each invSINEB2 sequences of mouse AS Uxt and cloned into the EcoRI/XhoI RE digested empty pRUF plasmid backbone. c-myc 5'UTR in pRUF construct was used a s positive control as previously described (Bisio et al., 2010, 2015). The respective chimera mutants (Luc-IRES) of invSINEB2 ED of AS UCHL1 SINEUP and HCV-IRES were gene synthesized and cloned into the

EcoRI/XhoI RE digested empty pRUF plasmid backbone. Empty pRUF and c-myc 5'UTR pRUF control constructs were obtained from the laboratory of Prof. Alberto Inga, University of Trento.

All respective cloning primers are listed in supplementary table 1. All clones were confirmed by sequencing with respective plasmids sequencing primers. Empty plasmids harbouring just the multiple cloning sites (MCS) were used as negative controls to normalize transfection effects in both SINEUP and IRES activity experiments. SINEUP constructs containing invSINEB2 ED to any gene targets were used as positive control checks in SINEUP activity assay. Likewise, c-myc 5'UTR pRUF construct was used as positive control checks in IRES-activity assay. Targeted GFP construct, peGFP-C2 (clonetech), pcDNA 3.1- (Thermo Fisher Scientific) and pCS2+ (addgene) expressing plasmids were obtained commercially and previously used in Carrieri et al., 2012, Zucchelli et al., 2015 and Indrieri et al., 2016.

3.3 Cell Lines and Transfection Conditions

HEK 293T/17 cells were obtained from ATCC (Cat. No. ATCC-CRL-11268 164 293T/17) and maintained in culture with Dulbecco's Modified Eagle Medium6 (DMEM, GIBCO by Life Technologies, Cat. No. 41090-028) supplemented with 10% fetal bovine serum-FBS (Euroclone, Cat. No. ECS0180L) and 1% penicillin/streptomycin (p/s) antibiotics, as suggested by the vendor. HepG2 cells were kindly provided by Professor Collavin L. from the University of Trieste, Italy (Lunardi et al., 2009). HepG2 were cultured in Eagle's minimal essential medium (SIGMA) supplemented with 10% FBS (Euroclone, Cat. No. ECS0180L), 1% p/s antibiotics, 1% GlutaMAX and 1% non-essential aminoacids. HeLa cells and U2OS cells were grown in the same medium as HEK 293T cells.

The experiments in all four cell lines were studied solely with Fugene HD (Roche) transfection reagent, following manufacture's instruction. In detail, $3x10^5$ cells/well were seeded in 6-well plates the day before transfection at 60% confluence and transfected with SINEUP or SINEUP-IRES or dual-luciferase plasmids using FuGENE HD reagent in the ratio of 1:3 (w/v). For SINEUP and SINEUP-IRESs targeting endogenous DJ1-mRNAs, the encoding plasmids were transfected at a dosage previously described for SINEUPs (Zucchelli et al., 2015). Specifically, 2µg of synthetic SINEUPs and SINEUP-IRES in various expressing plasmids, and SINEB2, human FRAM and MIR1b and c-myc 5'UTR pRUF and empty plasmids were used for transfection in SINEUP- and IRES-activity assays respectively. For SINEUP activity assay, cells were collected at 48h post-transfection for all three cell types and split in two equal

portions; one-portion for quantitative RT-PCR and the other portion for Western Blot (WB) analysis. For dual peGFP-miniSINEUPs encoding plasmid experiments, 1µg each of dual peGFP-miniSINEUP-GFP, dual peGFP-miniSINEUP-HCV-GFP and the empty dual-peGFP plasmid plus 1µg of empty pCDNA3.1- plasmid were used to maintain the conditions of constructs and FuGENE HD transfection ratio 1:3 (w/v). For the measure of SINEUP activity, quantification of RNA by qPCR and protein levels by Western Blot (WB) were obtained from the same transfection in each replica. For IRES activity assay, the same transfection conditions mentioned above were used, except that all collected cells were used either for the luciferase analysis or for the control RNA experiments; RT-PCR followed by two separate primer specific PCR and qPCR analysis.

3.4 Dual Luciferase Assays in Cell Lines

Specifically, 2µg of SINEB2 (inverted and direct sequences) or all other SINEUP EDs in pRUF, c-myc-IRES-pRUF and empty pRUF plasmids were used in the transfection. Empty and c-myc-IRES pRUF plasmids were kindly provided by Professor Alberto Inga from University of Trento, Italy. Cells were harvested 48h post-transfection and luciferase assay performed using the dual-luciferase reagent kit (Promega), according to the manufacturer's instructions. All cell lines lysis were performed with 1x Passive Lysis Buffer (PLB) and then cells were subjected to freeze-thaw cycle at -80^{o} C C to 25^{o} C. Cell lysates were then centrifuged at maximum speed for 30sec and used to measure Fluc and Rluc luminescence activities respectively. Luminescence was measured using a 96 Multilabel Plate luminometer (PerkinElmer) according the manufacturer's standard protocol. The empty pRUF reaction was used to normalize the transfection efficiency or background effects as previously described (Bisio et al., 2010, 2015). IRES activity was calculated as ratio of firefly luciferase (Fluc) activity relative to renilla luciferase (Rluc) activity as normalized against the empty pRUF vector transfection. Each luciferase activity was done in at least four independent duplicate-replicates and measured in relative light units (RLU). IRES activity control experiments were done to eliminate any possibility of cryptic promoter or alternative splicing events. Briefly, RNA was purified from harvested transfected cells and used for RT-PCR, followed by two separate PCR, and qPCR analysis. All measurements were done in duplicate transfections of at least four independent biological replicas.

3.5 RNA Isolation, Reverse Transcription and Quantitative real time-PCR (qRT-PCR)

Total RNA was extracted from cells with RNeasy Mini kit (Qiagen, #74106) according to the manufacturer's protocol. All RNA isolations were subjected to double DNase digestion; on-column DNase I treatment (Qiagen) according to manufacturer's protocol before RNA quantification, followed by in solution treatment with TURBO DNA-free Kit (Ambion) at 37^{o} C for 1 hour after the first RNA quantification to clean up all DNA contaminations. RNA quality was checked on formaldehyde agarose gel. A total of 1µg of RNA was reverse transcribed using the iSCRIPT cDNA Synthesis Kit (Bio-Rad) according to the manufacturer's protocol. qPCR reaction was performed on diluted RT-PCR products (1:20) using SYBR-Green PCR Master Mix (BioRad) and an iCycler IQ Real time PCR System (Bio-Rad). All relative expression was calculated by the 2^ΔΔCt method using GAPDH as normalizing fold control. Oligonucleotide sequences of primers used in this study for GFP and GAPDH (Carrieri et al., 2012), DJ1 (Foti et al., 2010) were previously described. All SINEUP-IRES-DJ1 and SINEUP-DJ1 RNA expressions were detected with the same primers designed on the 3' end of AS Uchl1 (mAS Uchl1 FWD and REV primers, Carrieri et al., 2012). MiniSINEUP-GFP and MiniSINEUP-DJ1 were also detected using the primers described in (Zucchelli et al., 2015), while miniSINEUP-IRES-GFP and MiniSINEUP-IRES-DJ1 primers are shown in the Supplementary Table 2.

For the IRES activity control experiments; total RNA was purified for RT-PCR as mentioned above, and then two separate PCR was done with a forward primer in the transcription start site (TSS) of the bicistronic Rluc-Fluc mRNA and two reverse primers Rluc and Fluc -rev to eliminate any possibility of alternative splicing, and qPCR measurement was performed to check for the possibility of cryptic promoter activities from the SINEB2 inserted sequences between renilla and firefly sequences. All PCR primers for IRES activity assays are previously described (Bisio et al., 2010, 2015).

3.6 SDS-PAGE and Western Blot Analysis

Cells were pelleted for 5min at 6000 rpm and then lysed in 2X SDS (Laemmli) sample buffer. Proteins were separated in 12% SDS– polyacrylamide gel and transferred to nitrocellulose membranes (Amersham™, Cat. No. GEH10600001) at 100V per 90min. All membranes were initially blocked with 5% (w/v) milk in TBST solution. For synthetic SINEUPs target proteins; DJ1 and GFP, immunoblotting was performed with the following primary antibodies: 1:500 diluted mouse monoclonal anti-DJ1 (ADI-KAM-SA100, Enzo Life Sciences), 1:800 diluted mouse monoclonal anti-GFP (632380, Clontech), and 1:25000 diluted anti-β-Actin peroxidase (A3854, SIGMA). For c-myc 5'UTR targeted genes, the immunoblotting conditions are listed in the appendix table. Signals were revealed after incubation with recommended secondary antibodies conjugated with horseradish peroxidase (Daco) by using enhanced chemiluminescence (ECL) detection reagent (RPN2105, GE Healthcare). Image detection was performed with Alliance LD2-77WL system (Uvitec, Cambridge) and quantification by using ImageJ Open Source software (http://imagej.net/), according to the program's manual. The proteins (DJ1 and GFP) image intensities were normalized to the B-Actin levels, used as loading control check.

3.7 Bioinformatic Analysis SINEUP EDs

IRESsite (Mokrejs et al., 2006 http://iresite.org/) search tool was used to check for sequence and structural motifs similarities between invSINEB2 of all reported natural SINEUPs from mouse genome (Carrieri et al., 2012) and all annotated experimentally proven viral and cellular IRES sequences in the IRESsite database. IRES sequences and structural motifs with good score, especially with invSINB2 of AS Uchl1 were selected for our synthetic SINEUP-IRES constructs design and experiments.

Web-server tool RegRNA 2.0 (Huang et al., 2006; http://regrna2.mbc.nctu.edu.tw/) was further used to analyze the sequence of mouse SINEUPs; AS Uchl1 and AS Uxt and human SINEUP (R12A-AS1 and ITFG2-AS1) for the presence of any known cis-regulatory elements; upstream open reading frames (uORF), IRES elements, terminal oligopyrimidine tract elements (TOPs) that mediates their function.

Another web-server tool SRAMP (Zhou et al., 2016; **http://www.cuilab.cn/sramp/)** was used to predict N^6-methyladenosine (m^6A) motifs in SINEUP sequences mouse SINEUPs; AS Uchl1 and AS Uxt and human SINEUP (R12A-AS1 and ITFG2-AS1) as potential cis-regulatory elements driving functionality in SINEUPs.

3.8 Statistical Analysis

All data were expressed as mean ± standard deviation on at least $n \geq 4$ replicas. Statistical analysis was performed by using Microsoft-Excel software. Statistically significant differences were assessed by two-tailed unpaired Student's *t*.test. Differences with $p < 0.05$ were considered significant.

4. RESULTS

4.1 Synthetic SINEUPs activity is dependent on both the embedded invSINEB2 ED and the overlapping BD targeting DJ1 mRNA in HEK 293T cells

As previously discovered, natural SINEUP of AS Uchl1 increase Uchl1 protein synthesis through an embedded inverted SINEB2 (invSINEB2) TE containing ED sequence and an overlapping BD sequence, which is antisense at the 5'UTR of Uchl1 (Carrieri et al., 2012). Therefore, deletion of both domains abolishes natural AS Uchl1 SINEUP ability to increase Uchl1 mRNA translation. This was confirmed by subsequent studies using synthetic miniSINEUP-GFP having just invSINEB2 as SINEUPs main ED and BD targeting GFPO mRNA (Zucchelli *et al*., 2015). Here, I studied the effects of these domains in synthetic full-length SINEUPs targeting endogenously expressed DJ1 gene by deleting either the SINEUP main ED, the invSINEB2 or BD one at a time, similar to what was previously done for natural AS Uchl1 on Uchl1 protein synthesis (Carrieri et al., 2012). I generated synthetic SINEUP-DJ1_BDΔED lacking the invSINEB2 ED, but retaining the overlapping BD to DJ1 mRNA, while synthetic SINEUP-ED had no BD and therefore they were not antisense to any specified mRNA (**Figure *14***). A SINEUP-DJ1 construct with complete invSINEB2 ED and BD targeting DJ1, was previously reported to have SINEUP activity on endogenous DJ1 protein synthesis when expressed from pcDNA3.1- (Zucchelli et al., 2015a). Here, the very same SINEUP-DJ1 was cloned into pCS2+ vector and used as a positive control, while empty pCS2+ vector was a negative control. By transfecting HEK 293T cells with these constructs, SINEUP-DJ1 activity was confirmed in increasing DJ1 protein levels of 1.6 folds on average, while no activity was observed when ED and BD sequences were deleted (**Figure *14***). Thus, overexpressed synthetic SINEUPs of both SINEUP-DJ1_BDΔED and SINEUP-ED have no effects on DJ1 protein synthesis. These data further confirmed and validated the importance of the invSINEB2 TE and BD sequences in driving SINEUP activity of both natural and synthetic SINEUPs.

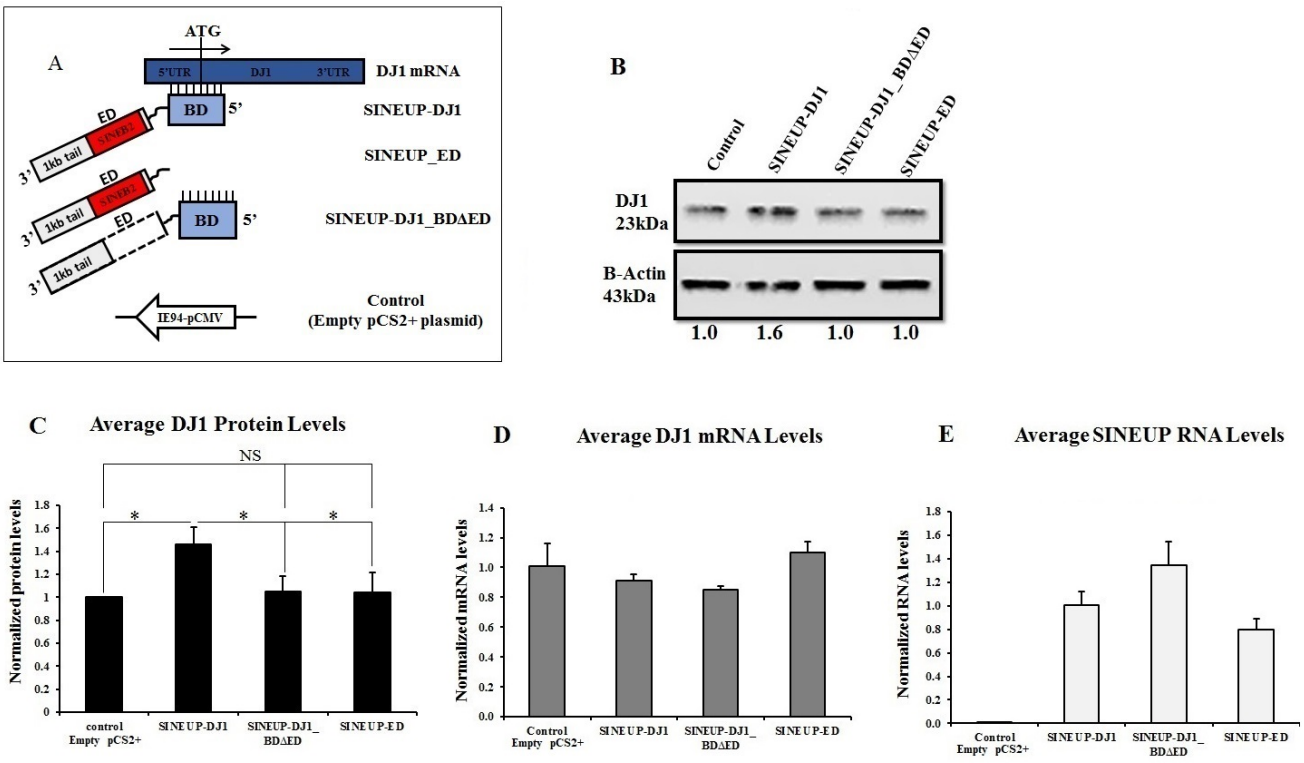


Figure 14 Synthetic full-length SINEUP-DJ1 increase DJ1 protein synthesis using an embedded invSINEB2 element in HEK 293T cells. (A) Scheme of synthetic SINEUP-DJ1 and mutants, showing full-length synthetic SINEUP with binding domain BD (light blue) targeting DJ1 mRNA (Blue) 5' UTR around the AUG. SINEUP main effector domain (ED), invSINEB2 element is shown in red within 1kb AS Uchl1 3' tail sequence. SINEUP-DJ1 mutants; SINEUP-DJ1_BDΔED that is Full-length synthetic SINEUP-DJ1 with deleted invSINEB2 sequence (white dotted box) and SINEUP-ED, which is Δ5'-AS Uchl1 sequence without any predicted BD are depicted. (B) Western blotting of DJ1 protein in HEK 293T cells was detected 48hrs post-transfection with the indicated Full-length synthetic SINEUP-DJ1 and mutant constructs expressed from pCS2+ plasmid as normalized against empty pCS2+ plasmid control. DJ1 protein detection was done with monoclonal anti-DJ1 antibody normalized to β-Actin that was used as loading control. (C) Average DJ1 protein fold change quantification plot (black bars) and comparison among the indicated constructs is shown . Expression of DJ1 mRNA (grey bar plot, D) and SINEUP constructs RNA (white bar plot, E) were monitored by qRT-PCR with respective specific primers. All plots indicate mean ± standard deviation, and representation of N=5 independent replicates $^{*}p < 0.05$, NS; not significant.

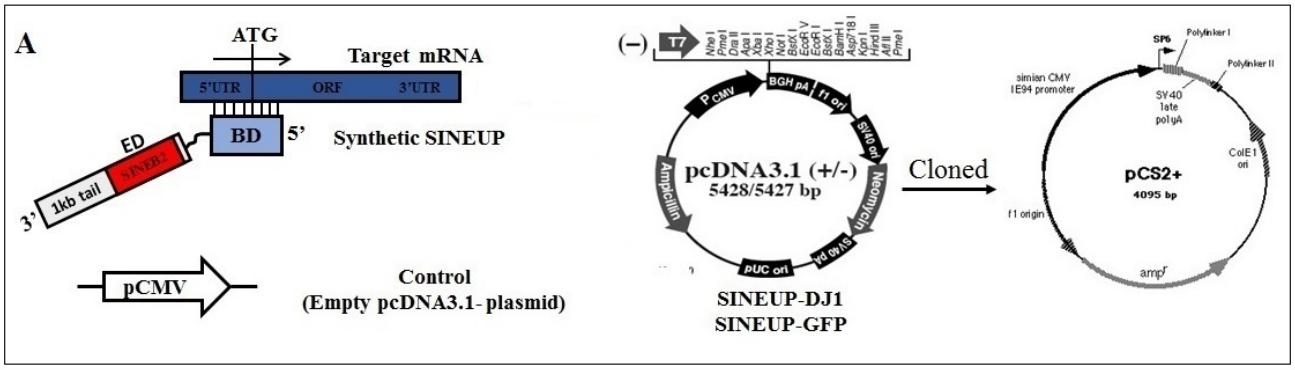
4.2 Optimization of Synthetic SINEUPs for applications

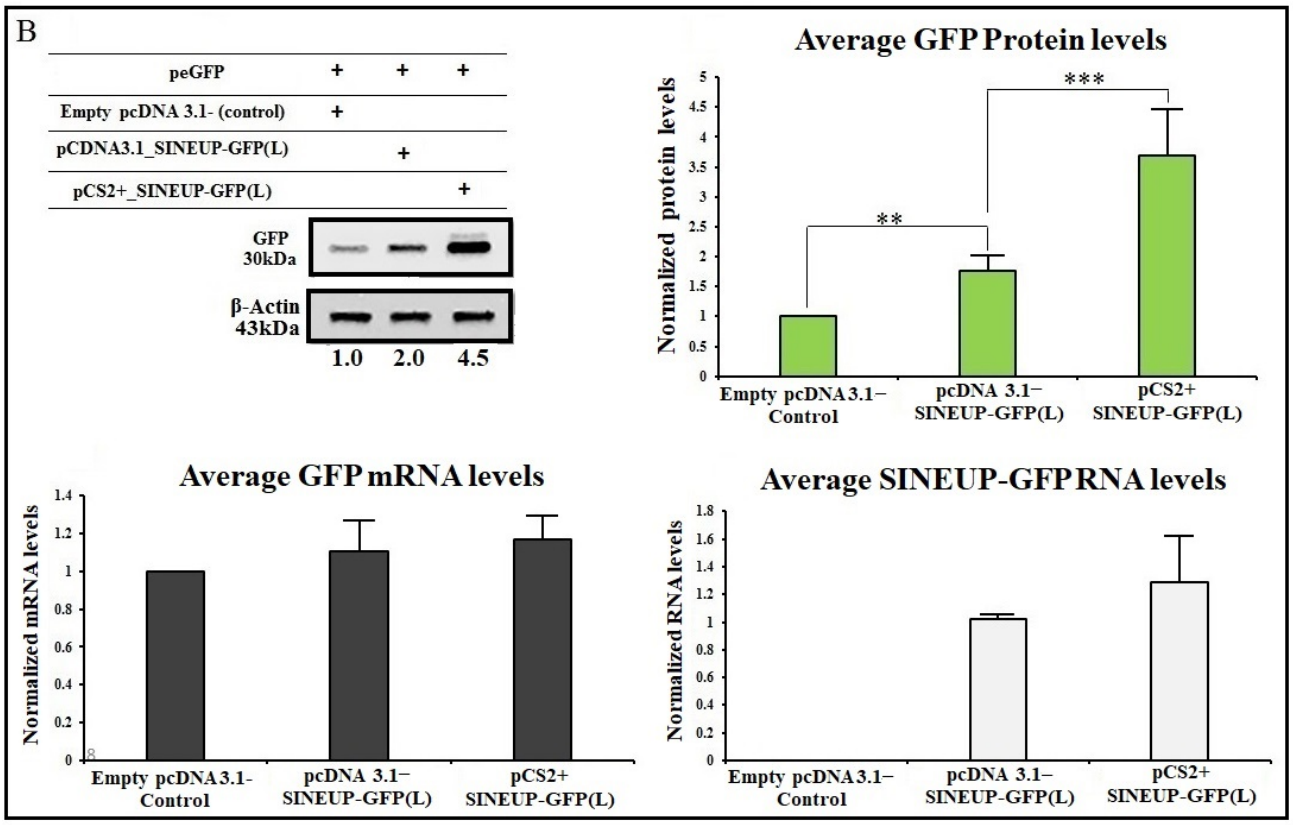
Synthetic SINEUPs ability to increase any targeted protein synthesis makes them useful as promising tools for applications in proteins manufacturing and RNA therapeutics for treating several diseases, including neuropathogenic haploinsufficiencies. Recently, miRNA and siRNA technologies have

become valuable molecular reagents used by both basic and applied researchers in inhibiting or knocking down gene expression in mammalian, fish, flies and other eukaryotic systems. However, there is currently no available counterpart technologies to activate or knock-up desired target-specific protein levels. Therefore, I aimed here to improve chances of delivery *in vivo* by creating the smallest functioning synthetic SINEUP within the small ncRNA length range and improve their expression.

4.2.1 Synthetic SINEUPs maintain SINEUP activity when expressed from different plasmid backbones

Previous works studied synthetic SINEUPs targeting overexpressed and endogenously express genes from pcDNA3.1- (Carrieri et al., 2012; Zucchelli et al., 2015a). Here, I tested whether synthetic SINEUPs could function when expressed from other plasmid backbones. I cloned synthetic SINEUPs-DJ1 and -GFP from pcDNA3.1- into pCS2+ plasmid and showed that they maintain their ability to increase endogenous DJ1 and overexpressed GFP proteins synthesis, respectively in HEK 293T cell lines *in vitro* (**Figure 15**). It must be noticed that each synthetic SINEUPs functions better when expressed in pCS2+ plasmid backbone as compared to pcDNA3.1- in HEK 293T cells *in vitro* (**Figure 15**B and C), suggesting that differences in plasmid features may influence SINEUP activity (**Figure 15**A).




B
peGFP + + +
Empty pcDNA 3.1- (control) +
pCDNA3.1_SINEUP-GFP(L) +
pCS2+_SINEUP-GFP(L) +
GFP 30kDa
β-Actin 43kDa
1.0 2.0 4.5
Average GFP Protein levels
Normalized protein levels
***
**
Empty pcDNA3.1– Control
pcDNA 3.1– SINEUP-GFP(L)
pCS2+ SINEUP-GFP(L)
Average GFP mRNA levels
Normalized mRNA levels
Empty pcDNA3.1- Control
pcDNA 3.1– SINEUP-GFP(L)
pCS2+ SINEUP-GFP(L)
Average SINEUP-GFP RNA levels
Normalized RNA levels
Empty pcDNA3.1– Control
pcDNA 3.1– SINEUP-GFP(L)
pCS2+ SINEUP-GFP(L)

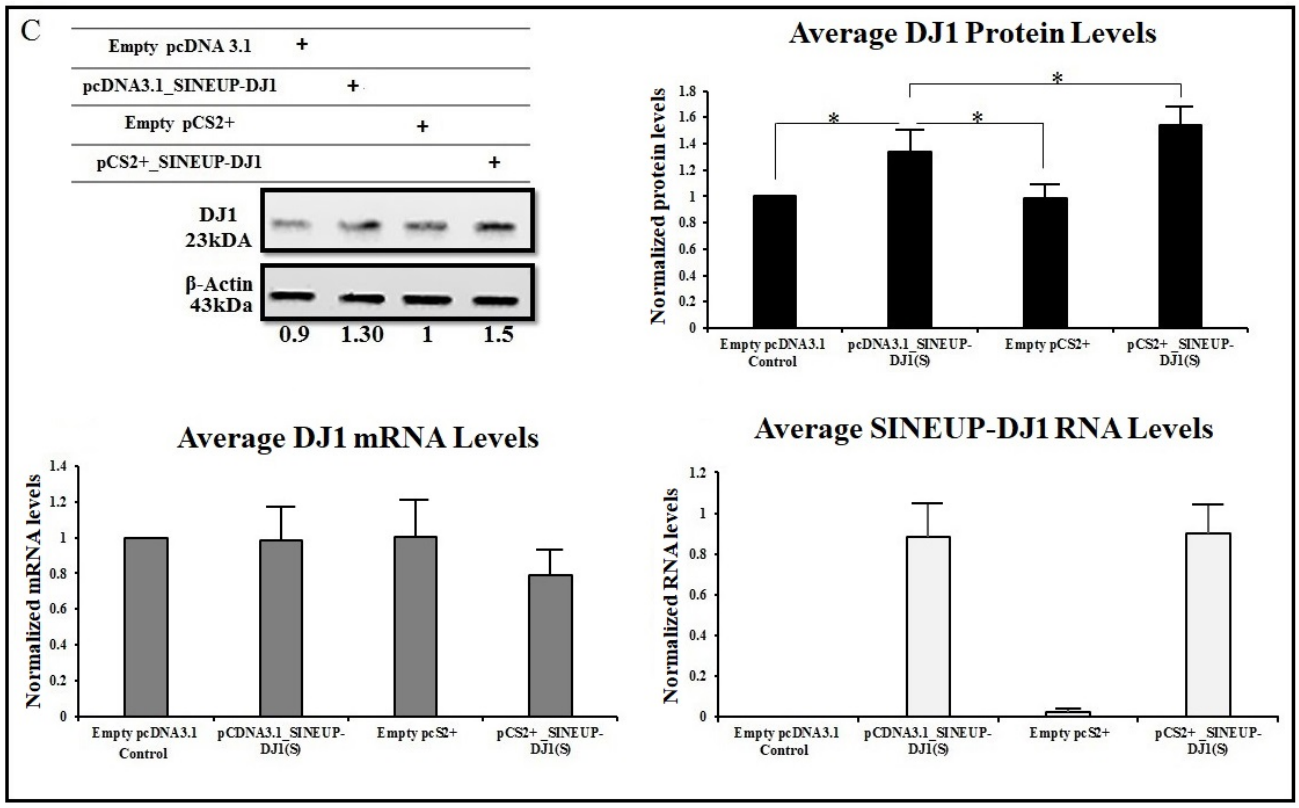


Figure 15 Synthetic SINEUPs function better when expressed from pCS2+ than pCDNA3.1-plasmid. (A) Scheme of synthetic Full-length SINEUP targeting endogenously expressed or overexpressed mRNAs (left). Full-length synthetic SINEUP-DJ1 and -GFP were cloned into pCS2+ plasmid (adapted from google images, right). Plasmid pcDNA3.1- and pCS2+ features are depicted, showing differences in plasmid size, promoter types and poly-A tail signals. Synthetic SINEUP activity in HEK 293T cells on overexpressed GFP (B) and endogenously expressed DJ1 (C) proteins were compared between the respective SINEUPs expressed from pcDNA3.1- and pCS2+ plasmids. SINEUP activity controls were checked with protein quantification from empty pcDNA3.1- backbone transfected cells. Quantified GFP (green bar plot, B) and DJ1 (dark bar plot, C) proteins levels were detected by Western blot with anti-GFP and -DJ1 antibody, respectively. GFP and DJ1 mRNA (grey bar plots, lower panel left) and SINEUP-GFP and DJ1 RNA (white bar plot, lower panel right) were quantified by qRT-PCR with respective DJ1 and GFP primer pairs and SINEUP detection primer pairs, respectively. SINEUP activity was calculated as the increase in protein levels from cells transfected with plasmids expressing synthetic SINEUPs as compared to empty pcDNA3.1- transfected controls. All estimated proteins were normalized against β-actin protein that was used as loading control. SINEUP activities were compared between the same synthetic SINEUP expressions from both plasmids. In all transfected samples, SINEUPs were expressed and the targeted DJ1 and GFP mRNAs were all stable statistically, $p>0.05$. All plots indicate mean ± standard deviation, and representation of at least N=5 independent replicates $^{*}p<0.05$, $^{**}p<0.01$, $^{***}p<0.001$.

4.2.2 Synthetic SINEUP BD length at the 5'UTR influence SINEUPs function in HEK 293T cells *in vitro*

The work of Zucchelli et al., 2015a, has established that different BD lengths; -40/+4 (S) and -40/+32 (L) respectively called BD short and long, targeting endogenously expressed DJ1 mRNA around the ATG are both functional. Here, I investigated these two different BD lengths on synthetic SINEUPs targeting overexpressed GFP mRNA. I cloned synthetic SINEUP-GFP by PCR with -40/+4 (S) and -40/+32 (L) identifying synthetic SINEUP-GFP(S) and SINEUP-GFP(L) constructs respectively (Figure *16*A). Both constructs were expressed from pCS2+ plasmid backbone. Transfection in HEK 293T cells showed increased in GFP protein synthesis (*Figure 16*B and C). Mean fold induction showed that SINEUP-GFP(S) works better than SINEUP-GFP(L), and this difference in SINEUP activity was statistically significant with $p < 0.05$ (*Figure 16*C). Here again, qRT-PCR showed statistically stable expressed GFP-mRNA (*Figure 16*D), indicating post-transcriptional increase in GFP protein synthesis. Both SINEUP-GFP -(S) and -(L) could be qPCR detected with SINEUP primers (*Figure 16*C).

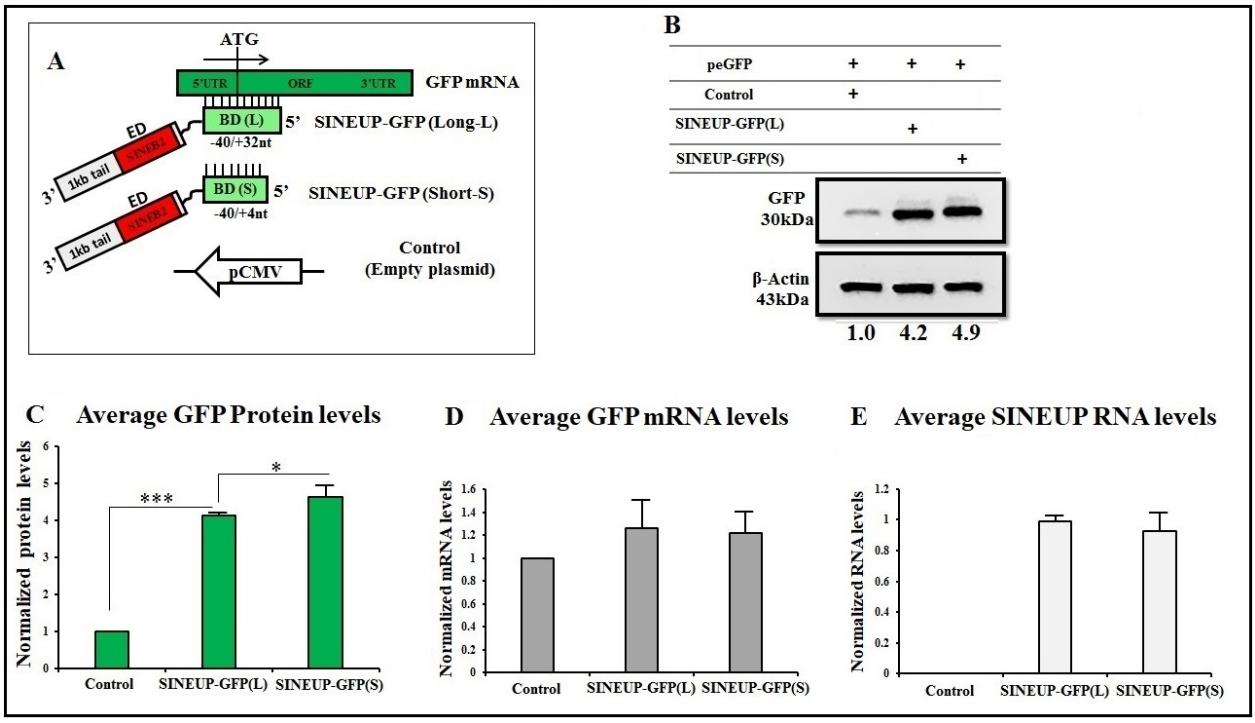
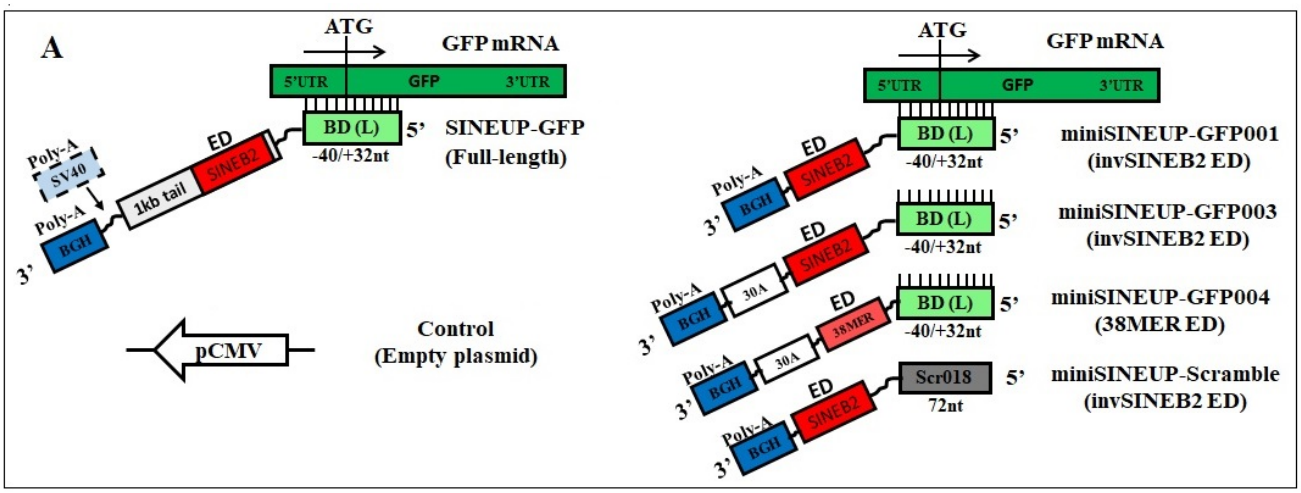


Figure 16 Synthetic SINEUP-GFP with shorter BD has better SINEUP activity than a longer version. Schematic representation of synthetic full-length SINEUP-GFP with longer (-40/+32) and shorter (-40/+4) BDs targeting overexpressed GFP mRNA. (B-E) HEK 293T cells were co-transfected with peGFP-C2 and SINEUP-GFP (S) or SINEUP-GFP (L) or control empty plasmid. Cells were harvested 48hours post-transfection and processed for GFP protein quantification (B and C) with anti-GFP antibody and RNA quantifications by qRT-PCR detects GFP mRNA (D) and SINEUP-GFP RNA (E) with respective primers. GFP fold induction shown on the Western blot image (B) was quantified relative to the empty plasmid control (C). Expressed GFP mRNA was invariant ($p<0.05$). All plots indicate mean ± standard deviation, and representation of at least N=5 independent replicates, $^{*}p<0.05$, $^{***}p<0.001$

4.2.3 3'-end poly-A tail influences synthetic full-length- and mini-SINEUPs activity

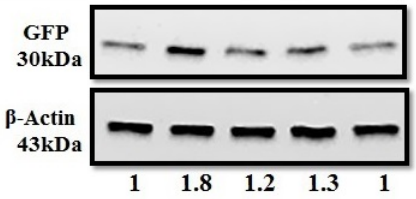
It is still unclear whether both natural and synthetic SINEUPs require poly-A tails for mediating SINEUP activity. Here, I aimed to establish the influence of poly-A sequences or signals on synthetic SINEUP activity. To this purpose, I first investigated the effects of the addition of 30 poly-A sequences downstream miniSINEUP-GFP and of a 38-MER sequence region within the invSINEB2 ED SL1 hairpin structure of AS Uchl1 (Podbevšek et al., 2018) in pcDNA3,1-. This plasmid backbone contains a BGH poly-A signal. The added 30 poly-A sequences to miniSINEUP-GFP and the 38-MER miniSINEUP-GFP respectively generated miniSINEUP-003 and -004 constructs, targeting overexpressed GFP-mRNA ***(Figure 17A, right).*** Interestingly, a 50% reduction in SINEUP activity was observed in miniSINEUP-003 and -004 constructs (**Figure *17***).

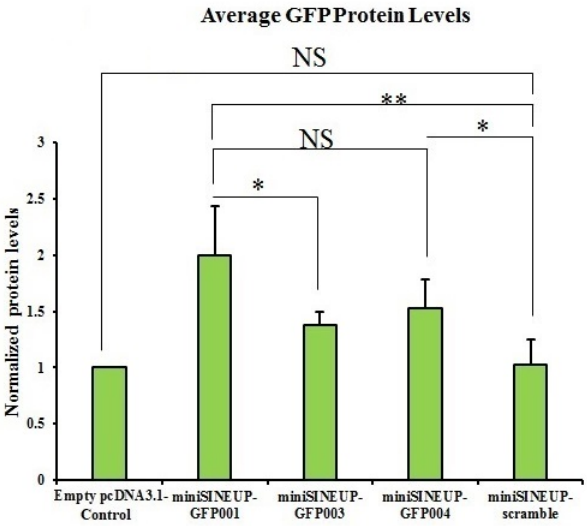
As previously reported above, synthetic SINEUPs targeting DJ1 and GFP mRNAs have better function when expressed in pCS2+ than pcDNA3.1- ***(Figure 15B***). Both pcDNA3.1 and pCS2+ vectors have RNA pol II promoters, but they differ for their, respectively, BGH and SV40 poly-A signals . I therefore reasoned that the differences in activity between vectors could be due to differences in poly-A tail lengths or strength with SV40 poly-A signal having better influence in SINEUPs activity. I tested this hypothesis by cloning a direct and a flipped SV40 poly-A site downstream SINEUP-GFP in pcDNA3.1- plasmid (**Figure *17***A, left). Indeed, when a direct but not the flipped SV40 poly-A site was present, an increase in SINEUP activity was observed (**Figure *17***C, upper panel). In the future I plan to study the effects of deleting the BGH poly-A site downstream the SINEUP-GFP/SV40 poly-A site in pcDNA3.1- ***(*Figure *17A, left)***. Furthermore, I plan to carry out a 3'RACE experiment on natural SINEUPs to identify their poly-A site usage and describe their poly-A tails.

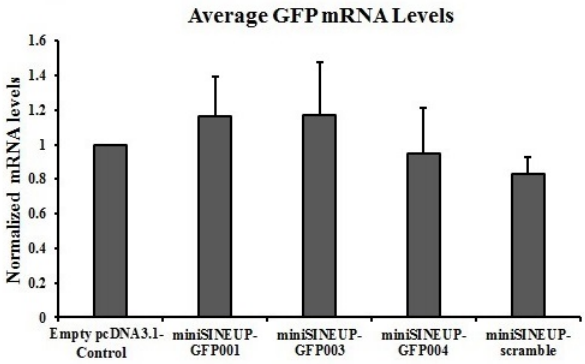


B

pEGFP	+	+	+	+	+
Empty pcDNA 3.1-	+				
MiniSINEUP-GFP001		+			
MiniSINEUP-GFP003			+		
MiniSINEUP-GFP004				+	
MiniSINEUP-Scramble					+







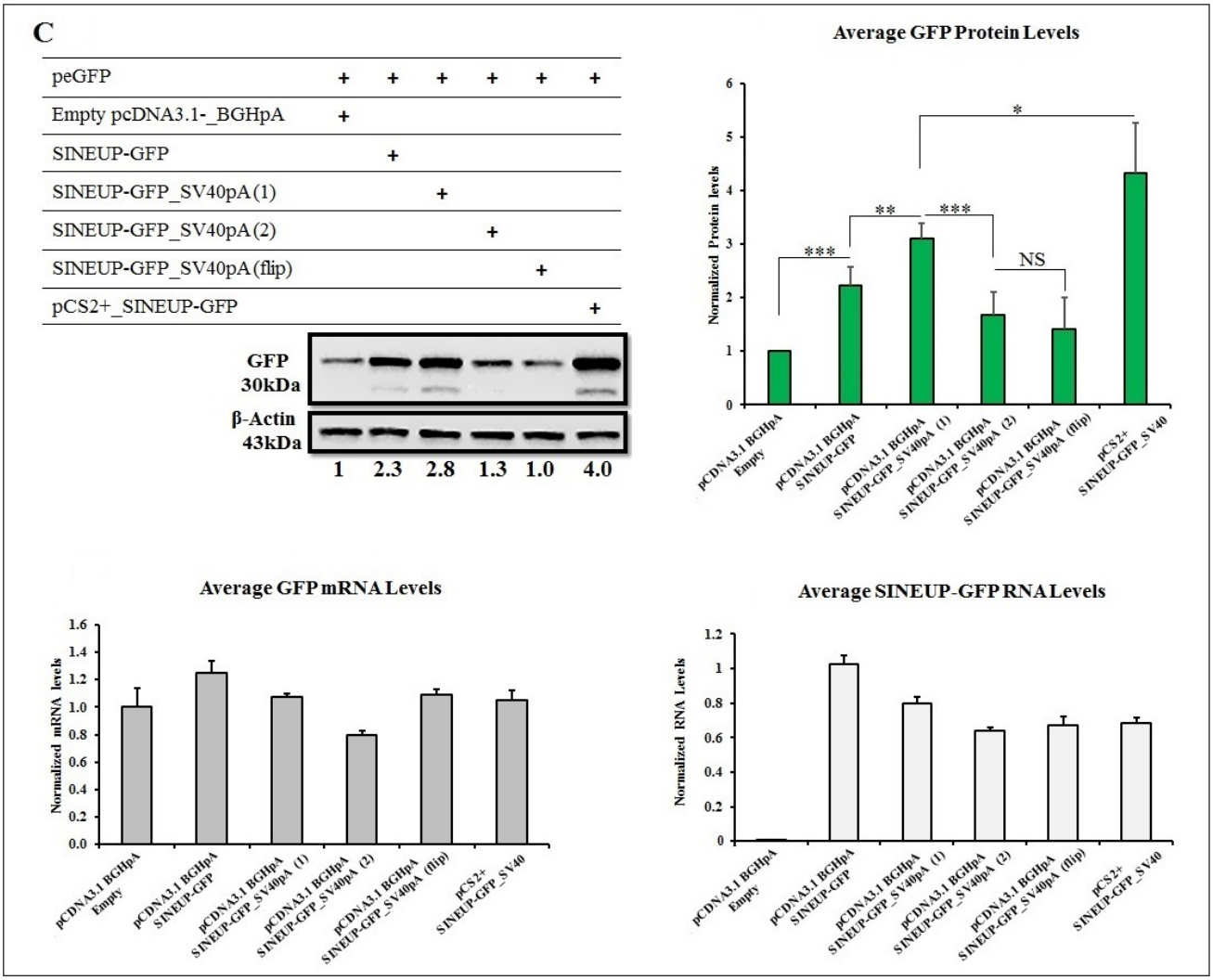
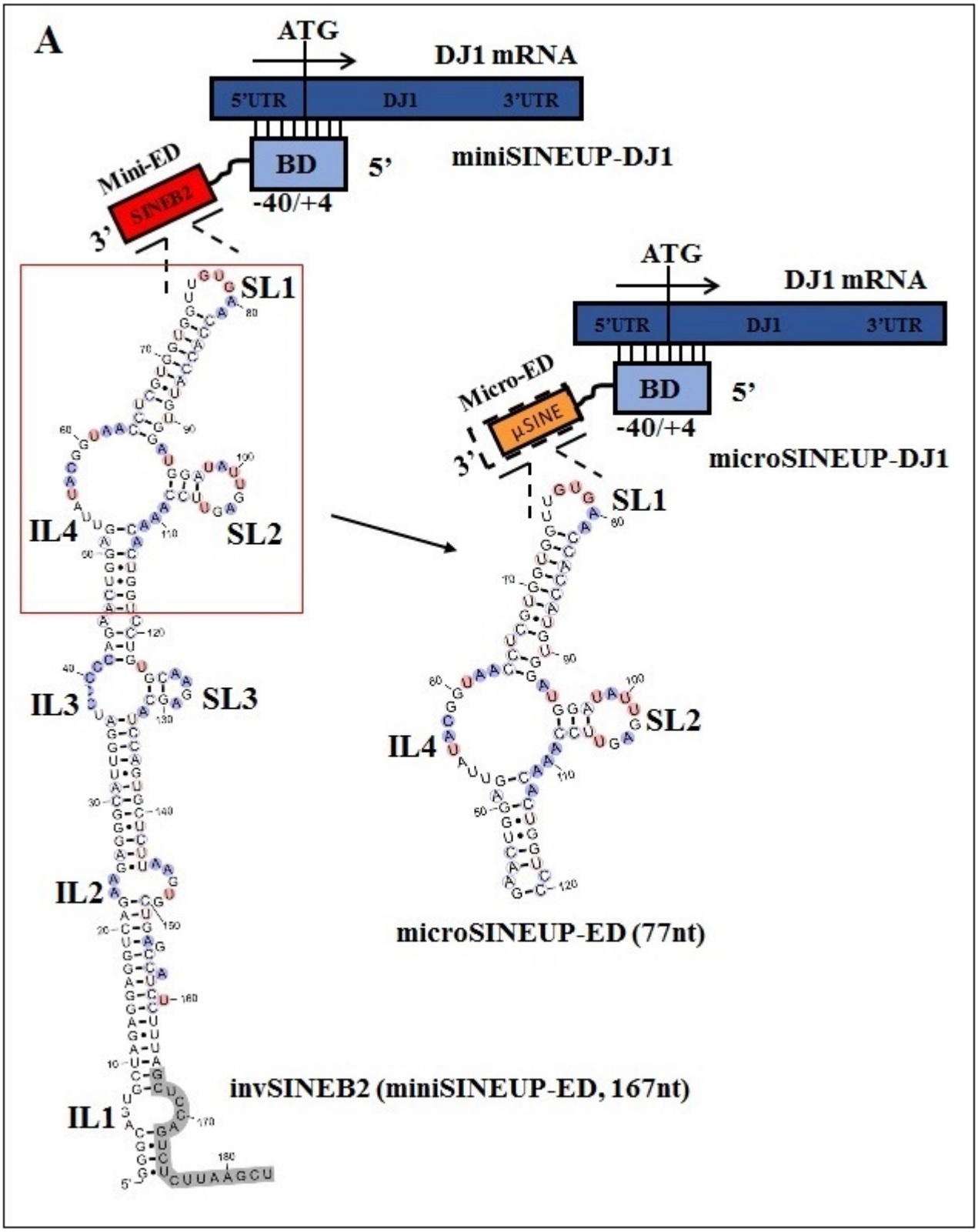


Figure 17 Downstream polyA-tail sequences influence synthetic SINEUP activity in HEK cells. (A) Scheme of synthetic SINEUPs targeting overexpressed GFP mRNAs (left). SV40 poly-A signal from pCS2+ plasmid was cloned under full-length synthetic SINEUP-GFP in pcDNA3.1- plasmid. The same full-length synthetic SINEUP-GFP in pcDNA3.1- was used as a positive control. 30 poly-A sequences were added to miniSINEUP-GFP001, 003 and 004 sequence cloned to pcDNA3.1-. (bottom). Synthetic miniSINEUP-scramble with BD sequence not targeting GFP mRNA (bottom) was used as additional negative control in the miniSINEUP-GFP studies, and empty pcDNA3.1- (bottom) was used as a control in both studies. (B and C) HEK 293T cells were co-transfected with peGFP-C2 and synthetic full-length SINEUP-GFP or miniSINEUP constructs or control empty plasmid. Cells were harvested 48hrs post-transfection and processed for GFP protein quantification with anti-GFP antibody (top) and RNA quantifications, GFP mRNA and SINEUP-GFP RNA (bottom) with respective primers. GFP fold induction shown on the Western blot images (top, right) were quantified relative to the empty plasmid control and the fold change average levels compared (top, right). Expressed GFP mRNA was stable ($p<0.05$). All plots indicate mean ± standard deviation, and representation of at least N=5 independent replicates, $^{*}p<0.05$, $^{**}p<0.01$, $^{***}p<0.001$.

4.2.4 MicroSINEUPs exclusively contain BD and a portion of the invSINEB2 sequence as ED

Synthetic miniSINEUP-GFP consisting of 250nt long transcript was previously shown to be active in increasing GFP protein synthesis (Zucchelli et al., 2015a). I then successfully targeted endogenously expressed DJ1 mRNA with miniSINEUP-DJ1. MiniSINEUPs are exclusively composed by a BD and the 167nt of the AS Uchl1 invSINEB2 as ED. Subsequently, I took advantage of the NMR-based structure of the invSINEB2 sequence that identified the SL1 hairpin of AS Uchl1 as the structural determinant mediating SINEUP function. By mutational analysis, I discovered that a 76nt SL1 hairpin-containing sequences of the invSINEB2 **(**Figure 18***A)*** was the shortest active ED. When it was combined with a BD, it was is able to increase protein synthesis of targeted mRNAs as a natural, long SINEUP. These RNA molecules were called microSINEUPs. Here, I synthesized microSINEUP-DJ1 by combining the 44nt BD targeting endogenous DJ1 mRNA with the identified 76nt ED sequence from invSINEB2, separated by a 19nt linker to create a functional synthetic microSINEUP-DJ1 of 140nt sequence (**Figure *18***A, right).

When transfected in HEK 293T cells, just like the miniSINEUP_DJ1 used as positive control, the newly synthesized microSINEUP-DJ1 stimulated a 1.7 fold increase of DJ1 protein levels without any effect on mRNA transcription levels ***(Figure 18B-D).*** Interestingly, the SINEUP activity of synthetic microSINEUP-DJ1 was comparable to the one of miniSINEUP-DJ1 ***(Figure 18B),*** indicating that microSINEUP ED retains all the activatory functions.


A
ATG
DJ1 mRNA
5'UTR
DJ1
3'UTR
BD
5'
-40/+4
miniSINEUP-DJ1
Mini-ED
SINEB2
3'
SL1
IL4
SL2
IL3
SL3
IL2
IL1
invSINEB2 (miniSINEUP-ED, 167nt)
ATG
DJ1 mRNA
5'UTR
DJ1
3'UTR
Micro-ED
μSINE
BD
5'
-40/+4
3'
microSINEUP-DJ1
SL1
IL4
SL2
microSINEUP-ED (77nt)

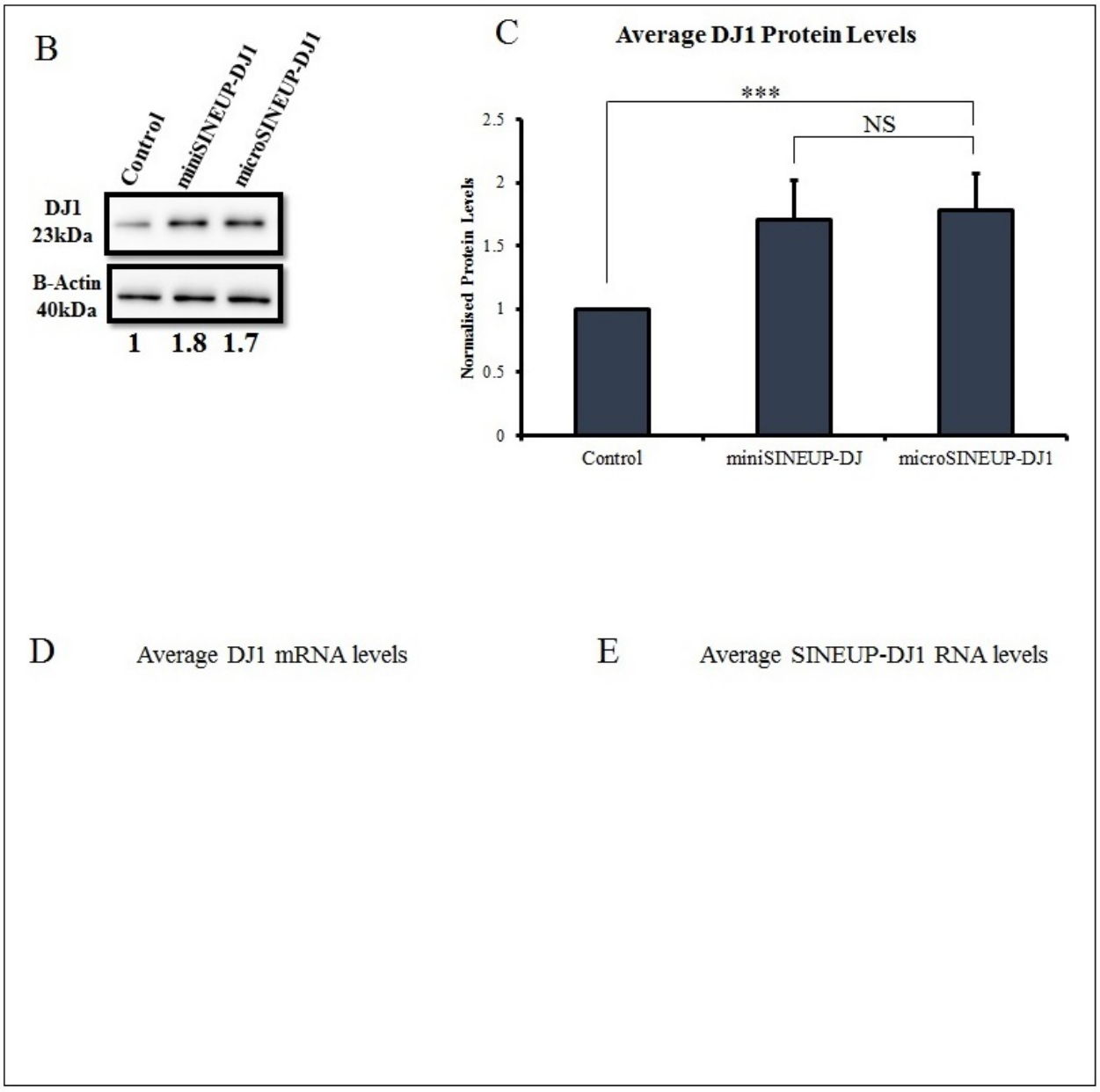
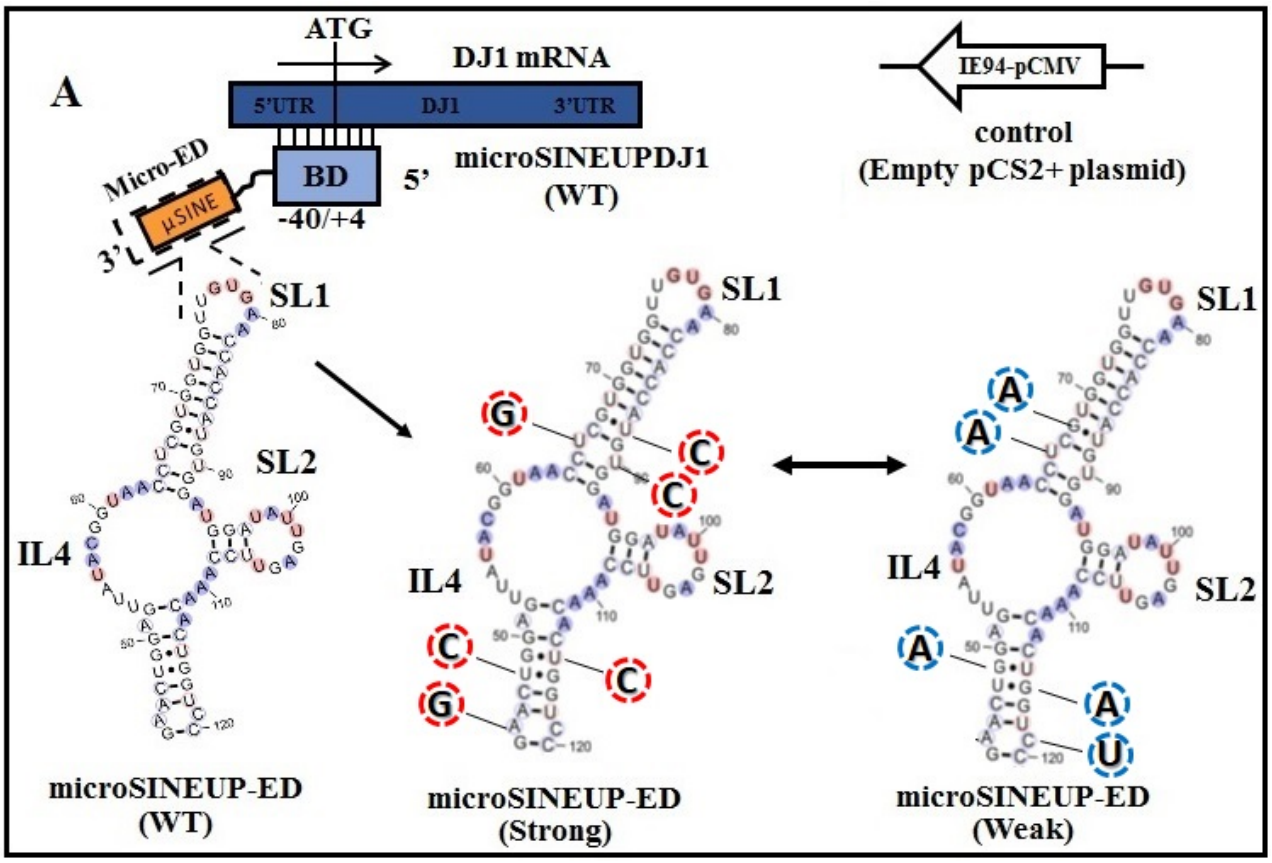


Figure 18 MicroSINEUP-DJ1 is active. (A) Scheme of domain organization in synthetic miniSINEUP-DJ1 with invSINEB2 ED, 167nt (red, top) and microSINEUP-DJ1 with micro-ED, 76nt (orange) sequence containing the stable SL1 hairpin structure. Secondary structure of invSINEB2 (modified from Podbevsek et al., 2018), showing the highlighted region that contains SL1 hairpin structure sequence (red box) that was used to generate microSINEUP-ED. Combining it with -40/+4 BD sequence (light blue) targeting DJ1 mRNA (top) to create synthetic microSINEUP-DJ1. (B) Western blotting of DJ1 protein in HEK 293T cells was detected 48hours post-transfection with the indicated synthetic mini- and micro-SINEUP-DJ1 constructs expressed from pCS2+ plasmid as normalized against empty pCS2+ plasmid control. DJ1 detection was carried out with monoclonal anti-DJ1 antibody a dn the quantification was normalized to β-Actin that was used as loading control. (C) Average DJ1 protein fold change quantification plots showing DJ1 protein upregulation by transfected plasmids expressing mini- and micro-SINEUP-DJ1. Expression of DJ1 mRNA (grey bar, D) and

SINEUP constructs RNA (white bars, E) were monitored by qRT-PCR with respective specific primers. All plots indicate mean ± standard deviation, and representation of N=5 independent replicates $^{***}p < 0.001$, NS; not significant.

4.2.5 Synthetic microSINEUP-ED with modified GC-content has better SINEUP activity

According to structural data, microSINEUPs presented five mismatches in pairing formation. Therefore I decided to investigate the effects of establishing a perfect match in all these pairing. To this purpose, I generated two modified versions of microSINEUPs: microSINEUP with "strong" ED with all GC pairs and microSINEUP with "weak" ED with all AU pairs. The microSINEUP "strong" was generated by converting all non-Watson-Crick base pairing in the original microSINEUP-DJ1 sequence to canonical Watson-Crick base pairs in the microSINEUP ED, by substituting all U for C nucleotides and A for G nucleotides to form complementary G-C base-pair, increasing the GC content of the microSINEUP-DJ1 as shown (**Figure *19***A, middle structure). Likewise, in microSINEUP-DJ1 "weak", all G and C nucleotides in non-Watson-Crick base-pairing in the ED with U were converted to A and U respectively, in order to form canonical U-A base-pairs (**Figure *19***A, right structure). All construct were cloned in pCS2+, using previously tested microSINEUP-DJ1and empty pCS2+ as controls. When transfected in HEK 293T cells, both microSINEUP-DJ1 constructs showed SINEUP activity on endogenously expressed DJ1 (**Figure *19***B). However, microSINEUP-DJ1 "strong" presented a 90% increase in activity respect to WT SINEUP.


A
ATG
DJ1 mRNA
5'UTR
DJ1
3'UTR
Micro-ED
μSINE
BD
-40/+4
5'
3'
microSINEUPDJ1
(WT)
IE94-pCMV
control
(Empty pCS2+ plasmid)
SL1
SL2
IL4
microSINEUP-ED
(WT)
microSINEUP-ED
(Strong)
microSINEUP-ED
(Weak)

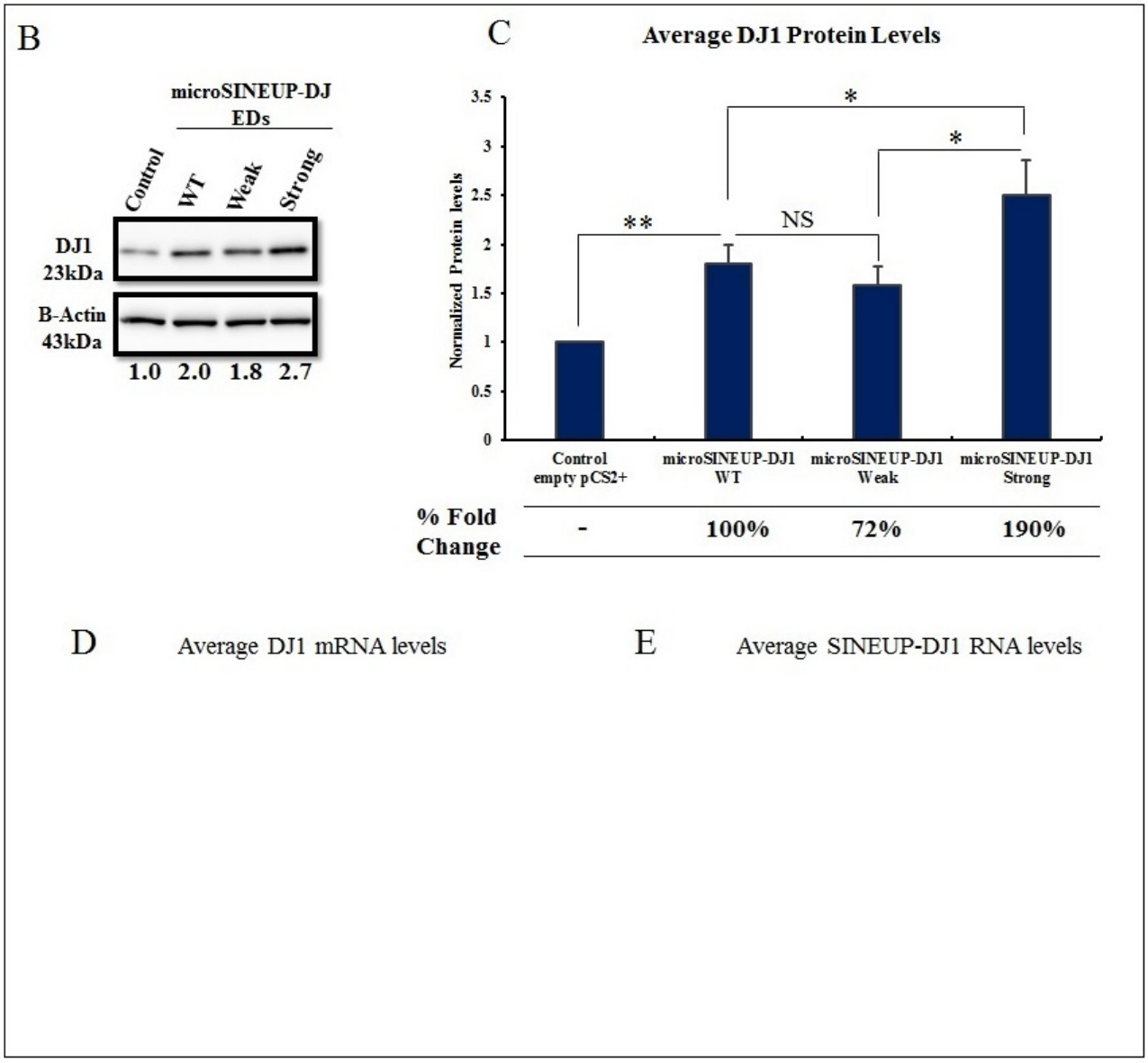


Figure 19 microSINEUP-DJ1 with modified Watson-Crick base pairing has increased SINEUP activity in HEK 293 cells. Scheme showing the microSINEUP-ED, 77nt (orange) and BD, 44nt (light blue) targeting DJ1 mRNA (blue). The region within the microSINEUP-ED where the modifications were made are indicated. The microSINEUP-ED, strong has modified GC content (circle in red) and weak has modified AT content (circle in blue). All modifications were made to convert non-Watson-Crick base pairing as well as non-complementary nucleotides to canonical Watson-Crick base pairs inside the micro-ED without disrupting the conserved SL1 hairpin-GUG. HEK 293T cells were transfected with microSINEUP-DJ1 -WT, -"strong" and –"weak" constructs expressed in pCS2+ plasmid. Control cells were transfected with an empty pCS2+ plasmid. (B) Western blotting of DJ1 protein was detected 48hours post-transfection, using monoclonal anti-DJ1 antibody with quantification normalized to β-Actin that was used as loading control. (C) Average DJ1 protein fold change quantification plots. Expression of DJ1 mRNA (grey bar, D) and SINEUP constructs RNA (white bars, E) were monitored by qRT-PCR with respective specific primers. All plots indicate mean ± standard deviation, and

representation of N=4 independent replicates, * p <0.05, **p <0.01 were considered significant and NS; not significant

4.2.6 Expressed Synthetic miniSINEUP-GFP from a peGFP pDual-gene construct has optimal function

For targeted overexpressed genes like GFP, co-transfection of plasmids expressing the respective synthetic SINEUP-GFP and targeted GFP genes are required. Therefore, I created a pDual gene construct harbouring both a miniSINEUP-GFP and the targeted GFP gene, under transcriptional control from RNA pol III and II promoters, respectively (Figure *20*A).

MiniSINEUP-GFP could show a very strong activation in the range of about 400% fold induction (Figure *20*B and C). There was no effect on GFP-mRNA transcription levels (Figure *20*D). As previously reported, a co-transfection of pcDNA3.1- expressing miniSINEUP-GFP plasmid and peGFP plasmid that expresses targeted GFP mRNA has an average of 2.4 fold GFP protein synthesis in HEK 293T cells (Zucchelli *et al*., 2015a). Taken together, these data suggest that pDual construct is a powerful molecular tool to take advantage of improved SINEUP activity.

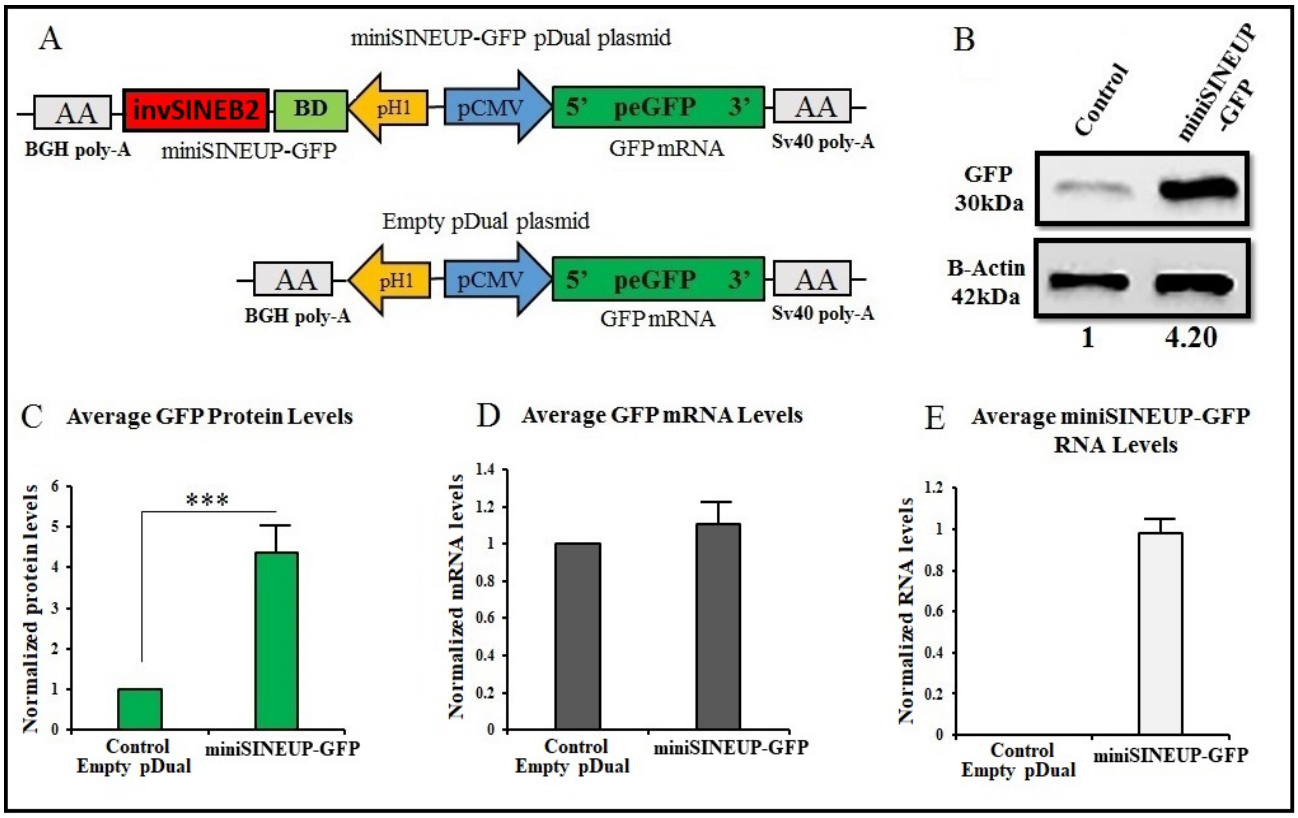


Figure 20 **miniSINEUP-GFP expressed from a pDual plasmid has optimal SINEUP activity in HEK 293T cells.** (A) Scheme of the pDual-miniSINEUP-GFP/peGFP plasmid construct showing miniSINEUP-GFP construct; invSINEB2 ED, 167nt (red) and BD, 72nt (light green) under the control of human H1 (orange), a pol III promoter (orange) and targeted GFP construct (green) that is expressed under CMV pol II promoter (blue). Both genes were expressed from the same pDual construct, with the miniSINEUP-GFP and GFP transcripts harbouring BGH poly-A and SV40 poly-A signals, respectively. The control empty pDual plasmid construct expresses only GFP mRNA but not the miniSINEUP-GFP. HEK 293T cells were transfected with miniSINEUP-GFP/peGFP pDual plasmid constructs. Control cells were transfected with an empty pDual plasmid, expressing only GFP mRNA. (B) Western blotting of GFP protein was detected 48hours post-transfection, using monoclonal anti-GFP antibody and the quantification normalized to β-Actin expression that was used as a loading control. (C) Average GFP protein fold change quantification with standard deviation plots. Expression of GFP mRNA (grey bar, D) and miniSINEUP-GFP RNA (white bars, E) were monitored by qRT-PCR with respective specific primers. All plots indicate mean ± standard deviation, and representation of N=4 independent replicates, $^{***}p < 0.01$ was indicated as significant result.

4.2.7 Synthetic SINEUP-GFP is functional in *Drosophila* S2 cell lines in vitro

So far, SINEUP activity has been shown in several mammalian systems including HEK 293T, HeLa, N2A, U2OS, HepG2, MN9D cell lines in *in vitro* studies. The fact that synthetic SINEUPs could enhance translation of both endogenously and overexpressed genes, makes them very promising as versatile molecular tools. This raises an interest to monitor the activity of synthetic SINEUPs in non-mammalian systems like yeast and flies. Here, I tested the versatility of synthetic SINEUP-GFP function in *Drosophila* S2 cell lines *in vitro*. In brief, full-length synthetic SINEUP-GFP and targeted peGFP were cloned in *Drosophila* expression pACTA vector and used for SINEUP activity assay. Cells were co-transfected with sense GFP and full-length synthetic SINEUP-GFP. Interestingly, synthetic SINEUP-GFP increased of two-fold GFP protein synthesis, post-transcriptionally in *Drosophila* S2 cell lines *in vitro* (Figure 21A and B), suggesting that pathways targeted by SINEUPs are conserved in flies. These data suggest a possible SINEUP activity dependence on evolutionary conserved structures of embedded TEs across the eukaryotic genome, since there are no known invSINEB2 elements in *Drosophila* species. Furthermore, this result proves the versatility of SINEUP technology for potential applications in other model organisms. In this context, it would be interesting to test for potential SINEUP activity in yeast cells as well.

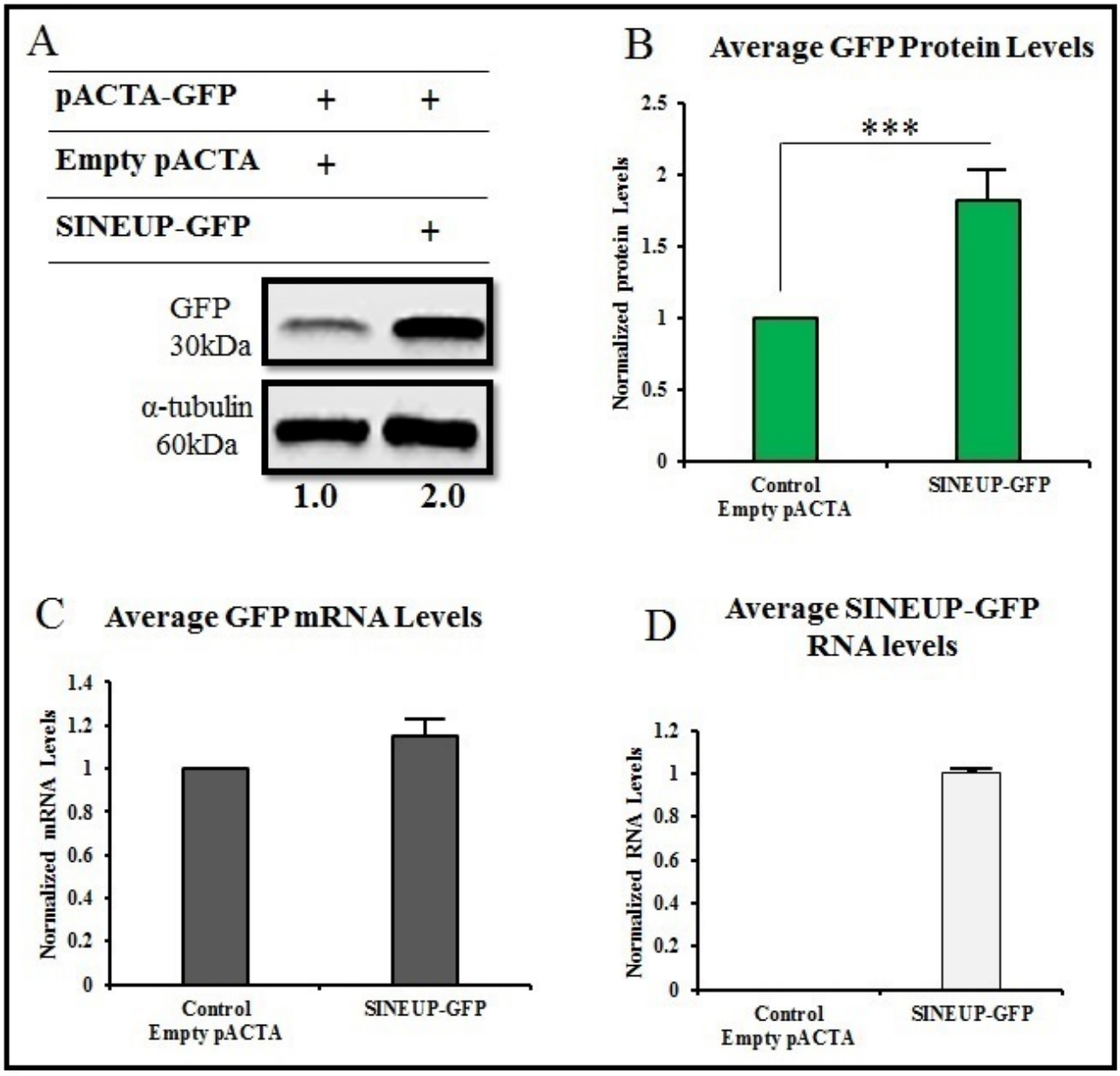


Figure 21 **Synthetic SINEUP-GFP is active in *Drosophila* cells in vitro.** Drosophila S2 cells were co-transfected with Drosophila pACTA plasmid constructs, expressing cloned synthetic miniSINEUP-GFP and GFP genes. Control cells were co-transfected with an empty pACTA and pACTA-GFP plasmids. (A) GFP protein was detected by westerbn blot 48hours post-transfection using monoclonal anti-GFP antibody with the quantification normalized to the β-actin protein. (B) Average GFP protein fold change plot were made from independent GFP quantified replicas in (A). Expression of GFP mRNA (grey bar, D) and miniSINEUP-GFP RNA (white bars, E) were monitored by qRT-PCR with

respective specific primers. All plots indicate mean ± standard deviation, and representation of N=5 independent replicates, $^{***}p < 0.001$

4.3 Translation regulatory pathways mediated by SINEUPs in activation of targeted proteins synthesis

Regulation of protein synthesis usually occurs at the level of translation initiation, which is a rate limiting step, where the translation machinery is recruited to the starting AUG (Hinnebusch, 2014; Hinnebusch and Lorsch, 2012; Sonenberg and Hinnebusch, 2009). Cap-dependent translation involves recognition and binding of the m^7G-cap by eIF4E through eIF4F-complex assembly, which is subjected to various regulations by eIF4E-binding proteins (4E-BPs) under key signaling pathways such as mTOR (Fraser, 2015; Roux and Topisirovic, 2018). This eIF4F-complex assembly formation is inhibited by 4E-BPs when hypo-phosphorylated through inhibition of mTORC1 by molecules such as rapamycin (Sonenberg and Hinnebusch, 2009), leading to cap-dependent translation inhibition and activation of cap-independent translation pathways.

Interestingly, AS Uchl1 activity is triggered by mTORC1 inhibition by rapamycin (Carrieri et al., 2012). Currently, cap-independent translation pathways are known for viral and cellular IRESs (Hellen and Sarnow, 2001; Sonenberg and Hinnebusch, 2009), and 3' and 5' UTR CITEs (Coots et al., 2017; Meyer et al., 2015; Terenin et al., 2013; Zhou et al., 2015), acting as cis-regulatory elements. These elements can be predicted from sequences features and tested experimentally for their ability to promote Cap-independent translation initiation.

Since natural SINEUPs and IRESs are active upon stress, I hypothesized that SINEUP activity involves an IRES-like mechanism. To this purpose, I tested bioinformatically whether the invSINEB2 sequence of natural SINEUPs shares features with known viral and cellular IRES. Then I verified experimentally functional similarities between the invSINEB2 EDs of SINEUPs and IRES using IRES- and SINEUP-activity assays.

While so far there are no known consensus sequence and structural motifs that define a sequence for having IRES-like activity in cells (Baird et al., 2007), I bioinformatically compared the invSINEB2 sequences of natural SINEUPs reported in Carrieri et al., 2012 with experimentally verified IRES sequences reported in the IRESsite database (Mokrejs et al., 2006; http://iresite.org/IRESite_web.php)

for similarities in both sequence and structural motifs. Interestingly, I obtained alignments with good scores and *E*-values for the large majority of invSINEB2 sequences of natural SINEUPs with the annotated cellular and viral IRESs motifs from the IRESsite database ***(Table 3)***.

The invSINEB2 of As Uchl1 was further analyzed, since it was the SINEUP ED with the largest number of statistically significant alignment to annotated IRESs reported in the IRESsite database. By taking advantage of the RegRNA browser (Chang et al., 2013 http://regrna.mbc.nctu.edu.tw/index1.php), IRES sequences were identified within the invSINEB2 (highlighted in violet) (***Figure 22***) **(appendix**

***Figure 36*)**. By the use of this bioinformatics tool, other known *cis*-acting regulatory elements were identified: a terminal oligopyrimidine tracts (TOPs) and upstream open reading frames (uORFs). Finally, using the SRAM browser (Zhou et al., 2016; http://www.cuilab.cn/sramp/#predSRAMP), N6-methyladenosine (m^6A) RRACH motifs (highlighted in green) was predicted in the invSINEB2 sequence as potential 5' UTR CITEs that could additionally enhance invSINEB2 ED activity **(*Figure 22* and appendix**

***Figure 36*)**.

Aside all these bioinformatics predictions of IRES-like motifs and other *cis*-regulatory elements in SINEUP-EDs that could potentially be responsible for SINEUP activity, I was particular aware that bioinformatics prediction of these *cis*-acting regulatory elements in SINEUPs and correlation with IRES-like activity are not conclusive and need experimental validation (Baird et al., 2007; Thompson, 2012).

4.3.1 The SINEUP-IRES hypothesis

Here I propose the SINEUP-IRES hypothesis that both natural and synthetic SINEUPs use IRES-like internal ribosome entry mechanisms to upregulate translation of target mRNAs. I therefore designed 3 main experimental models to test my hypothesis as follows;

- Model I: I tested whether SINEUP ED works as IRES in IRES-activity assays.
- Model II: I tested whether IRESs are SINEUP ED in SINEUP activity assays.
- model III: I searched for naturally occurring SINEUPs as embedded in the 5'UTR of viral and cellular IRES-containing mRNAs.

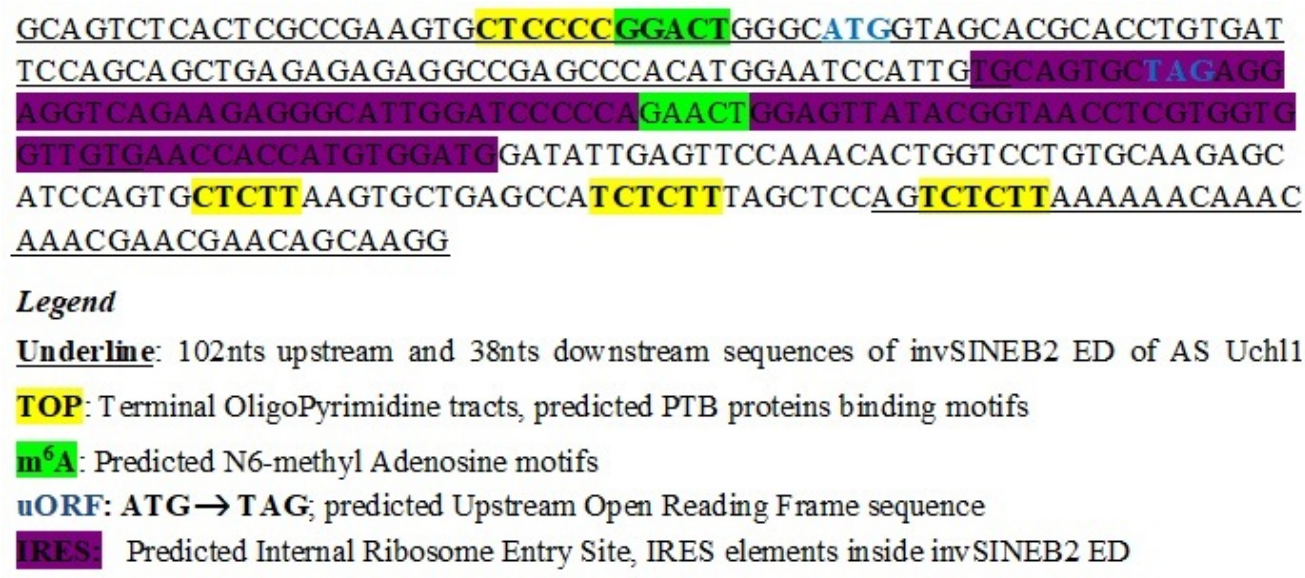


Figure 22 **Predicted cis-acting regulatory elements in invSINEB2 with spacer sequence from SINEUP of AS Uchl1**. Snapshot of inverted SINEB2 ED (167nt) with underlined upstream (102nt) and downstream (38nt) spacer sequences from AS Uchl1. Highlighted are predicted cis-regulatory elements within the 307nt SINEUP-ED sequence that was used in IRES-activity assay. In yellow are predicted TOP sequences which are targets of PTB proteins, in green are predicted 5'UTR m^6A motifs which can act as cap-independent translation enhancer (CITE) to promoted cap-independent translation activation and in violet is the predicted IRES sequence within the 167nt invSINEB2 that could mediate cap-independent translation of proteins in IRES-activity assay. ATG-TAG, highlights the start and stop codons of a predicted upstream open reading frame (uORF) sequence that could also recruit translation competent ribosome. uORF, TOPs and IRES motifs were predicted with (Chang et al., 2013; http://regrna.mbc.nctu.edu.tw/index1.php) browser, while m^6A motifs were predicted with (Zhou et al., 2016; http://www.cuilab.cn/sramp/#predSRAMP) browser.

Table 3 Sequence and structural motifs alignment score of invSINEB2 of SINEUPs and IRES sequence annotated in IRESsite database (Mokrejs et al., 2006; http://iresite.org/IRESite_web.php)

Annotated SINEUP	AS to Gene	SINE/B2 Portion	Alignment	IRES-Type	Score	E-Value
AK078321.1	AS Uchl1					
AK029359.1	AS Uxt		EMCV-R virus	Viral	26	0.054
AK019925.1			ELG1 gene	Cellular	26	0.038
			HAV_HM175 virus	Viral	23	0.43
AK032194.1	Nars2		HIV-1 virus	Viral	23	0.39
AK032215.1	Nudt9		none			
AK034331.1	n/a		PV_type3_Leon virus	Viral	21	0.71
AK035015.1	Nrm		IGF2_leader2 gene	Cellular	21	0.65
			EMCV-R virus	Viral	21	0.65
AK035406.1	Sv2b		ELG1 gene	Cellular	33	2.00E-04
AK041236.1	Ccdc88a		none			
AK041654.1	Rcc		none			
AK041742.1	Abhd11		AQP4 gene	Cellular	23	0.37
AK042861.1	Wfdc5		HAV_HM175 virus	Viral	24	0.11
AK044205.1	Rhod		none			
AK045677.1	Eln		none			
AK047213.1	Uhmk1		HAV_HM175 virus	Viral	23	0.38
AK048309.1	Epb4.9		ELG1 gene	Cellular	26	0.035
AK053130.1	Rabgap11	SINE/B2 159-275	none			
		SINE/B2 1207-1363	HAV_HM175 virus		21	0.95
AK054076.1	Gadd45a	SINE/B2 277-407	PSIV_IGR virus	Viral	21	0.78
			MTG8a gene	Cellular	21	0.78
			HAV_HM175 virus	Viral	21	0.78
			EMCV-R virus	Viral	21	0.78
			Apaf-1 gene	Cellular	21	0.78
		SINE/B2 1747-1860	none			
AK078161.1	Nck1	SINE/B2 1134-1337	ELG1 gene	Cellular	24	0.11
			CVB3 virus	Viral	23	0.37
AK080749.1	Pgls	SINE/B2 1081-1179	HAV_HM175 virus	Viral	23	0.17
AK135599.1	Ednra	SINE/B2 228-409	none			
		SINE/B2 2428-2606	none			

AK143014.1	Cdkn2aip	SINE/B2 1470-1683	MTG8a gene	Cellular	24	0.12
			EMCV-R virus	Viral	24	0.12
			CVB3 virus	Viral	23	0.39
AK143784.1	Txnip	SINE/B2 1389-1495	HAV_HM175 virus	Viral	23	0.18
			EMCV-R virus	Viral	23	0.18
			PV_type1_Mahoney virus	Viral	21	0.61
AK145079.1	Gsk3b	SINE/B2 2953-3143	HAV_HM175 virus	Viral	23	0.35
			ELG1 gene	Cellular	23	0.35
AK149843.1	Cmtm6	SINE/B2 2001-2167	none			
		SINE/B2 2270-2439	none			
AK163105.1	E4f1	SINE/B2 1080-1254	none			
AK165234.1	Dtx3	SINE/B2 1401-1588	none			
		SINE/B2 1926-2064	none			
		SINE/B2 2153-2335	ELG1 gene	Cellular	26	0.03
AK169421.1	n/a	SINE/B2 853-1043	CrPV_IGR virus	Viral	23	0.35
		SINE/B2 1247-1420	EMCV-R virus	Viral	26	0.028
		SINE/B2 1742-1936	CVB3 virus	Viral	24	0.11
			IGF2_leader2 gene	Cellular	23	0.36
			ELG1 gene	Cellular	23	0.36

4.3.2 Model I: SINEUP ED is an IRESs

Here, I tested whether the invSINEB2 sequence of AS Uchl1 has IRES activity in a bicistronic reporter assay.

4.3.2.1 InvSINEB2 of AS Uchl1 presents IRES activity in a bicistronic assay.

Here, I tested whether the invSINEB2 sequence works as IRES **(*Figure 22* and** Error! Reference source not found.**).** To this purpose, I took advantage of the widely used bicistronic reporter assay in which a single transcript is expected to be transcribed, but the translation of renilla luciferase (Rluc) is via cap-dependent mechanisms, whereas firefly luciferase (Fluc) is only translated if preceded by an IRES-like sequence through cap-independent translation mechanisms (Thompson, 2012; Van Eden et al., 2004).

In this experiment, the 167nt sequences of direct and inverted SINEB2 were cloned between Renilla luciferase (Rluc) and Firefly luciferase (Fluc) reporters in a pRUF plasmid. The empty pRUF plasmid and constructs containing inverted (inv) and direct (dir) SINEB2 and c-myc-IRES cloned between Rluc and Fluc were used in transient transfection into HEK 293T cells ***(Figure 22A).*** Cells were harvested 48 hours post transfection, and subjected to the Dual-GLo assay system (Promega). c-myc-IRES pRUF and empty pRUF plasmids transfections were used as positive and negative IRES-activity controls, respectively, as previously reported (Bisio et al., 2010, 2015). IRES-activity is defined as a relative measure of Fluc/Rluc activities. Interestingly, Idetected IRES-activity mediated by invSINEB2 sequence but not dirSINEB2 sequence as compared to c-myc-IRES positive control sequence ***(Figure 22D).*** Furthermore, when I doubled the number of invSINEB2 sequence, I increased significantly ($p<0.001$) the fold induction of IRES-like activity mediated by the invSINEB2 ***(Figure 22D).***

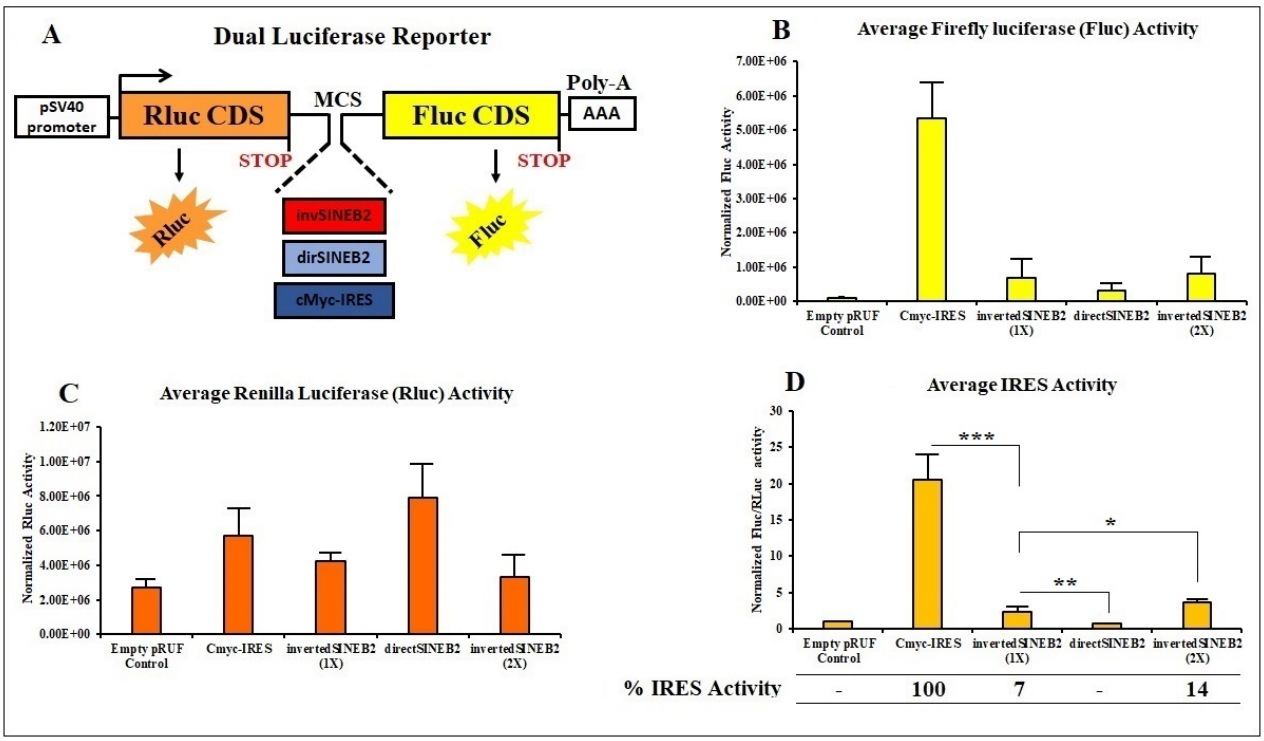


Figure 23 I**nverted SINEB2 sequence of AS Uchl1 SINEUP has IRES-activity. (**A) Schematic representation of pRUF-dual luciferase reporter vectors, containing empty pRUF multiple cloning sites (MCS), cloned inverted (red) and direct SINEB2 (light blue) of AS Uchl1 and c-myc-IRES (blue) sequences that are flanked by Renilla luciferase (Rluc CDS, brown) cDNA upstream and firefly luciferase (Fluc CDS, yellow) cDNA-downstream, all under the influence of constitutive pSV40 promoter. Inverted SINEB2 clone were tested as a single (1x) and double (2x) sequence. Each pRUF construct produces a single bicistronic-transcript with Rluc protein translated in a cap-dependent manner, while Fluc protein could be translated in a cap-independent manner. The empty pRUF-MCS and pRUF- c-myc-IRES constructs were used as negative and positive controls respectively. (B and C) Measured quantities of luciferase activities performed in HEK 293T cells transiently transfected with the respective individual pRUF dual-luciferase reporters. Results are shown as average of Rluc (B) and Fluc (C) measured activities, which are quantified from raw data of at least four independent biological replicas and each performed in duplicates. (D) IRES-activity was calculated as ratio of Fluc (C) to Rluc (B) measured activities in relative light units, and normalized against empty pRUF-MCS vector background activities. Empty pRUF activity value was set at 1. InvSINEB2 (2x) has twice the activity of the InvSINEB2 (1x). Error bars show standard deviation (SD) from four biological replicates. Percentage IRES-activity were estimated relative to the measured activity of the c-myc-IRES construct, which was arbitrarily set to 100%. invSINEB2 (1x) and (2x) but not dirSINEB2 had 7 and 14% respectively of c-myc-IRES activity (D). All plots indicate mean ± standard deviation, and representation of N=4 independent duplicate-replicas, $^{*}p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$ were considered significant.

4.3.2.2 InvSINEB2 but not dirSINEB2 with surrounding sequences in AS Uchl1 has optimal IRES-activity

I then hypothesized that additional flanking sequences to invSINEB2 may be required for its correct folding. Therefore I generated invSINEB2 and dirSINEB2 with flanking upstream and downstream sequence from AS Uchl1 SINEUP-ED in pRUF vector ***(Figure 24A)*** and performed the dual-luciferase reporter assay as previously described. The 102nt upstream and 38nt downstream sequences were selected avoiding the inclusion of the upstream BD and downstream embedded Alu repeat sequences from SINEUP of AS Uchl1**(*Figure 22*).** Here again, c-myc-IRES in pRUF and empty pRUF plasmid were used as positive and negative controls respectively. Fluc and Rluc activities were respectively measured ***(Figure 24B and C).*** Interestingly, a significant increase in Fluc activity by invSINEB2 with surrounded spacer sequences was measured, indicating cap-independent translation of Fluc proteins mediated by both c-myc-IRES and invSINEB2, but not dirSINEB2 ***(Figure 24B).*** Importantly, the invSINEB2-spacer sequence presented a significantly higher IRES-activity ($p<0.05$) than c-myc-IRES ***(Figure 37).*** These results were confirmed with the same experiments performed in U2OS cell lines ***(Appendix Figure 37).*** Here again, there was significant higher IRES-activity ($p<0.05$) mediated by invSINEB2-spacer as compared to c-myc-IRES ***(Figure 37).*** Altogether, our data confirmed that invSINEB2 but not dirSINEB2 with spacer sequences from AS Uchl1 has IRES-activity in a bicistronic reporter assay.

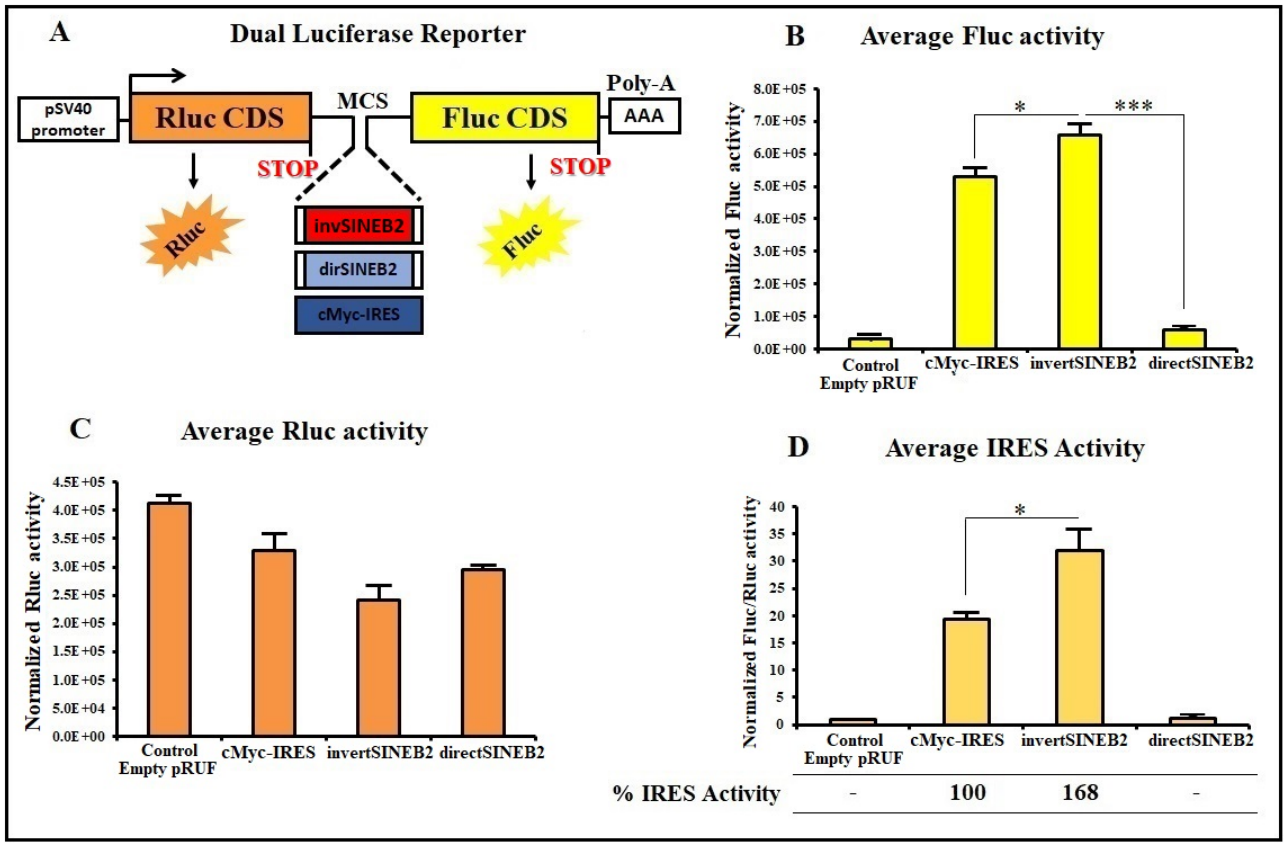


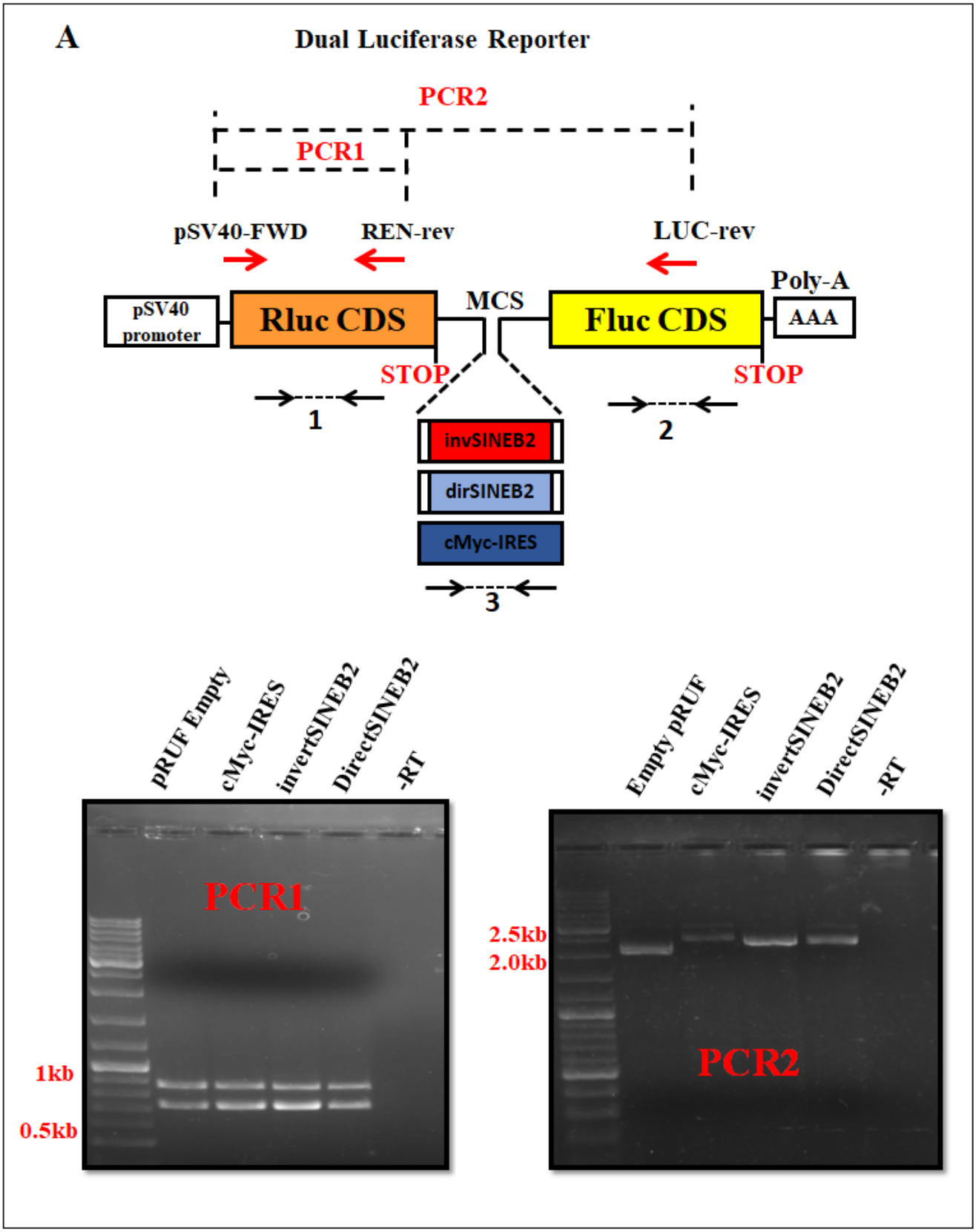
Figure 24 **InvSINEB2 with surrounding sequences in AS Uchl1 SINEUP has optimal IRES-activity. (**A) Schematic representation of pRUF-dual luciferase reporter vectors: containing empty pRUF multiple cloning sites (MSC), cloned inverted (red) and direct SINEB2 (light blue) with up- and down-stream sequences (white) of AS Uchl1 SINEUP-ED and c-myc-IRES (blue) sequences flanked by Renilla luciferase (Rluc CDS, brown) cDNA upstream and firefly luciferase (Fluc CDS, yellow) cDNA downstream, all under the influence of constitutive pSV40 promoter. Each vector produces a single bicistronic-transcript with Rluc protein translated in a cap-dependent manner, while Fluc protein could be translated in a cap-independent manner. The empty pRUF-MCS and pRUF- c-myc-IRES constructs were used as negative and positive controls respectively. (B and C) Measured quantities of luciferase activities were performed after 48hours in HEK 293T cells transiently transfected with the respective individual pRUF dual-luciferase reporter constructs. Results are shown as average of Rluc (B) and Fluc (C) measured activities, quantified from raw data of at least four independent biological replicas, each performed in duplicates. (C) Measured Fluc activities are due to Fluc cap-independent protein synthesis induced by the inv- and dir-SINEB2 with spacer sequences and c-myc-IRES inserted sequence. (D) IRES-activity was calculated as ratio of Fluc (C) to Rluc (B) measured activities in relative light units, and normalized against empty pRUF-MCS plasmid background activities. Empty pRUF activity value was set at 1. c-myc-IRES activity was arbitrarily set at 100%. Percentage IRES-activities were estimated relative to the measured activity of the c-myc-IRES construct. InvSINEB2 but not dirSINEB2 with spacer sequences has optimal IRES-activity, about 68% more on average than c-myc-IRES. Error bars show standard deviation (SD) from four biological replicates. All plots indicate mean ± SD, and representation of N=4 independent duplicate-replicas; $^{*}p < 0.05$ and $^{***}p < 0.001$ were considered significant.

4.3.2.3 InvSINEB2 mediated cap-independent translation of Fluc protein synthesis is neither due to cryptic promoter activity nor alternative splicing events

Drawbacks of the bicistronic reporter assay in determining cap-independent translation initiation are possible events of alternative splicing and cryptic promoter activities that could be mediated by the inserted invSINEB2-spacer sequences between Rluc and Fluc cDNAs. These are the so-called technical artifacts of the bicistronic reporter assay (Thompson, 2012). Cryptic promoter and alternative splicing events could generate a false positive result (Van Eden, 2004). To rule out these possibilities, I use approaches reported before (Bisio et al., 2010, 2015), since our empty pRUF and c-myc-IRES controls were used in these reported works. First, I checked for the presence of full-length transcripts comprising of Rluc-insert-Fluc sequence produced by transfection of each c-myc-IRES-, invSINEB2-spacer-, dirSINEB2-spacer-pRUF constructs in both HEK 293T and U2O cells by RT-qPCR. In brief, total RNA was extracted from transfected cells and the various cDNAs were reverse-transcribed. Two sets of PCR, 1 and 2 were separately performed, using 2 pairs of primers as shown ***(Figure 25A, upper panel).*** The forward primer (pSV40-FWD) was the same for PCR 1 and 2, and was chosen to anneal close to the transcription start site. In PCR 1, the reverse primer (REN-rev) anneals to Rluc cDNA and in PCR 2, the reverse primer 2 (LUC-rev) anneals to Fluc cDNA as shown schematically ***(Figure 25A, upper panel).*** The PCR 2 may allow for the detection of full-length transcripts. From both PCR, amplicons of expected sizes were obtained as shown in the gel image ***(Figure 25A, lower panel).*** Indeed, PCR2 confirmed a single full-length transcript of 2kb, 2.4kb and 2.3kb respective sizes for the empty pRUF, c-myc-IRES-pRUF and inv/dirSINEB2-spacer-pRUF constructs, proving the absence of splicing events mediated by the inv/dirSINEB2-spacer sequences.

The Fluc activity could also be due to invSINEB2-spacer sequence mediating cryptic promoter events, resulting in the production of monocistronic Fluc transcripts. Therefore, I performed qRTPCR from the same cDNA synthesized by RT-PCR to measure the relative amounts of expressed Rluc, Fluc, c-myc-IRES and SINE sequences. The Rluc-insert-Fluc transcripts were amplified and quantified at 3 regions as indicated by dotted arrows; 1 and 2 inside Rluc and Fluc respectively, and 3 inside the inserted SINE and c-myc-IRES sequences, using their respective primers ***(Figure 25A, upper panel).*** The estimated results are plotted as normalization ratios between Fluc/Rluc, SINEB2/Rluc and c-myc-IRES/Rluc ***(Figure 25B),*** showing no significant increase in Fluc transcription levels and hence absence of a possible cryptic promoter events. Quantified Fluc/Rluc transcription results of the dirSINEB2-inserted-

pRUF construct could also be used as an additional control since promoters, unlike enhancers, promotes transcription of genes in a directional dependent manner (Bisio et al., 2015). Therefore, one would expect a different Fluc/Rluc qRTPCR ratio ***(Figure 25B, lower panel, right),*** in the presence of a criptic promoter. Similar results were obtained using U2OS cells ***(appendix Figure 37E and F).*** Altogether, I have shown that the invSINEB2 sequence of SINEUPs has IRES-like activity.


A
Dual Luciferase Reporter
PCR2
PCR1
pSV40-FWD
REN-rev
LUC-rev
Poly-A
pSV40 promoter
Rluc CDS
MCS
Fluc CDS
AAA
STOP
STOP
1
2
invSINEB2
dirSINEB2
cMyc-IRES
3
pRUF Empty
cMyc-IRES
invertSINEB2
DirectSINEB2
-RT
PCR1
1kb
0.5kb
Empty pRUF
cMyc-IRES
invertSINEB2
DirectSINEB2
-RT
2.5kb
2.0kb
PCR2

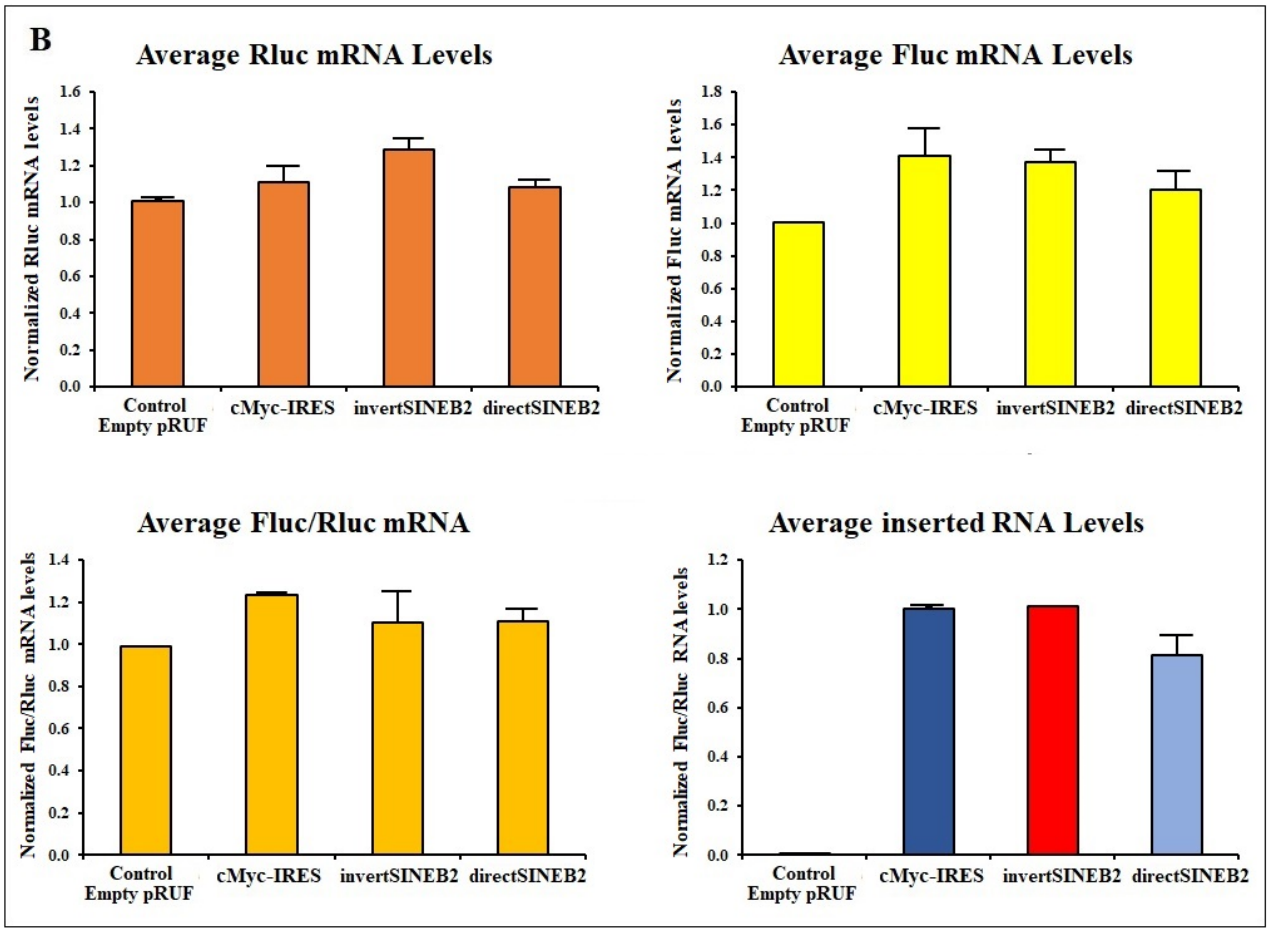


Figure 25 **Inverted SINEB2 has no cryptic promoter function and does not promote alternative splicing.** Control experiments were carried out to exclude the possibilities of either cryptic promoter activities at the DNA level or post-transcriptional alternative mRNA splicing events to be mediated by the inserted SINEB2 sequences between Rluc and Fluc cDNAs. (A) The empty pRUF-MCS and pRUF-c-myc IRES constructs were used as control checks. Total RNAs were extracted upon 48hours from transfection of HEK 293T cells and reversed transcribed by RT-PCR to cDNAs. Then two series of normal PCR 1 and 2 were performed as shown schematically, using a pSV40 forward primer (forward direction red arrow) that anneal downstream the transcription start site but upstream the small intron in the pRUF vector. Combining independently with a reverse primer 1 in renilla (REN-rev, reverse direction red arrow on top of Rluc CDS construct) and 2 in firefly (LUC-rev, reverse direction red arrow on top of Fluc CDS construct) resulted in PCR 1 and 2 products respectively; shown as electropherogram images (lower panel A). I obtained DNA amplicons of the expected sizes, indicating that both reporters were translated from a single bicistronic transcripts. The visible double bands on PCR 1 image are short and long amplicons made from process and un-processed transcripts, which arise due to the presence of an intronic sequence downstream the pSV40 promoter of the pRUF vector backbone. (B) qPCR was performed from the same RT-PCR products to quantify the schematic regions marked by dotted-black arrows 1, 2 and 3 within Rluc, Fluc and the inserted sequences respectively. The mRNA quantities of Rluc (brown bars, upper panel B left) relative to Fluc (yellow bars, upper panel B right) as normalized by the empty pRUF were not statistically different, showing no shorter transcripts containing only Fluc mRNA, indicating cap-independent translation of Fluc protein from a bicistronic transcript (figure 11B) and absence of cryptic promoter possibilities. qRT-PCR also

confirmed the expression of the inserted c-myc-IRES (blue), invSINEB2 (red) and dirSINEB2 (light blue) sequences shown as respective blue, red and light blue bars (lower panel B, right). Error bars show standard deviation (SD) from four biological replicates. All plots indicate mean ± SD, and representation of N=4 independent replicas

4.3.3 Model II: IRESs can function as ED in SINEUPs

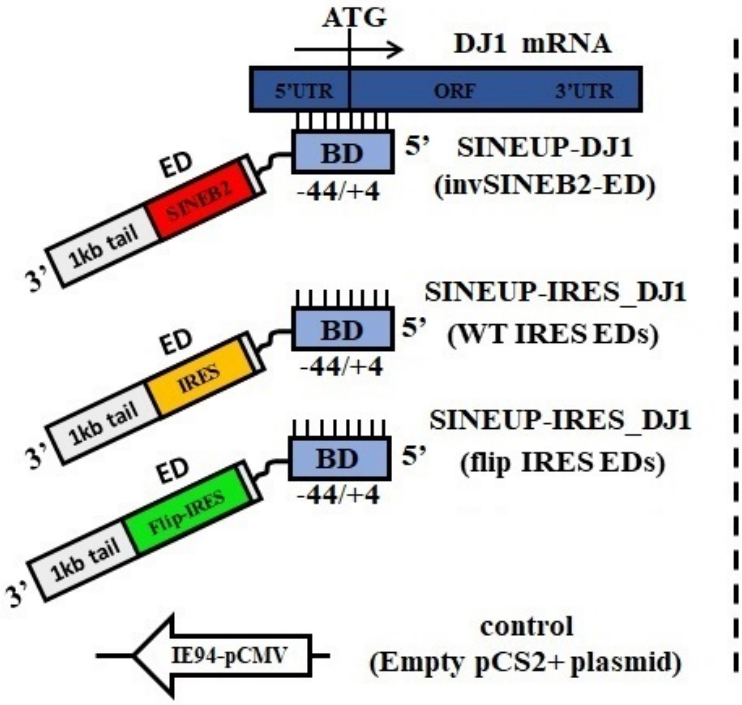
Here, I tested if any cellular and viral IRES sequence or structural motifs reported in the IRESsite (***Table 3;*** Mokrejs et al., 2006; http://iresite.org/IRESite_web.php) database can functionally substitute the invSINEB2 sequence as ED in a synthetic SINEUP.

4.3.3.1 Synthetic SINEUP-DJ1s containing viral and cellular IRES as ED increase endogenous DJ1 protein synthesis in mammalian cells *in vitro*

I generated a new group of synthetic SINEUP-DJ1 by replacing invSINEB2 sequence with annotated IRES and cloned them into pCS2+ expressing plasmid (Carrieri et al., 2012, Zucchelli et al., 2015a). Viral and cellular IRES are listed **(Figure 26A, right)** and were cloned in direct and flipped orientations in Δ5'-AS Uchl1 backbone sequences joined to a 44nt long BD antisense to endogenously expressed DJ1 mRNA ***(Figure 26A)***, as described previously **(*Figure 14*)**. Transient expression of the SINEUP-IRES-DJ1 constructs in HEK 293T cells caused strong and reproducible increase in DJ1 protein synthesis comparable to SINEUP-DJ1 activity ***(Figure 26B and C).*** There was no appreciable increase in DJ1 mRNA from qRT-PCR, indicating a post-transcriptional SINEUP effect ***(Figure 26D, upper panel).*** Noticeably, SINEUP-IRES containing EMCV and CrPV IRES sequences have stronger increase in DJ1 protein synthesis than the canonical SINEUP-DJ1 activity. Surprisingly, I observed lower but significant SINEUP activity for the constructs with flipped IRES sequences as EDs ***(Figure 26B and C),***

Similar results were observed in HepG2 cells ***(Appendix Figure 38).*** Taken together, these data suggest that cellular and viral IRESs have alternative SINEUP-like activity in increasing targeted proteins synthesis.

A

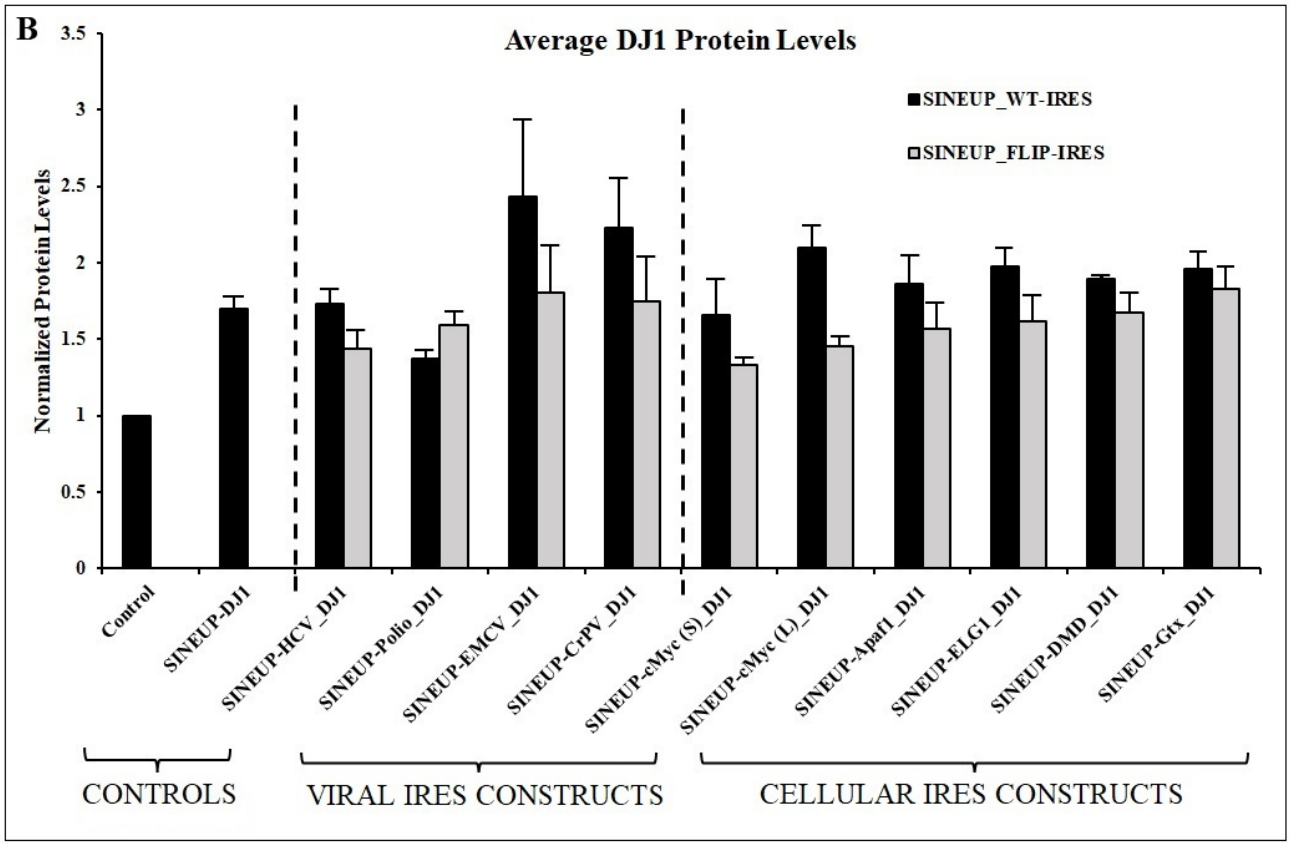


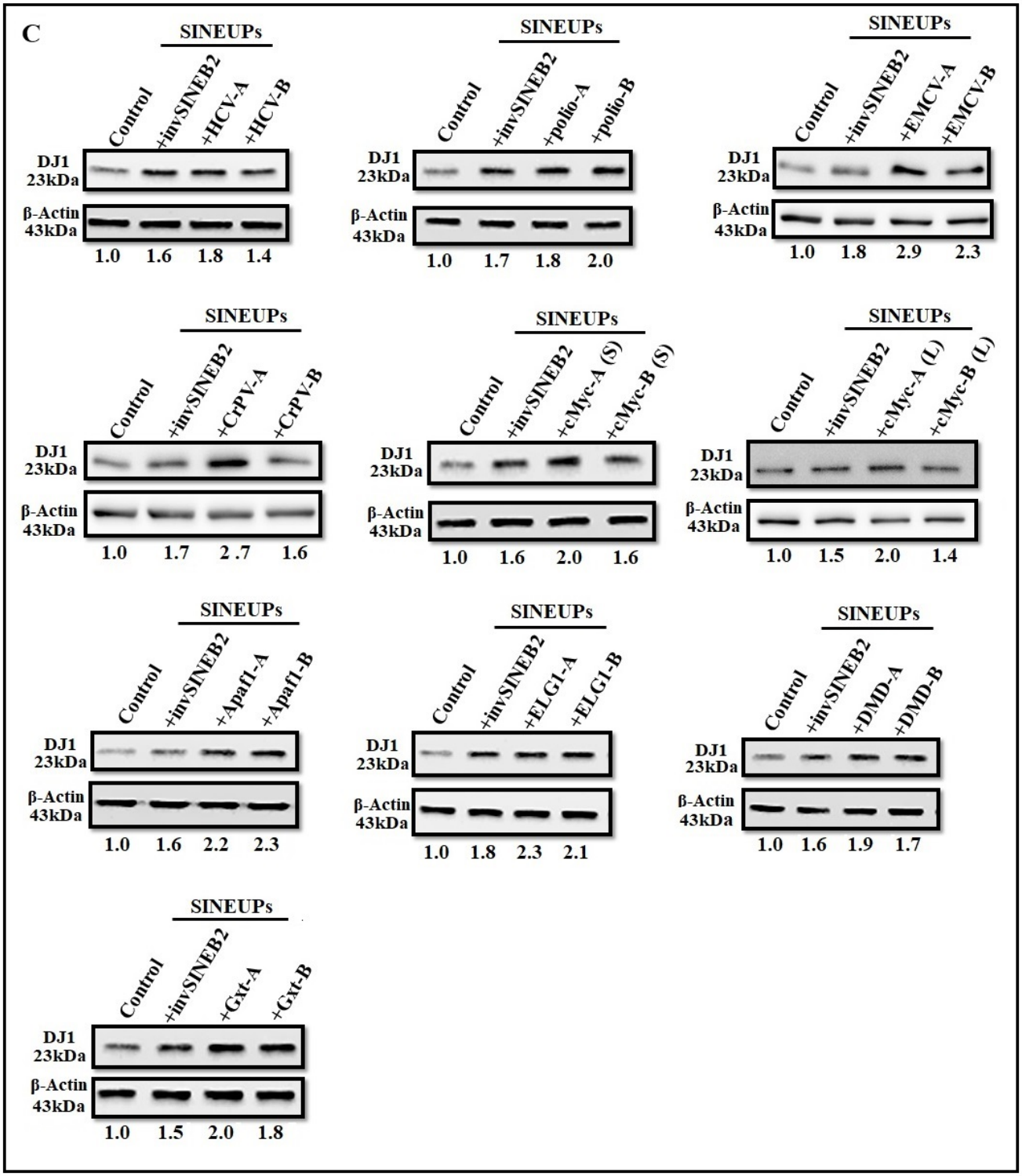
IRESs Effector Domains (EDs) in synthetic SINEUP constructs

Origin	IRES type	Size (bp)
Viral	HCV	383
	Polio	312
	EMCV	576
	CrPV	192
Cellular	cMyc	41 & 395
	Apaf1	231
	ELG1	460
	DMD	71
Synthetic	Gxt	180

Key
A: wild type IRES ED
B: flip IRES ED

B




C
SINEUPs
Control
+invSINEB2
+HCV-A
+HCV-B
DJ1 23kDa
β-Actin 43kDa
1.0 1.6 1.8 1.4
SINEUPs
Control
+invSINEB2
+polio-A
+polio-B
DJ1 23kDa
β-Actin 43kDa
1.0 1.7 1.8 2.0
SINEUPs
Control
+invSINEB2
+EMCV-A
+EMCV-B
DJ1 23kDa
β-Actin 43kDa
1.0 1.8 2.9 2.3
SINEUPs
Control
+invSINEB2
+CrPV-A
+CrPV-B
DJ1 23kDa
β-Actin 43kDa
1.0 1.7 2 .7 1.6
SINEUPs
Control
+invSINEB2
+cMyc-A (S)
+cMyc-B (S)
DJ1 23kDa
β-Actin 43kDa
1.0 1.6 2.0 1.6
SINEUPs
Control
+invSINEB2
+cMyc-A (L)
+cMyc-B (L)
DJ1 23kDa
β-Actin 43kDa
1.0 1.5 2.0 1.4
SINEUPs
Control
+invSINEB2
+Apaf1-A
+Apaf1-B
DJ1 23kDa
β-Actin 43kDa
1.0 1.6 2.2 2.3
SINEUPs
Control
+invSINEB2
+ELG1-A
+ELG1-B
DJ1 23kDa
β-Actin 43kDa
1.0 1.8 2.3 2.1
SINEUPs
Control
+invSINEB2
+DMD-A
+DMD-B
DJ1 23kDa
β-Actin 43kDa
1.0 1.6 1.9 1.7
SINEUPs
Control
+invSINEB2
+Gxt-A
+Gxt-B
DJ1 23kDa
β-Actin 43kDa
1.0 1.5 2.0 1.8

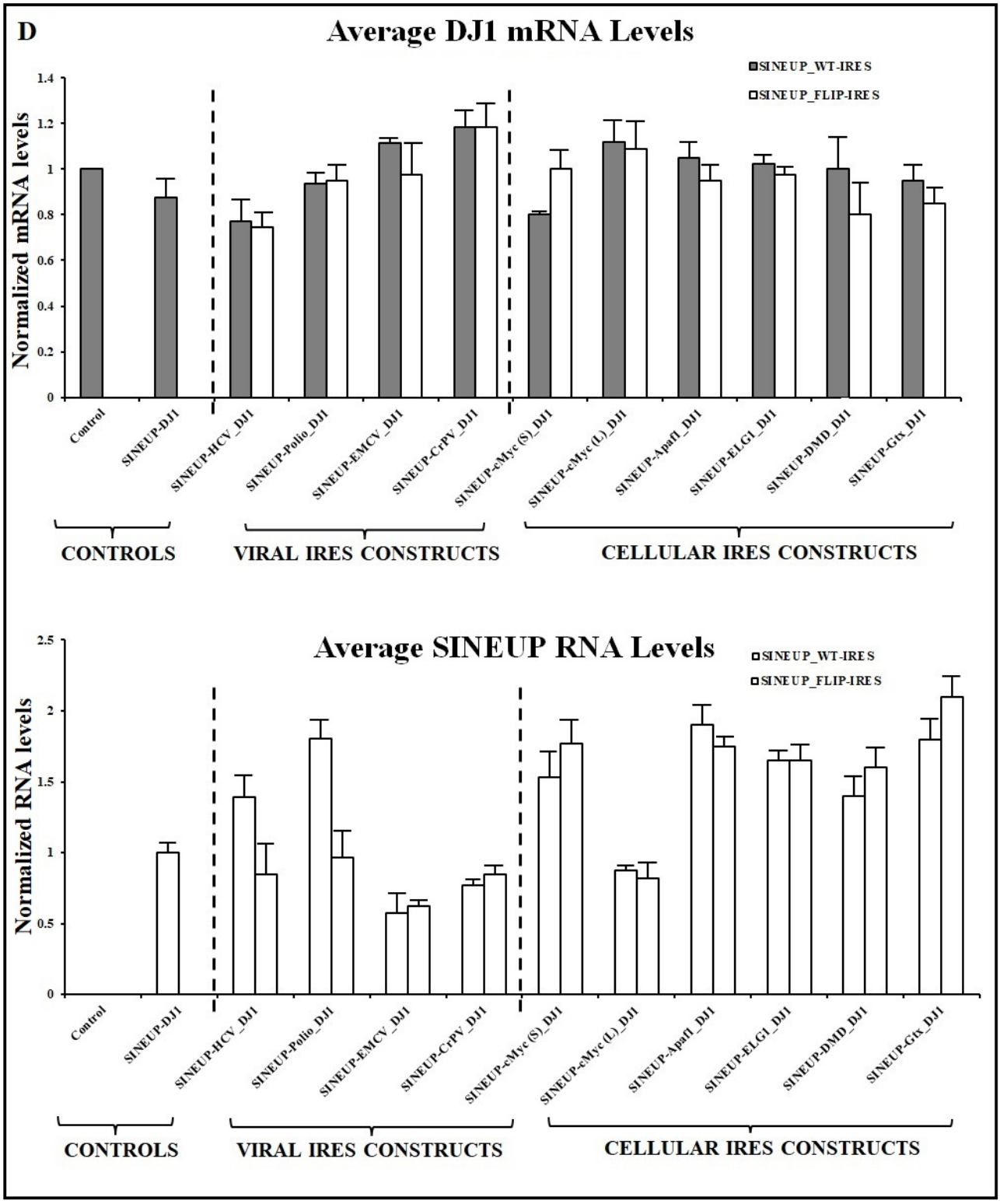


Figure 26 **Synthetic SINEUPs of viral and cellular IRESs increase DJ1 protein synthesis in cells. (A)** Schematic constructs overview of SINEUP-DJ1 and SINEUP-IRES_DJ1 of viral and cellular origins, targeting endogenously expressed DJ1 mRNA (left, blue bar). All SINEUPs were constructed using natural SINEUPs modular architecture to comprise of a short BD of 44nt that is antisense to DJ1

mRNA and a long Δ5'-AS Uchl1 backbone sequence containing EDs of invSINEB2 of AS Uchl1 (in red) for SINEUP-DJ1 construct and all the listed IRES sequences (as tabulated, panel A, right) in direct (Wild-type; WT IRES EDs, orange) and reverted (flip IRES EDs, green) orientations for the SINEUP-IRES_DJ1 constructs, using pCS2+ plasmid as expression vector. (B and C) The SINEUP constructs were transiently expressed in HEK 293T cells and SINEUP activity measured 48 hours post-transfection by WB and qRT-PCR analysis. Western blot analysis was carried out using anti-DJ1 antibody to detect and quantify DJ1 proteins as fold change of no SINEUP-empty pCS2+ vector control. β-actin was used as loading control. (B) Mean DJ1 fold change plot of at least 5 independent replicates are shown. Plots in black represent activities of SINEUP-DJ1 and SINEUP-WT IRES EDs, and empty pCS2+ plasmid control, while SINEUP-flip IRES sequences are plotted in white. (C) Represented western blot analysis showing differential DJ1 protein fold change. (D) RT-qPCR was used to quantify DJ1 mRNA (grey bars for the SINEUP with WT EDs and white bar for SINEUP with flip IRES EDs, upper panel) and compared with no SINEUP-empty vector control, which shows no statistical significant levels (mean ± S.D, $p<0.05$, sample t.Test vs no SINEUP-empty pCS2+ plasmid control). This indicates post-transcriptional upregulation of DJ1 protein synthesis that was mediated by the SINEUP constructs. Synthetic SINEUP RNA sequences (white plot, lower panel B) were also detected and measured by RT-qPCR with primers designed to the 3'tail of Δ5'-ASUchl1 backbone sequence. Empty pCS2+ values in all measurement was arbitrarely set to 1. As shown from the lower panel B plot, some SINEUP constructs were expressed at higher levels than others.

4.3.3.2 Viral and cellular IRESs are portable SINEUP EDs to increase endogenously expressed targeted proteins synthesis

I previously shown that miniSINEUPs were able to maintain SINEUP activity by the exclusive combination of BD and ED sequences (Schein et al., 2016; Zucchelli et al., 2015a). Here, I wanted to probe further the SINEUP-like activity of the IRES sequences (Figure 26), by synthesizing miniSINEUP with representative IRES sequences as EDs. Therefore, I generated miniSINEUP-IRES-DJ1 constructs, consisting of the IRES sequences from Polio Virus and c-myc ***(Figure 27A).*** All constructs were cloned into pCS2+ plasmid backbone. Transient transfection in HEK 293T cells led to a marked increase in DJ1 protein synthesis post-transcriptionally ***(Figure 27B and C),*** indicating that IRES sequences of both cellular and viral origins can be used in synthetic miniSINEUPs. These results were confirmed with similar SINEUP activity on another target gene, the GFP mRNA, when they were expressed from a different plasmid backbone, the pDual -peGFP plasmid ***(appendix Figure 39).***

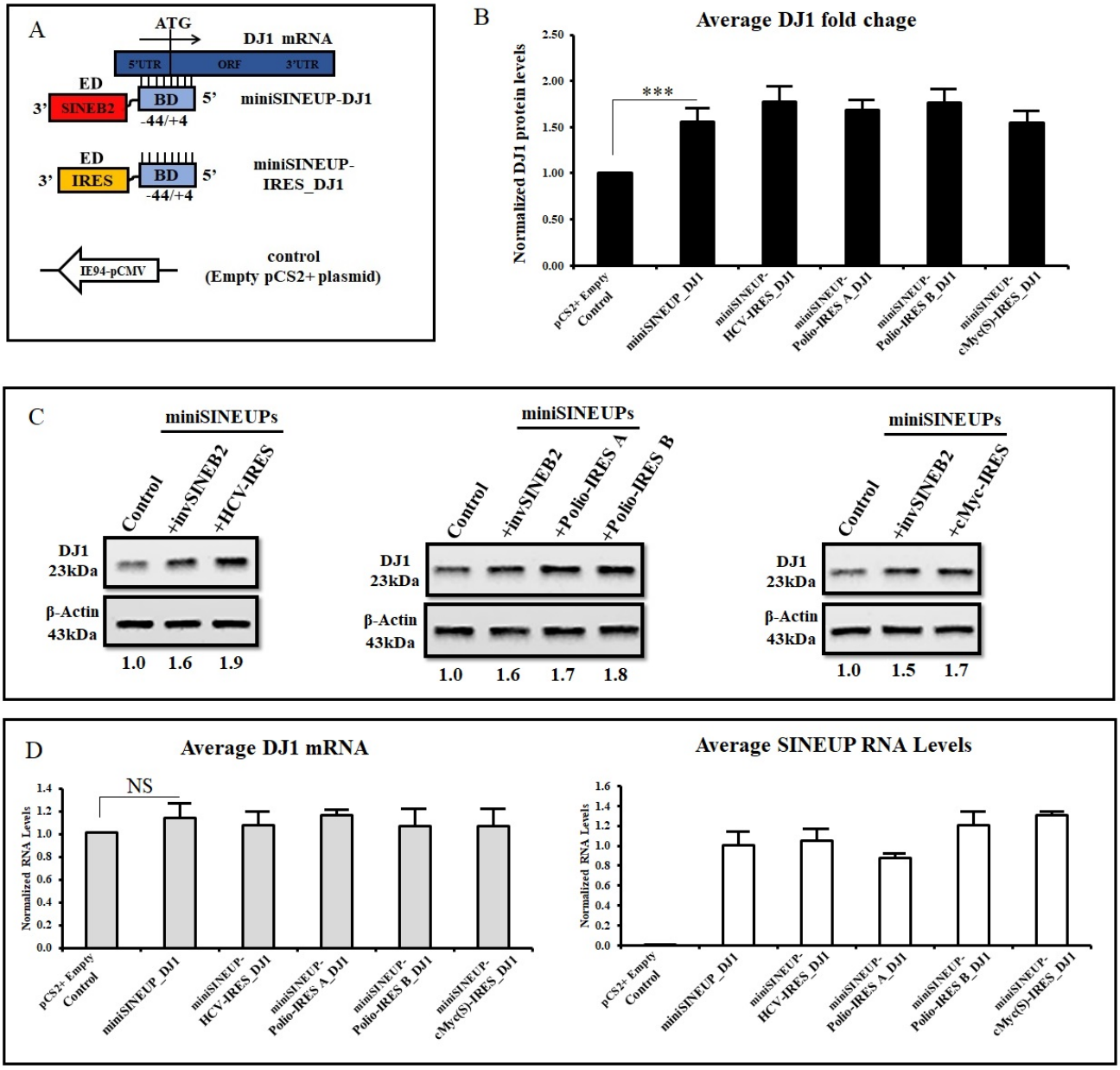


Figure 27 **IRES sequences can act as EDs in miniSINEUPs. (A)** Structural organization of miniSINEUPs to endogenous expressed DJ1 mRNA (blue scheme). Antisense to DJ1 mRNA (blue) at the 5'UTR is the 44nts BD (light blue) and EDs of AS Uchl1-invSINEB2 (red) for miniSINEUP-DJ1 and HCV-, Polio- and c-myc-IRES sequences (orange) for the miniSINEUP-IRES_DJ1 constructs, all cloned in pCS2+ plasmid backbone. MiniSINEUP-DJ1 and empty pCS2+ plasmids were used as experimental positive and negative controls respectively. (B and C) DJ1 protein levels were measured in HEK 293T cells 48hours post transient transfection. WB analysis as described before showed increase in DJ1 protein levels mediated by each of the tested IRES sequences. (B) Mean DJ1-protein intensity plots for each miniSINEUP-IRES activities that are quantified from WB analysis as normalized to fold change of no SINEUP, empty pCS2+ plasmid control. (C) Represented WB analysis showing quantified DJ1 proteins for each tested IRES sequence as compared to miniSINEUP-DJ1 with invSINEB2 ED. β-actin was used as normalized loading control. Here again, flipped polio IRES sequence has SINEUP activity that is comparable to direct IRES sequence. (D) RT-qPCR analysis as

described before showed stable DJ1 mRNA (grey bar, left), indicating post-transcriptional increase in DJ1 protein synthesis. Both invSINEB2 and each IRES EDs RNA (white bar, right) were detectable and measured by RT-qPCR with their respective primer pairs (appendix table 2). Data indicate mean ± SD of at least 5 independent biological replicas; $p > 0.05$ was considered not significant (NS) while $^{***}p < 0.001$ was considered significant.

4.3.3.3 HCV-IRES as ED in miniSINEUP-GFP increases targeted GFP protein synthesis in a pDual plasmid construct

HCV-IRES sequence was tested for its ability to function as ED in driving SINEUP activity of overexpressed GFP protein synthesis from a pDual plasmid. Similar to miniSINEUP-HCV-DJ1 construct, the 383nts HCV-IRES ED sequence separated by a 26nt linker sequence was attached to a 72nts BD targeting GFP mRNA, and then cloned into a pDual-peGFP plasmid backbone to produce miniSINEUP-HCV-IRES_GFP/peGFP-pDual plasmid construct ***(Figure 28A).*** Previously reported miniSINEUP-GFP/peGFP-pDual construct was used as a positive transfection control ***(Figure 20).*** miniSINEUP-HCV-IRES_GFP showed a statistically significant activity although smaller than miniSINEUP-GFP ($p < 0.05$). Interestingly, miniSINEUP-HCV-IRES_GFP was less expressed. It remains unclear whether lower expression is the cause of smaller activity.

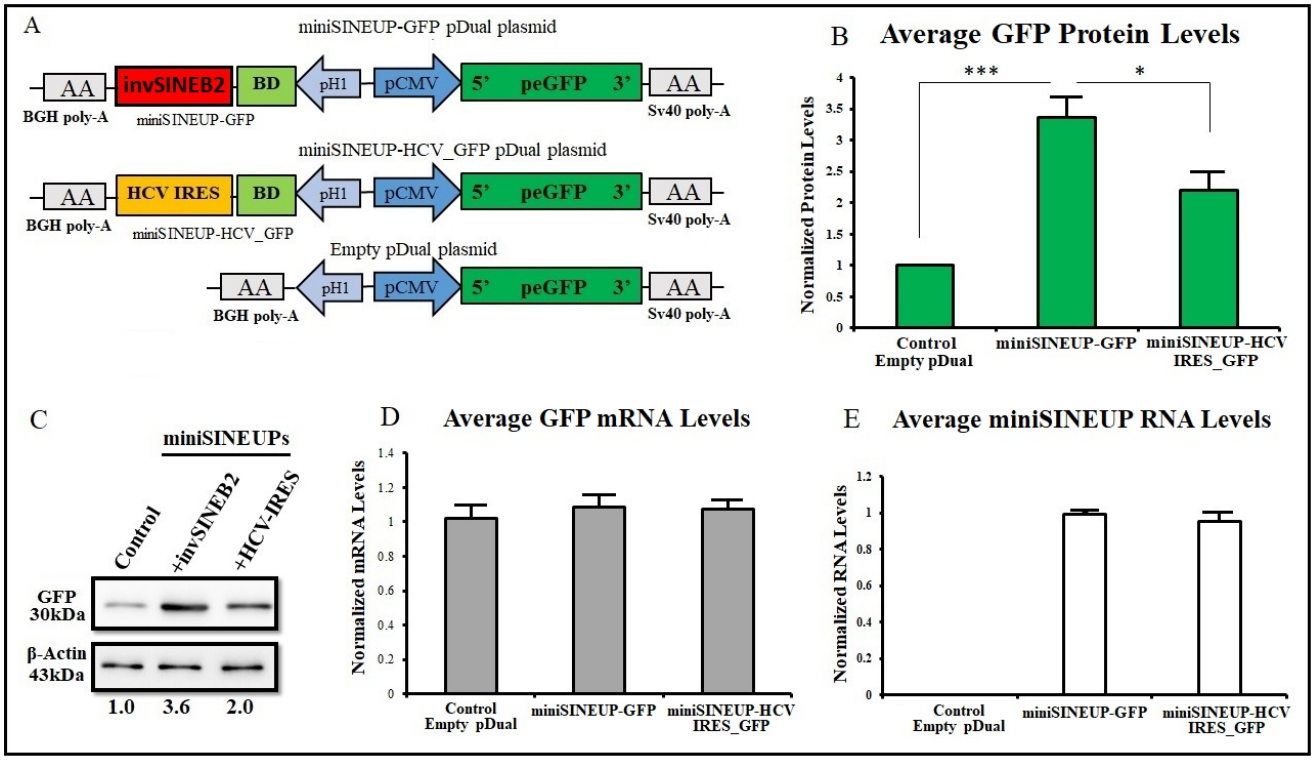


Figure 28 **HCV-IRES ED in a pDual plasmid construct increase targeted GFP protein synthesis in HEK 293T cells.** (A) Schematic representation of S/AS pDual-peGFP plasmid constructs showing respective invSINEB2 (red) and HCV-IRES (orange) sequences as sole SINEUP-EDs plus 72nt BD sequence (light green) that is antisense to GFP (green) at the 5'UTR and are under H1 and CMV promoters, respectively. The control *empty* pDual plasmid construct expresses only GFP mRNA but not the miniSINEUPs. SINEUP activity was performed by transient transfection of HEK 293T cells with the respective pDual plasmid constructs, WB (B and C) and qRT-PCR (D and C) analysis showed post-transcriptional GFP protein increase as seen on the WB image detected with anti-GFP antibody (C) and mean fold change intensity plot (green bars, B). WB quantification of GFP was normalized to β-Actin expression that was used as a loading control. qRT-PCR quantification is shown for GFP mRNA (grey bar, D) and miniSINEUP RNAs (white bar, E), respectively. The detected miniSINEUP-GFP and miniSINEUP-HCV-IRES_GFP RNAs were separated quantified with respective primer pairs as normalized by the empty pDual control. Data indicate mean ± SD of at least 5 independent biological replicas; $^{*}p < 0.05$ and $^{***}p < 0.001$ were considered significant.

4.3.4 Model III: IRES elements in natural transcripts has SINEUP-like activities on targeted protein synthesis in *trans*

From Model I and II, invSINEB2 TE and IRES elements complement each other in translational activation. Thus, viral and cellular IRES-containing transcripts increase translation of other expressed transcript targets in *trans*, though S/AS pairing and inter-transcript interactions. Using c-myc as a representative transcript containing an IRES sequence acting as ED in SINEUP, I bioinformatically

searched for stretches of sequence homology between the 5'UTR of c-myc and cellular mRNAs in antisense orientation. These can be potentially natural targets of SINEUP-like activity of natural c-myc mRNAs.

4.3.4.1 Ectopic expression of IRES-containing full-length c-myc cDNA mediates SINEUP-like function in increasing targeted proteins synthesis in *trans* in mammalian cell lines

Until now, it is known that IRES elements in both cellular and viral genes act as *cis*-regulatory elements to promote cap-independent translation of mRNAs in eukaryotic systems. Both viral and cellular IRESs allow internal recruitments of ribosomes to activate translation in both capped and uncapped mRNAs, especially under conditions that inhibit cap-dependent protein synthesis like in stress, G2/M phase of the cell-cycle or during cellular apoptosis to ensure the continual expression of key proteins needed for cell maintenance, growth and survival (Barrett et al., 2012; Komar and Hatzoglou, 2005). However, mRNAs of many growth factors, tumor suppressors, oncogenes, and many other proteins involved in stress-response and programmed cell-death regulations, which contains no known IRES elements or other CITEs, are also encoded in a cap-independent ways by unknown regulatory mechanisms to maintain their protein expression levels (Holcik and Sonenberg, 2005; Jackson et al., 2010; Liu and Qian, 2014; Sajjanar et al., 2017). Furthermore, global mapping of eukaryotic transcriptome reveals extensive RNA-RNA pairing, including coding mRNA-mRNA and ncRNA-mRNA pairings with unknown functions (Aw et al., 2016; Lu et al., 2016; Sharma et al., 2016). Previous reports have suggested possible inter-transcripts interactions and regulation in *trans* by cis-acting elements located at 3' untranslated regions (Macdonald et al. 2016; Reveal et al., 2010)

Our data suggests that IRES-containing transcripts might have evolved the ability to activate translation in *trans* increasing protein synthesis of other genes through S/AS mRNA-mRNA pairing. Therefore, I hypothesized that IRES-containing transcripts may have SINEUP-like activity in *trans* to partially overlapping targeted mRNAs. To test this hypothesis, I relied on two consecutive approaches: i. by the use of IRES-containing c-myc full-length transcript as representative mRNA, I bioinformatically searched for sequence pairing in antisense conformation to c-myc in 5'UTRs of protein coding mRNAs. Genes that exhibited partial pairing at the 3'UTR and CDS as well as ncRNA transcripts were discarded, since I have not yet established SINEUP activity for S/AS pairing at these mRNA regions. A list of potential targets is presented in Table 2. ii. the c-myc IRES-containing transcripts ability to post-

transcriptionally increase targeted protein synthesis were screened in mammalian cell lines *in vitro*. In brief, HEK 293T and U2OS cells were transiently transfected with pcDNA3.1- plasmid construct overexpressing full-length c-myc cDNA, using empty pcDNA3.1- plasmid transfection as control. Cells were harvested 48hours post-transfection, split in two portions; one-half for protein quantification and the other half for RNA quantification. Overexpression of full-length c-myc cDNA ***(Figure 30E)*** caused increased in targeted proteins levels post-transcriptionally in both U2OS cells ***(Figure 30B and C)*** and HEK 293T cells ***(Appendix Figure 40A and B)*** *in vitro*. Importantly, despite the fact that c-myc is a transcription factor, no transcriptional activation of these genes was revealed ***(Figure 29D).*** For JAG2 gene, which has been reported to be a transcription target of c-myc protein (Yustein et al., 2010), variable results on mRNA quantities were observed when qRTPCR was carried out for the long transcript variant. There is about 50% increase in JAG2 total mRNA quantification in HEK 293T cells ***(Appendix Figure 40C, left plot)*** but not in U2OS cells ***(Figure 29D),*** suggesting JAG2 can be both a transcriptional and translational target (Yustein et al., 2010) and translational target of c-myc.

Table 4 ***in Silico probing of c-myc IRES "SINEUP-like" activities target genes genomic localization and features as annotated in the NCBI database***

Annotated Gene	Gene Name	Gene pairing region	Transcript type	Position	Genomic location	Orientation	Alignment length	% Alignmnet	E-value	cMyc pairing region
ENST00000347004	JAG2	5'UTR	protein_coding	157-297	14:105168528-105168668	Reverse	144	65.97	0.014	IRES
ENST00000460865	RAB6B	5'UTR	protein_coding	240-311	3:133895526-133895597	Reverse	76	72.37	0.014	IRES
ENST00000433688	C11orf95	1st exon	protein_coding	120-183	11:63768496-63768559	Reverse	66	72.73	0.1	IRES
ENST00000256925	CABLES1	5'UTR	protein_coding	3760-3787	18:23259225-23259252	Reverse	28	92.86	0.087	IRES
ENST00000399816	UBE2QL1	5'UTR	protein_coding	184-201	5:6448806-6448823	Reverse	18	100	0.57	BD
ENST00000399816	UBE2QL1	5'UTR	protein_coding	166-205	5:6448806-6448824	Reverse	40	85	0.005	BD
ENST00000539361	MYRF	internal exon (nc)	retained_intron	1764-1785	11:61783950-61783971	Reverse	22	100	0.002	BD
ENST00000564062	ARIH1	1st coding exon	protein_coding	217-235	15:72474861-72474879	Reverse	19	100	0.14	BD
ENST00000575477	PAFAH1B1	1st exon (nc)	processed_transcript	87-178	17:2593901-2593992	Reverse	101	68.32	0.0002	BD
ENST00000397195	PAFAH1B1	5'UTR	protein_coding	156-247	17:2593901-2593992	Reverse	101	68.32	0.0002	BD
ENST00000393555	DYRK2	5'UTR	protein_coding	313-360	12:67649033-67649080	Reverse	48	79.17	0.034	BD
ENST00000468291	ZIC2	last exon (nc)	processed_transcript	378-455	13:99985415-99985492	Reverse	82	68.29	0.25	BD
ENST00000620342	ZIC2	last exon	protein_coding	1329-1407	13:99985415-99985492	Reverse	83	68.67	0.092	BD
ENST00000446285	ODC1	1st exon (nc)	processed_transcript	73-114	2:10448196-10448237	Reverse	43	79.07	0.67	BD
ENST00000234111	ODC1	5'UTR	protein_coding	268-309	2:10448196-10448237	Reverse	43	79.07	0.67	BD
ENST00000495910	NRF1	5'UTR	protein_coding	6056-6075	17:31252125-31252144	Reverse	20	100	0.1	BD
ENST00000220496	DNAJC17	5'UTR	protein_coding	1000-1020	15:40767866-40767886	Reverse	21	100	0.028	3'UTR
ENST00000309539	OLR1	5'UTR	protein_coding	1954-2063	12:10158700-10158809	Reverse	110	66.36	0.018	3'UTR
ENST00000533834	RNPC3	5'UTR	protein_coding	3260-3359	1:103552380-103552479	Reverse	106	66.04	0.018	3'UTR
ENST00000610980	IBTK	5'UTR	protein_coding	4749-4819	6:82170615-82170685	Reverse	72	73.61	0.018	3'UTR
ENST00000295213	SPATA18	5'UTR	protein_coding	3644-3702	4:52096540-52096598	Reverse	60	78.33	0.018	3'UTR
ENST00000398559	TMF1	5'UTR	protein_coding	3616-3730	3:69022946-69023060	Reverse	115	68.7	0.0001	3'UTR
ENST00000238497	VPS4B	5'UTR	protein_coding	1873-1995	18:63390519-63390641	Reverse	138	65.22	0.0003	3'UTR
ENST00000356126	PSMD12	5'UTR	protein_coding	2542-2626	17:67339696-67339780	Reverse	87	72.41	0.001	3'UTR
ENST00000526335	C11orf54	5'UTR	protein_coding	28-117	11:93746512-93746601	Forward	92	69.57	0.002	3'UTR
ENST00000357917	RHOBTB1	5'UTR	protein_coding	2535-2612	10:60871165-60871242	Reverse	82	70.73	0.002	3'UTR
ENST00000392519	TMEM121	5'UTR	protein_coding	620-696	14:105529290-105529366	Reverse	80	70	0.053	CDS
ENST00000325468	GYLTL1B	5'UTR	protein_coding	128-187	11:45922905-45922964	Reverse	60	76.67	0.053	CDS
ENST00000263754	KAT2B	5'UTR	protein_coding	168-282	3:20040239-20040354	Reverse	117	63.25	0.14	CDS
ENST00000543491	PCDH7	5'UTR	protein_coding	404-595	4:30721340-30721522	Reverse	198	61.62	0.38	CDS

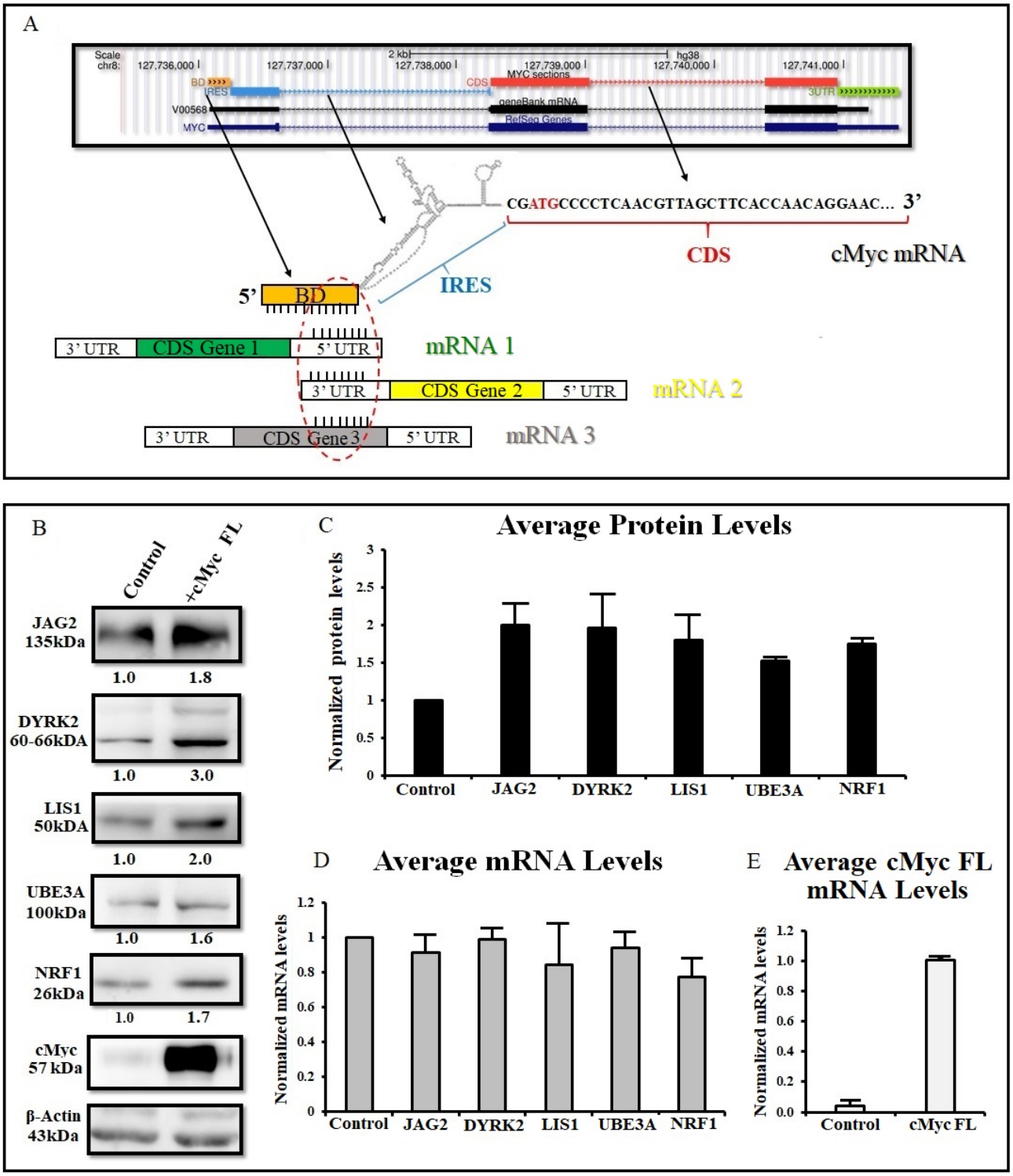


Figure 29 **IRES-containing full-length c-myc cDNA has SINEUP-like function in *trans* on several targets in U2OS cells. (A)** Schematics of human c-myc gene locus with natural transcript variants, showing upstream 5' UTR with IRES sequence (blue), coding sequence (CDS, red) and downstream 3' UTR sequence (green). Nucleotide numbering refers to UCSC-mm10 browser. Predicted at the TSS of the annotated c-myc mRNA reference sequence is the BD region (orange) preceding the IRES region at the 5' UTR. Arrows are pointed to a highlighted version of c-myc full length transcript used in our

studies, depicting the BD interacting to predicted targeted genes at various genomic loci, 5'UTR, CDS and 3'UTR. In this study, only genes with predicted 5'UTR pairing were verified experimentally. c-myc full-length transcript cDNA was gene synthesized, cloned into pcDNA3.1- plasmid and used to transiently transfect U2OS cells. Transfected cells were harvested 48hours and processed for detecting SINEUP-like activity on targeted genes. Control cells were transfected with empty pcDNA3.1- plasmid. (B) Western blotting of targeted JAG2, DYRK2, UBE3A, LIS1 and NRF1 proteins were detected using respective antibodies. The quantification was normalized to detected β-actin that was used as loading control. (C) Average targeted protein fold change quantification plot normalized to the empty plasmid control. The empty pcDNA3.1- plasmid control was arbitrarily set to 1. (D and E) Expression of targeted mRNA (grey bar, panel D) and c-myc full-length constructs RNA (white bars, panel E) were monitored by qRT-PCR with respective specific primers. All plots indicate mean ± standard deviation, and representation of N=5 independent replicates * $p < 0.05$, $^{**}p < 0.01$ were considered significant, NS; not significant

4.3.4.2 c-myc 5' UTR globally increase targeted proteins synthesis post-transcriptionally in cell lines.

c-myc protein is a transcription factor activating the expression of target genes. Our data ***(Figure 1)*** is suggesting a novel function for c-myc mRNA in increasing translation of targeted mRNAs, through partial mRNA-mRNA interactions and IRES activity in *trans*. To better characterized this novel role of c-myc mRNA and separate it from the protein function, I generated three transcript mutants: i. c-myc -deltaC, which lacks the DNA binding domain region in the encoded protein by deleting the c-terminal domain sequence and therefore the loss of transcriptional activity: ii. c-myc-5'UTR containing the 5'UTR sequence and therefore the predicted BD and IRES sequences, and iii. c-myc-IRES, the 5'UTR sequence containing only the IRES without the preceding BD region (Figure *30*A). Experiments were performed as described before.

Interestingly, I observed a statistically significant increase in the targeted proteins levels mediated by the c-myc 5'UTR sequence, which was consistent among different replicates. Similar results could be seen in both U2OS cells ***(Figure 30B)*** and HEK 293T cells ***(Appendix Figure 41A),*** confirming the previous data that the increase in targeted protein synthesis was post-transcriptional and at the level of translation, since the c-myc-deltaC and c-myc-5'UTR mutants could not elicit any transcriptional regulatory function ***(Figure 30C)***. Overall, additional experiments are needed to proof *in vivo* that c-myc mRNA and targeted mRNAs really interacts and/or are both associated together within the translating polysomes as a confirmation of the observed reported results in mammalian cells.

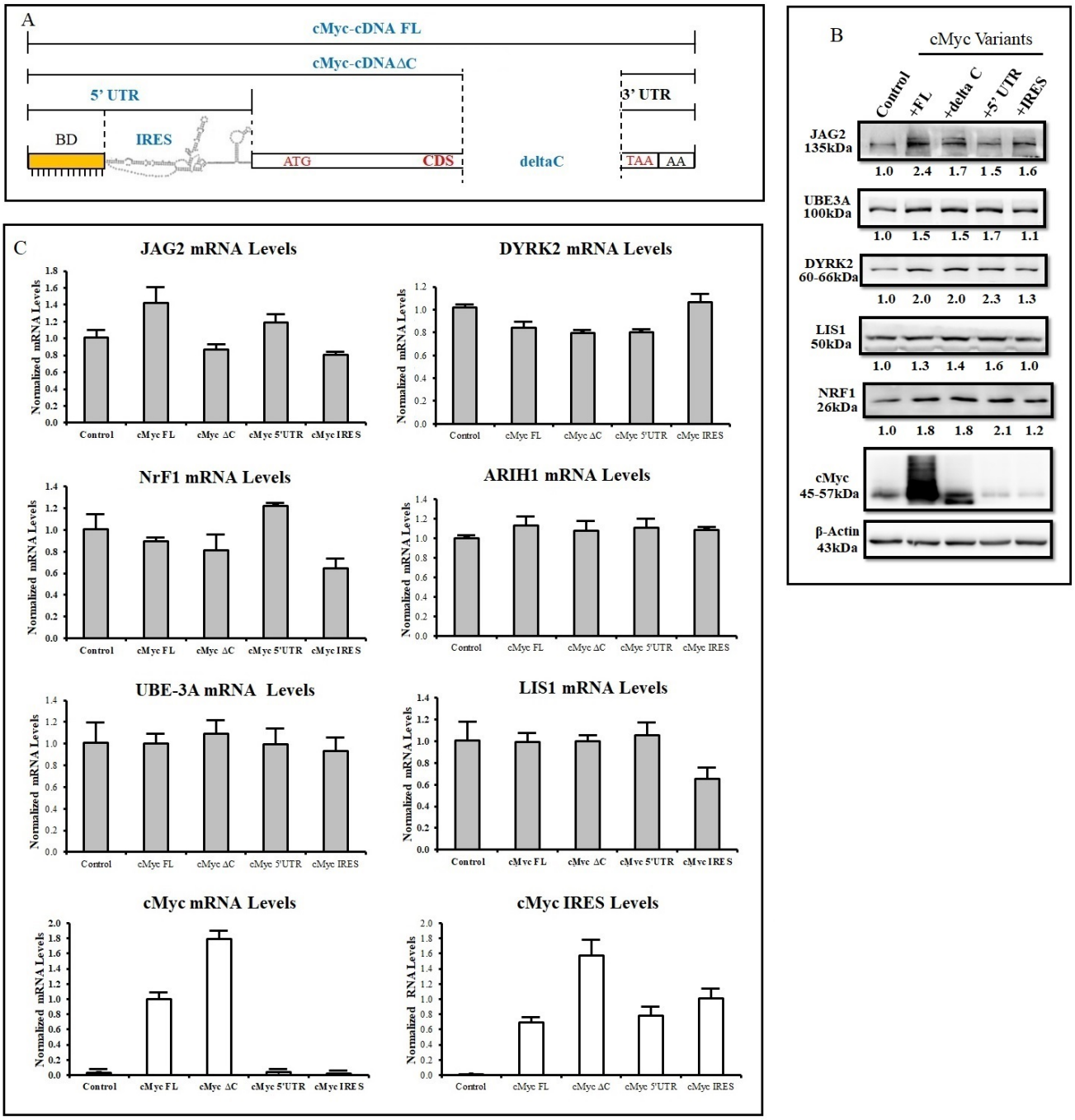


Figure 30 **c-myc 5' UTR has novel role in increasing other proteins synthesis post-transcriptionally in U2OS cell lines *in vitro*. (A)** Schematics of human c-myc full length and mutants cDNA. The mutants comprise of c-myc-ΔC, -IRES and -5'UTR sequences. The c-myc 5'UTR has the predicted BD preceding the IRES sequence while the c-mycΔC has the C-terminal domain deleted and hence the synthesized c-myc protein lost it transcription activation function. c-myc-IRES sequence is predicted to pair to JAG2, while all other targeted genes would pair at the BD. The c-myc full-length transcript and mutants cDNAs were gene synthesized, cloned into pcDNA3.1- plasmid and used to transiently transfect U2OS cells. Transfected cells were harvested after 48hours and processed for detecting SINEUP-like activity in trans on targeted genes. Control U2OS cells were transfected with empty pcDNA3.1- plasmid. (B) Western blotting of targeted JAG2, DYRK2, UBE3A, LIS1 and NRF1

were carried out using specific antibodies and bands were normalized to β-actin used as loading control. (C) Expression of targeted mRNAs (grey bar plots, upper panel C) and c-myc full-length mRNA and IRES RNAs (white bar plots, lower panel C) were monitored by qRT-PCR with respective specific primers. All plots indicate mean ± standard deviation, and representation of N=5 independent replicates, NS; not significant

4.4 Structural and functional commonalities between HCV-IRES and invSINEB2

4.4.1 SINEUP-HCV-IRES mutants disrupts HCV-IRES ability to increase targeted DJ1 protein synthesis in mammalian HEK 293T cell line in vitro

A long list of mutations haw been found in IRES sequences to partially or totally inhibit its activity. For HCV, there are known cases of HCV-IRES mutants from clinical isolates (Barría et al., 2009; Kieft et al., 1999) and synthetically induced cases (Kieft et al., 2001; Matsuda and Mauro, 2014; Wilson et al., 2000) that have detrimental defects on HCV-IRES wild-type (WT) activity and hence viral virulence. I have shown previously that HCV-IRES can function as ED in SINEUP-DJ1 and SINEUP-GFP to increase targeted DJ1 and GFP protein synthesis ***(Figure 26-Figure*** 28***)***. I also showed that the invSINEB2 element works as IRES in IRES-activity assay ***(Figure 23-Figure*** 25***)***. The mechanism of HCV-IRES WT function is known to be dependent on the coordination of modular domain (D) structures I-III **(*Figure 31*A)** with DIIId-GGG loop directly interacting with the 18S-rRNA to form a translation initiation complex that recruit eIF3 through DIIIb-c (Fraser and Doudna, 2007; Jubin et al., 2000; Lukavsky, 2009; Wilson et al., 2000). Based on the known details of HCV-IRES WT mechanism to activate translation without the cap-eIF4F complex, I was interested to investigate the activity of these mutations as ED in a SINEUP activity assay (Fraser and Doudna, 2007; Kieft et al., 1999; Matsuda and Mauro, 2014; Wilson et al., 2000).

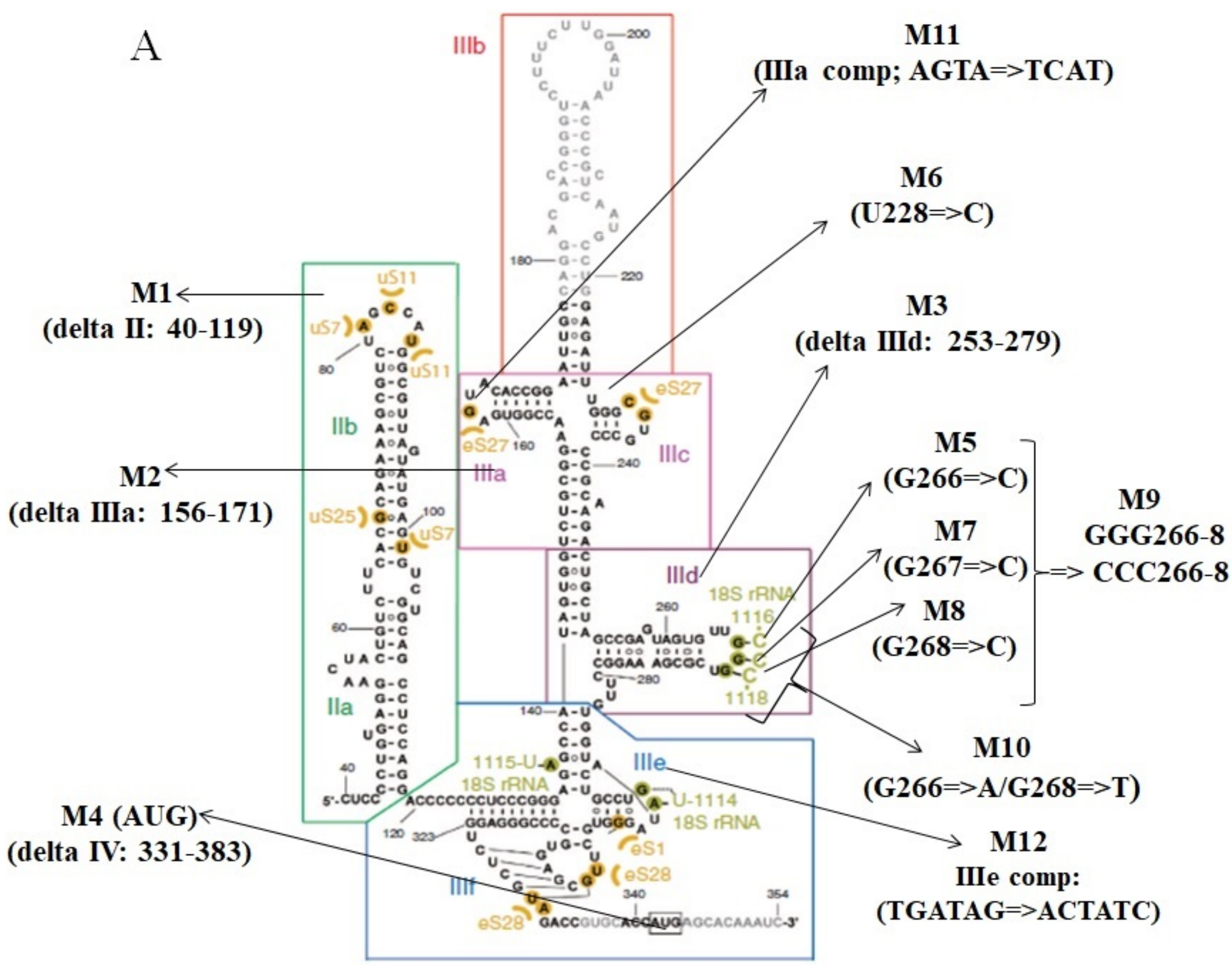
I therefore generated 12 known HCV-IRES mutants in SINEUP-HCV-IRES-DJ1 including deletions, reverse complementary and single to triple point mutations. The mutations were in the RPS25 (Landry et al., 2009), eIF3 and 40S ribosome binding sites in the HCV-IRES (WT) structural D sequences (Fraser and Doudna, 2007; Kieft et al., 1999; Matsuda and Mauro, 2014; Wilson et al., 2000) **(*Figure 31*A)**. Here, I report the effects of these SINEUP-HCV-IRES mutants (M1-M12) on DJ1 protein synthesis in HEK 293T cell lines *in vitro*, using SINEUP-HCV-IRES WT construct and empty pCS2+ plasmid as positive and negative controls respectively. Upon transient transfection of HEK 293T cells and

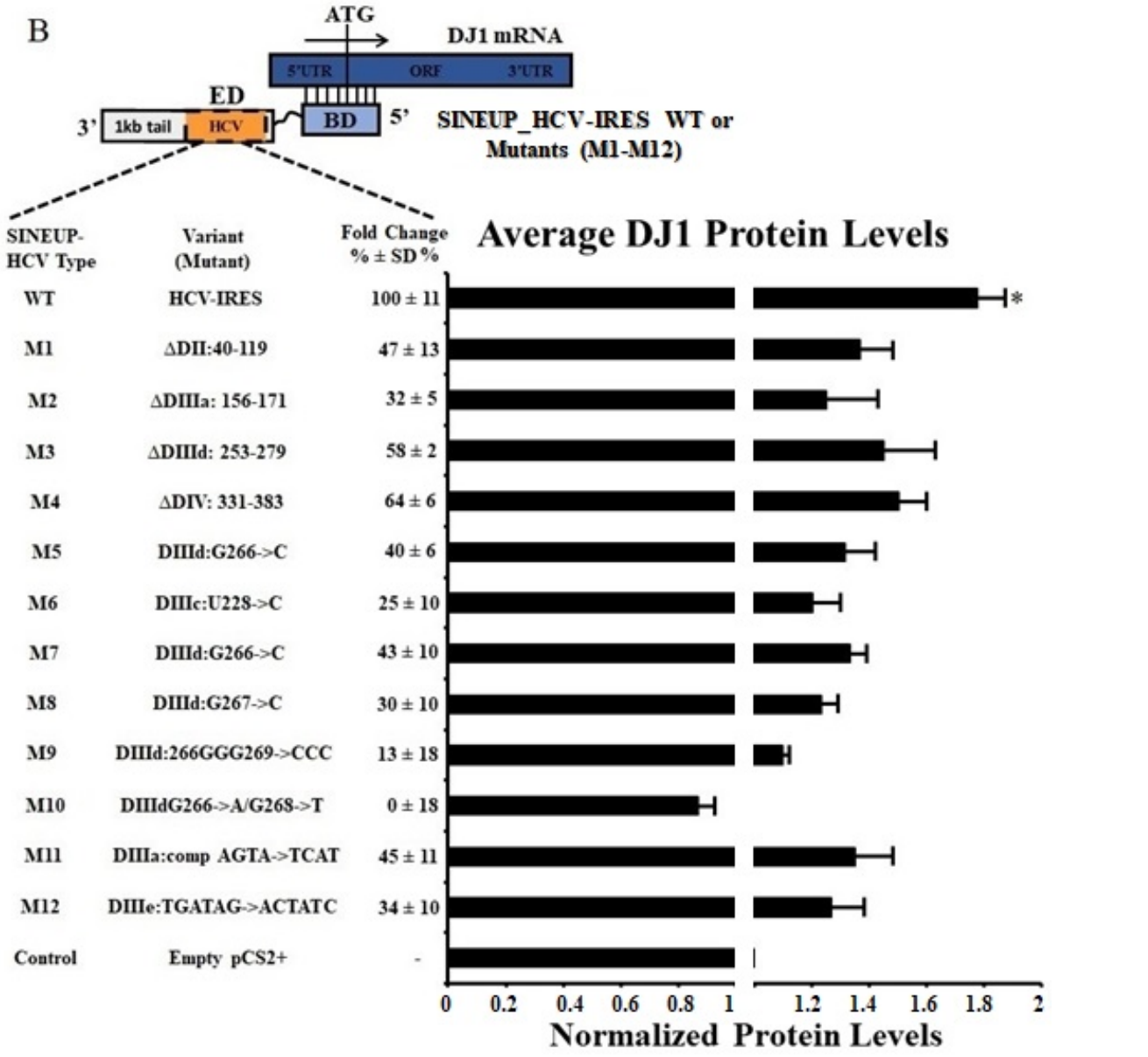
quantification of DJ1 protein after 48hours post-transfection, the SINEUP activity of the WT IRES was on average 1.8 fold compared to the empty pCS2+ plasmid control and arbitrary set at 100 ± SD % **(*Figure 31*B and C)**. Interestingly, mutations that were previously reported to disrupt function in IRES-activity assay (Kieft et al., 2001, Matsuda and Mauro et al., 2014) also disrupted their SINEUP activity. Notably, the G-to-C single point mutations; ^{266}CGG^{268}, ^{266}GCG^{268} and ^{266}GGC^{268} respectively for M5, M7 and M8 in the DIIId apical loop ***(Figure 31A)*** respectively has 41, 43 and 30% of the WT SINEUP activity, while the DIIId-^{266}CCC^{268} triple base M9 mutation killed completely its SINEUP activity ***(Figure 31B).*** Another double point mutation DIIId-^{266}AGT^{268} in the same apical loop, M10 **(*Figure 31*A)** that was reported from HCV-virus clinical isolates (Barría et al., 2009), also killed completely SINEUP activity ***(Figure 31B)***.

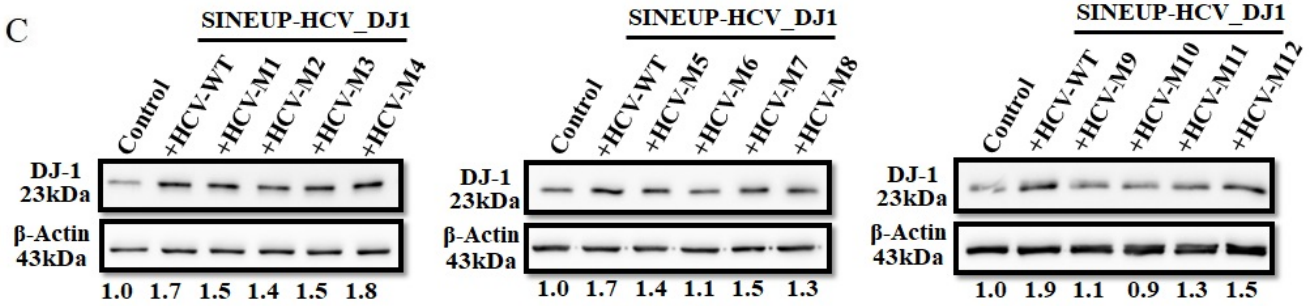
Interestingly, M3, that includes DIIId apical loop, showed 58% of the WT SINEUP activity. Most of deletion mutants (M1 of DII, M2 of DIIIa and M4 of DIIIf (Ji et al., 2004)) had residual 47, 43 and 65% of WT SINEUP activity, respectively ***(Figure 31)***. A single point U228C substitution mutation, M6 at the DIIIabc junction in the eIF3 interacting region, showed only 25% of WT SINEUP activity (figure 18A and B) ($p>0.05$). This mutation is another clinical isolate that was shown to affect IRES folding, eIF3 and 40S binding (Kieft et al., 1999, 2001; Matsuda and Mauro et al., 2014). Although, it does not disrupt 48S complex formation, it inhibits 80S ribosomes association (Ji et al., 2004). The observed 25% residual SINEUP activity may represent background errors due to the less sensitivity in using WB assay to detect protein expression as compared to the more sensitive IRES-activity assay.

I also studied reverse complementary mutations in the DIIIa apical loop, M11 mutation in an area that is suggested to be involved in binding eIF3, and M12 in DIIIe apical loop (Figure *31*A) of another 40S ribosome-pairing site (Kieft et al., 1999, 2001). Previously, both M11 and M12 mutations have reported just 10% WT activity (Kieft et al., 2001), using IRES-activity assay. However, I have observed 45 and 37% of WT activity, respectively (Figure *31*B), indicating that there was severe, but not complete impairment in the propensity of eIF3 and 40S ribosomal subunit bindings. In fact, our data on M2 and M11 are suggesting that DIIIa mutations alone are not completely deleterious to eIF3 binding since the observed 32 and 45% activities are in agreement with the indication by Kieft et al., 2001. For eIF3 binding, both IIIa and IIIb are required. The observed SINEUP activity of the WT and mutants are all post-transcriptional events since qRT-PCR results on DJ1 mRNA are statistically stable ***(Figure 31D, grey plots, upper panel)***. The WT and mutants SINEUP RNAs were also detectable by qRT-PCR ***(Figure 31D, white plots, lower panel)***.

Overall, I am reporting, using SINEUP activity assay, translational defects for the HCV-IRES deletion (M1, M2, M3 and M4), substitution (M5, M6, M7, M8, M9 and M10) and reverse complementary (M11 and M12) mutants on targeted DJ1 protein synthesis. Although, I observed less sensitivity in using WB to quantify proteins as a measure of SINEUP activity compared to dual luciferase reporter in IRES-activity assays, I have confirmed that mutations that disrupt or kill HCV-IRES WT function, can also affect SINEUP activity. This work show that mutations in known eIf3 binding sites and 40S ribosome internal entry sites are very detrimental to HCV-IRES in *trans* (Barría et al., 2009; Ji et al., 2004; Kieft et al., 1999, 2001; Matsuda and Mauro, 2014).




B
ATG
DJ1 mRNA
5'UTR
ORF
3'UTR
ED
3'
1kb tail
HCV
BD
5'
SINEUP_HCV-IRES WT or Mutants (M1-M12)
SINEUP-HCV Type
Variant (Mutant)
Fold Change % ± SD %
Average DJ1 Protein Levels
WT HCV-IRES 100 ± 11
M1 ΔDII:40-119 47 ± 13
M2 ΔDIIIa: 156-171 32 ± 5
M3 ΔDIIId: 253-279 58 ± 2
M4 ΔDIV: 331-383 64 ± 6
M5 DIIId:G266->C 40 ± 6
M6 DIIIc:U228->C 25 ± 10
M7 DIIId:G266->C 43 ± 10
M8 DIIId:G267->C 30 ± 10
M9 DIIId:266GGG269->CCC 13 ± 18
M10 DIIIdG266->A/G268->T 0 ± 18
M11 DIIIa:comp AGTA->TCAT 45 ± 11
M12 DIIIe:TGATAG->ACTATC 34 ± 10
Control Empty pCS2+ -
*
0 0.2 0.4 0.6 0.8 1 1.2 1.4 1.6 1.8 2
Normalized Protein Levels


C
SINEUP-HCV_DJ1
Control
+HCV-WT
+HCV-M1
+HCV-M2
+HCV-M3
+HCV-M4
DJ-1 23kDa
β-Actin 43kDa
1.0 1.7 1.5 1.4 1.5 1.8
SINEUP-HCV_DJ1
Control
+HCV-WT
+HCV-M5
+HCV-M6
+HCV-M7
+HCV-M8
DJ-1 23kDa
β-Actin 43kDa
1.0 1.7 1.4 1.1 1.5 1.3
SINEUP-HCV_DJ1
Control
+HCV-WT
+HCV-M9
+HCV-M10
+HCV-M11
+HCV-M12
DJ-1 23kDa
β-Actin 43kDa
1.0 1.9 1.1 0.9 1.3 1.5

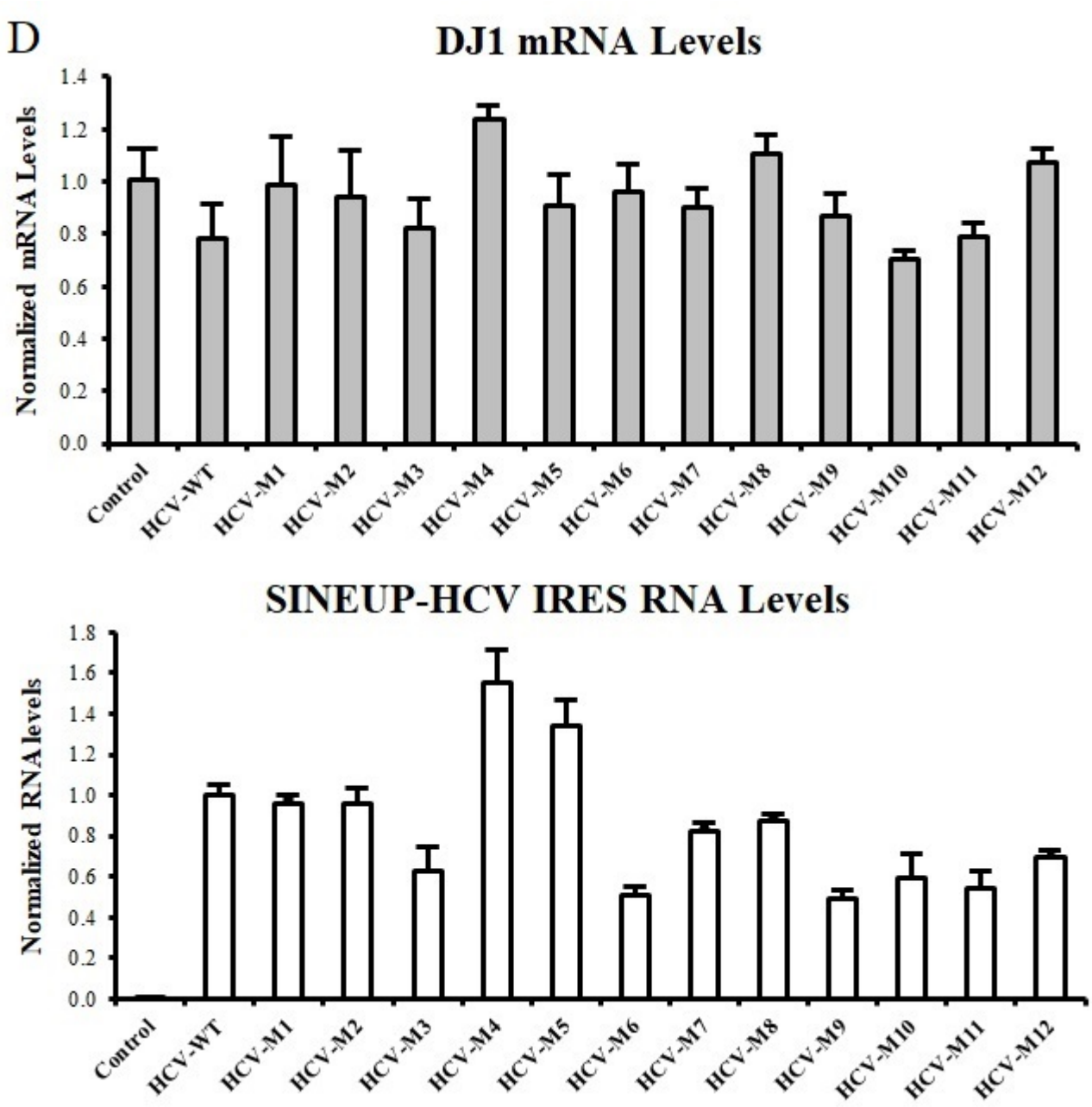


Figure 31 **HCV-IRES mutants disrupt WT SINEUP activity function in HEK 293T cells. (A)** Schematic representation of HCV IRES secondary structure, showing in details with black arrows the deletion mutation M1-M4, substitution mutations M5-M10 and complementation mutations M11-M12 within individual structural domains, DII-III, defined with enclosed lines (modified from Quade et al., 2015). (B and C) SINEUP-HCV activity assays were performed with transfection with the respective plasmid constructs; showing scheme of SINEUP 3' tail and inserted EDs of HCV-IRES WT and mutants (M1-12, orange) with 44nt BD (light blue) that is antisense to the 5'UTR of DJ1 mRNA (blue). The individual WT and mutant constructs were gene synthesized and expressed from pCS2+ plasmid. The WT plasmid construct and empty pCS2+ plasmid was used as transfection positive and negative controls respectively. Transient transfection in HEK 293T cells and WB analysis after 48hours with anti-DJ antibody showed the WT ability to increase DJ1 protein synthesis disruption to variable degrees by the M1-12 mutants. SINEUP activity of WT and each mutants are plotted as mean fold-

induction intensities of at least 5 biological replicas as normalized from empty pCS2+ transfection with it value arbitrarily set at 1. Each HCV-IRES mutant activity relative to the WT activity was calculated and shown as percentage DJ1 protein fold induction in the respective plots. the WT activity was arbitrarily set at 100 ± SD%. (C) Represented individual WB quantified images are shown. The quantification was normalized to detected β-Actin that was used as loading control. (D) As a control check for translational and not transcriptional phenomenon, RT-qPCR was performed to quantify and compare relative to the empty pCS2+ transfection control, the DJ1 mRNA (grey plot, upper panel) quantities, which is stable statistically. The SINEUP-HCV WT and mutants, M1-12 RNAs could also be detected and quantified (light plot, lower panel). Error bars show SD from at least four-independent replicates. An asterisk indicates statistical significance between empty pCS2+ transfection control and WT activity; * $p < 0.05$.

4.4.2 Mutations in invSINEB2 SL1 hairpin structure inhibit translational activities

Cis-acting regulatory elements have been implicated to control translation initiation that is considered to be the rate-limiting step in any protein synthesis. Protein synthesis by the ribosome takes place at a sequence-dependent manner, which could be lowered by RNA secondary structural motifs such as junctions, hairpins and pseudoknots (Chen et al., 2016). Hairpin, junction and pseudoknot structures play regulatory roles in IRES-dependent translation in eukaryotes, since their interaction with the ribosome causes conformational dynamics that controls the ability of the ribosome to bind to mRNAs, influencing protein synthesis (Chen et al., 2016). These structures also bind different proteins that either enhance or repress translation of the mRNA targets (Barrett et al., 2012).

As previously reported in this work, IRES elements work as invSINEB2 and vice versa. These findings together with previous works (Carrieri et al., 2012, Zucchelli et al., 2015, Podbevsek et al., 2018) have implicated invSINEB2 as the main component driving activation of translation and hence SINEUP activity on targeted proteins synthesis. To assess how invSINEB2 regulates translation activation, the secondary structure of invSINEB2 of AS Uchl1 was first probed, using chemical footprinting and reported in the work of (Podbevšek et al., 2018). The structure of invSINEB2 exhibits high GC content with a stable SL1 hairpin structure with it loop regions exhibiting dynamic properties of reactivity with external chemical agents. This together with functional studies confirmed the terminal SL1 hairpin structure as the main structural determinant required for SINEUP of AS Uchl1 function (Podbevšek et al., 2018). Taken together, this suggests that invSINEB2 through the dynamic SL1 hairpin structure potentially act as RNA scaffolds that binds to the 40S ribosome to internally recruit the ribosomal initiation complexes for the synthesis of targeted proteins, utilizing more specifically HCV-IRES-like mechanism.

HCV-IRES also form stable domain structures, which are overall not globular, but dynamic with helical junctions and loops that specifically recruit the translation machinery for IRES-activity (Kieft et al., 199, 2001). I have reported in this work that triple ^{266}GGG^{268} to ^{266}CCC^{268} mutation in the DIIId loop of the HCV-IRES WT (M9 mutant) disrupts SINEUP activity when substituted for invSINEB2 ED in SINEUP activity assay ***(Figure 31)***. This is similar to the disruption of HCV IRES-activity in a bicistronic reporter assay. The HCV-IRES DIIId ^{266}GGG^{268} directly binds to the 18S-rRNA ^{1115}AUUCCCA^{1120} helix (h) 26 of the 40S ribosome (Kieft et al., 1999, Mauro and Matsuda, 2014). From sequence comparisons, similar nucleotides could be seen conserved between invSINEB2 SL1 hairpin loop of AS Uchl1 and HCV DIIId loop structures **(*Figure 32*A).** Thus, the SL1 hairpin structure has from positions ^{74}GUUGUG^{79} nucleotides similar to HCV-IRES DIIId loop at positions ^{253}GUUGGG^{268} with only the respective U^{78} and G^{267} nucleotides mismatch (***Figure 32****A*). Interestingly, not as strong as the G nucleotide, the U nucleotide can basically form non-Watson-Crick base pair with C nucleotide (Varani and McClain, 2000, Leontis et al., 2002), suggesting potential similarities in function between invSINEB2 SL1 hairpin-^{77}GUG^{79} and in HCV-IRES DIIId loop at positions ^{266}GGG^{268} in recruiting the 40S ribosome.

I therefore hypothesized that invSINEB2 through SL1 hairpin -^{77}GUG^{79} structure recruits the 40S ribosome to activate translation of targeted mRNAs, using HCV-IRES-like mechanism. For testing this experimentally, various point mutations (C3 and C4) within invSINEB2 SL1 hairpin-^{77}GUG^{79} were made. I also designed chimera mutants (C1 and C2) generated by swapping SL1 hairpin for HCV-IRES DIIId of ^{266}GGG^{268} WT and ^{266}CCC^{268} M9 mutant as shown in figure 19A. Conversely, I substituted HCV-IRES DIIId for SL1 hairpin -GUG WT, -GGG and -CCC to generate chimera mutants C5, C6 and C7 respectively (figure 19A). SINEUP constructs containing invSINEB2 EDs; SL1 WT and mutants C1-C4 as well as HCV-IRES EDs with DIIId WT and mutants C5-C7 were synthesized and cloned into pCS2+ plasmid. Experiments were performed using as positive control synthetic SINEUP-DJ1 (invSINEB2 SL1-^{77}GUG^{79} hairpin) and here called the SL1-^{77}GUG^{79} WT and empty pCS2+ plasmid as transfection negative controls ***(Figure 1 and Figure*** 26***).***

Here the measured SINEUP activity was on average 1.7 fold and arbitrary set at 100 ± SD%. As shown in figure 18B plots, the chimera mutant C1 and C2 (that has invSINEB2 SL1 hairpin substituted for HCV-IRES DIIId ^{266}GGG^{268} WT and mutant ^{266}CCC^{268} M9) resulted in 75% and 63% respectively of the SL1-^{77}GUG^{79} WT activity, a result which is not statistically significant ($p > 0.05$) when compared between these two mutants. However, comparing both chimera mutant C1 and C2 activities against the SL1-^{77}GUG^{79} WT, there are statistical significant differences in activities with $p<0.05$ and $p<0.01$ values,

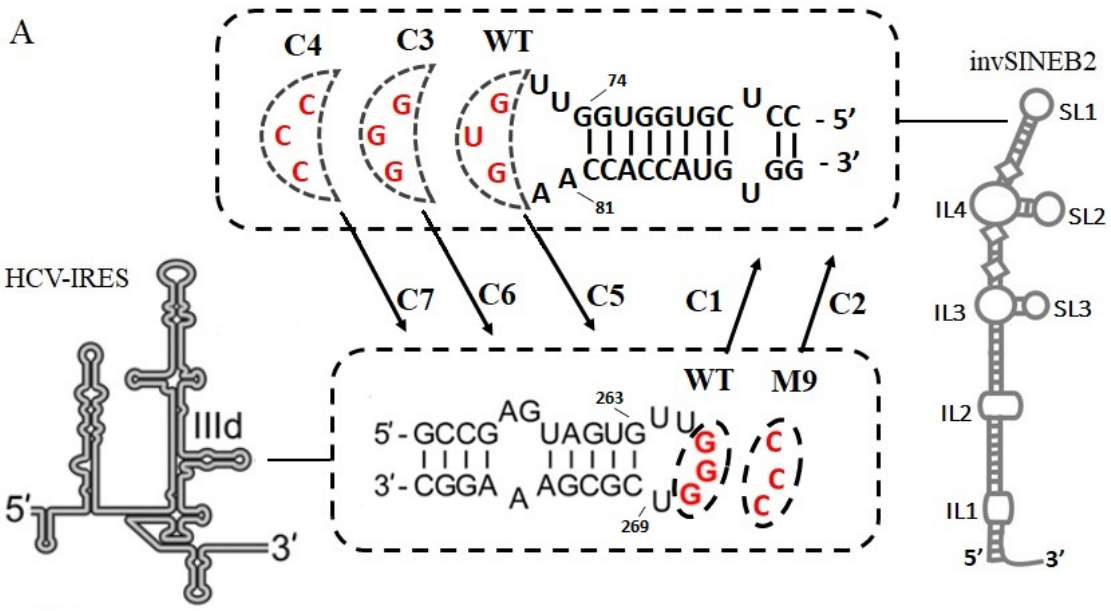
respectively. The lost in 25% of SL1-^{77}GUG^{79} WT activity by mutant C1 could suggest that the entire SL1 hairpin structure is important for driving invSINEB2 mediated SINEUP activity. This suggestion is supported by the fact that the mutant C2, which has HCV-IRES DIIId-^{266}CCC^{268} M9 retains 63% of the WT function, a loss of only 37% from SL1-^{77}GUG^{79} WT activity. However, these results suggest similarities in recruiting 40S ribosome between SL1 hairpin-^{77}GUG^{79} WT and HCV-IRES DIIId-^{266}GGG^{268} WT.

Similar to SL1 chimera mutants C1 and C2, HCV-IRES chimera mutants C5 and C7, which respectively have HCV-IRES DIIId substituted for invSINEB2 SL1 hairpin –^{77}GUG^{79} WT and –^{77}CCC^{79} mutants retained 82% and 63% of SL1-^{77}GUG^{79} WT activity (*Figure 32*B). Exactly as in C2, there was 37% loss of WT activity in mutant C7, indicating translation inhibition by all chimera mutants containing -CCC hairpin loops, suggesting a strong importance of the -^{77}GUG^{79} WT loop in mediating SINEUP activity. The similarities in results between mutants C1 and C5 that have substituted WT structures or C2 and C7 with mutant structures ***(Figure 32A and B)*** are further proof that the entire SL1 hairpin–^{77}GUG^{79} WT and HCV-IRES D-IIId WT structures are needed for maintaining the respective invSINEB2 and HCV-IRES mediated SINEUP activity. Interestingly, chimera mutant C6 that has HCV-IRES DIIId replaced with SL1 hairpin-GGG loop structure retained 95% of SL1-^{77}GUG^{79} WT activity. Although not statistically significant, this result is 9% higher than the HCV-IRES DIII-GGG WT activity **(*Figure 32*B).** Altogether, the result indicates the strong dependence on -^{266}GGG^{268} triplet nucleotide by HCV-IRES DIIId structure in mediated SINEUP activity. This may account for the reason why between chimera mutants C5 and C6, G to U nucleotide substitution in C5 resulted in 13% less activity as compared to C6 (Figure *32*B), suggesting that canonical U^{78} in SL1- ^{77}GUG^{79} WT structure strongly influence how the SL1 hairpin structure interacts with the 40S ribosome to mediate SINEUP activity.

It seems from the chimera mutants that the conserved SL1 hairpin-GUG WT and the DIIId-GGG WT structures are respectively needed to maintain the full invSINEB2 and HCV-IRES mediated activities. To characterize better the importance of SL1 hairpin -^{77}GUG^{79} structure I generated two constructs: i. C3, having SL1 hairpin -^{77}GGG^{79} structure in which the U^{78} was substituted with G to depict the HCV-IRES DIIId ^{266}GGG^{268} WT; ii. the C4 with the -^{77}CCC^{79} mutation to inhibit binding to the conserved 18S-rRNA ^{1116}CCC^{1118} h26, as additional control (Figure *32*A). Upon transient transfection in HEK 293T cells, SINEUP activities were 80 and 51% for C3 and C4, respectively. The loss by the C4 mutant of 49% of WT activity, strongly confirms the model that the SL1 hairpin-^{77}GUG^{79} structure interacts with the 40S ribosome in a similar fashion as used by HCV-IRES DIIId-GGG. In addition, other potential cis-

regulatory elements or structures within the invSINEB2 secondary structure or surrounding sequences in Δ5'-AS Uchl1 backbone might also play roles in the 40S ribosome recruitment **(*Figure 22* and**

***Figure 36*).** Taken together, the mutant data suggest that the terminal SL1-^{77}GUG^{79} WT of invSINEB2 use HCV-IRES like pathways to bind the ribosome in mediating SINEUP activity on targeted protein synthesis and the conserved ^{78}U probably has a crucial role to play during interactions.



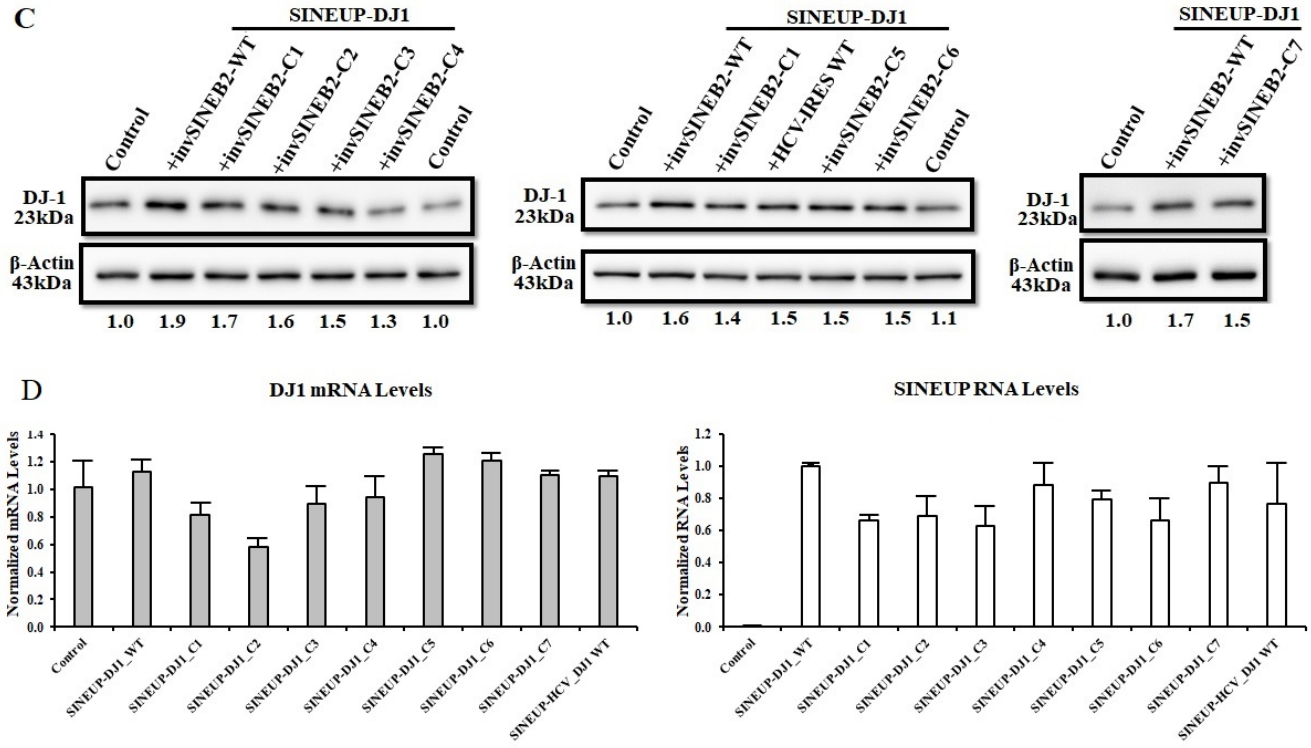


Figure 32 M**utations in invSINEB2 SL1 hairpin structure disrupt WT SINEUP activity in HEK 293T cells. (A)** Schematic representation of HCV IRES (left, modified from Mauro and Matsuda, 2014) and invSINEB2 of AS Uchl1 (right) secondary structures. Showing in details with dotted box are the sequence of SL1 hairpin structures of invSINEB2 and DIIId structures of HCV-IRES. The WT and various mutations at the loop regions are indicated in red. Just like HCV-IRES WT triplet -^{77}GUG^{79} nucleotides, SL1 hairpin triplet -^{77}GUG^{79} was hypothesized to bind to 18S rRNA helix 26. Black arrows indicate chimera mutants, which were generated by swapping invSINEB2 SL1 hairpin WT or mutants structures with HCV-IRES DIIId WT and mutant structures. Chimera mutants C1 and C2 were generated by substituting SL1 hairpin for HCV-IRES DIIId-^{266}GGG^{268} WT and -^{266}CCC^{268} M9 sequences. C3 and C4 were generated by substituting the triplet SL1-^{77}GUG^{79} for -^{77}GGG^{79} and -^{77}CCC^{79} nucleotides, respectively to depict the HCV-IRES WT and mutant M9 -DIIId loops. SINEUP chimera mutants C5, C6 and C7 were generated by substituting HCV-IRES DIIId for SL1 hairpin-^{77}GUG^{79} WT, and mutant -^{77}GGG^{79} and -^{77}CCC^{79} structures respectively as shown by the arrow. SINEUP activity assay was performed by transfecting HEK 293T cells with the respective plasmid constructs; showing scheme of SINEUP 3' tail and inserted EDs of invSINEB2 WT and -chimera mutants (C1-C7, red) with 44nt BD (light blue) that is antisense to DJ1 mRNA at the 5'UTR (blue). The individual WT and mutant constructs were gene synthesized and expressed from pCS2+ plasmid backbone. The SINEUP-DJ1 with invSINEB2 SL1 hairpin-^{77}GUG^{79} WT plasmid construct and empty pCS2+ plasmid was used as transfection positive and negative controls respectively. Transient transfection in HEK 293T cells and WB analysis after 48hours with anti-DJ antibody showed DJ1 protein synthesis disruption to variable degrees by the mutants. SINEUP activity of WT and each mutant are plotted as mean fold-induction intensities ± SD of at least 5 biological replicas as normalized from empty pCS2+ transfection with it value arbitrarily set at 1. Each chimera mutant activity relative to the WT activity was calculated and shown as percentage fold change. The WT activity was arbitrarily set at (100 ± SD)%. (C) Represented individual WB quantified images are shown. The quantification was normalized to β-Actin that was used as loading control. (D) As a control

check for post-transcriptional phenomenon, RT-qPCR was performed to quantify and compare relative to the empty pCS2+ transfection control, the DJ1 mRNA (grey plot, left panel) quantities, which is stable statistically. The SINEUP RNAs could also be detected and quantified (white plot, right panel). Error bars show SD from at least five-independent replicates. An asterisk indicates statistical significance between empty pCS2+ transfection control and WT activity; * $p < 0.05$.

4.4.3 invSINEB2 SL1 hairpin mutants inhibit cap-independent Fluc protein synthesis

Structurally, sequence similarities between invSINEB2 SL1 hairpin-^{74}GUUGUG^{79} structure and HCV-IRES DIIId-^{263}GUUGGG^{268} made us hypothesized that invSINEB2 through SL1 hairpin -^{77}GUG^{79} structure recruits the 40S ribosome to activate translation of targeted mRNAs, using a HCV-IRES-like mechanism. Using SINEUP activity assay, I showed that chimera and substitution mutations involving both structures disrupted WT SINEUP activity ***(Figure 32),*** providing evidence that the SL1 hairpin structure potentially interacts at the same 18S rRNA h26 site of the 40S ribosome as the HCV-IRES DIIId structure in mediating protein synthesis, and the conserved ^{78}U probably has a crucial role to play during interactions.

As a further proof of principle for the importance of SL1 hairpin structure, here, I tested in IRES-activity assay some of the mutants that were used in the SINEUP activity assay using HEK 293T and U2OS cell lines. This is very important because the construct used in this part of the experiment is a short SINEUP ED, containing the conserved invSINEB2 TE. I tested two chimera mutants named Luc-IRES C1 and C2, which has HCV-IRES DIIId-^{266}GGG^{268} WT and -^{266}CCC^{268} mutant structures respectively substituted for the SL1 hairpin structure-^{74}GUUGUG^{79} WT in invSINB2-spacer sequence to generate a pRUF construct similar to the SINEUP chimera mutants C1 and C2 invSINEB2 EDs used in SINEUP activity assay ***(Figure 32A).*** The other mutants include two substitution mutants, Luc-IRES C3 and C4 that has SL1 hairpin -^{76}GUG^{78} substituted for -^{76}GGG^{78} and -^{76}CCC^{78} triplet nucleotides respectively, depicting the chimera mutants C3 and C4 invSINEB2 EDs, respectively used in SINEUP activity assay ***(Figure 32A).*** The measure of IRES-activity showed similar pattern of WT activity disruption as reported in HEK 293T, using SINEUP activity assay ***(Figure 32B).***

As it should be expected, the Luc-IRES C1 retained 77% of WT IRES-activity ***(Figure 33),*** which is similar to 75% of WT activity that was observed in SINEUP activity assay ***(Figure 32B).*** This was similar for C3, which retained 82 and 80% of WT activity in SINEUP- and IRES-activity assays,

respectively ***(Figure 32B Figure 33).*** There was just 13% activity difference between the C2 mutant in SINEUP- and IRES-activity assays ***(Figure 32B Figure 33).*** The loss of 23% and 25% of WT IRES-activity by the Luc-IRES C1 and C2 mutants respectively confirmed our SINEUP activity results for chimera mutants C1 and C2 ***(Figure 32B).*** The similarities in the Luc-IRES chimera C1 and C2 mutants IRES-activities could also confirm our previous suggestion from SINEUP activity results for C1 and C2 mutants that structural rearrangement possibilities might have led to new structures with novel interactions that might not involve DIIId- ^{263}GUUGGG^{268}and -^{263}GUUCCC^{268}. I could hence conclude that the substituted HCV-IRES DIIId-^{263}GUUGGG^{268} WT could not recapitulate SL1 hairpin-^{74}GUUGUG^{79} mediated activities, likewise the substituted DIIId-^{263}GUUCCC^{268} mutant could not kill SL1 hairpin-^{74}GUUGUG^{79} activities.

The IRES-activity of the substitution mutants, Luc-IRES C3 and C4 were 82 and 65% respectively of the invSINEB2 SL1 hairpin GUG WT activity ***(Figure 33).*** Here again, the obtained C3 mutant IRES-activity confirmed the SINEUP activity result (figure 19B and 20A). However, there was only 35% loss of WT activity in IRES-activity assay for Luc-IRES C4 as compared to 49% loss in SINEUP activity assay for the same corresponding chimera mutant C4. The 15% difference may partly be attributed to technicalities that is due to differences in the two assays. However, the fact that the results in both assays were the same for Chimera mutants C1 and C3 with WT loop-^{74}GUUGGG^{79} but not chimera mutants C2 and C4 with mutant loops -^{74}GUUCCC^{79} could suggest something more biological, which could be influence from surrounding sequences around the invSINEB2 TE that differs in both assay. However, the retention of 65% of WT IRES-activity by the Luc-IRES C4 mutant means that the nucleotides ^{76}U and ^{80}AA^{81} in the SL1 hairpin ^{75}UUCCCAA^{81} are probably interacting with nucleotides A^{1119} and ^{1115}UU^{1116} respectively in the 18S-rRNA ^{1114}UUCCCA^{1119} h26 through alternative interactions involving reverse complementary base-pairing. This interaction may probably be weaker as compared to the WT with SL1 hairpin-GUG loop, leading to the observed 65% and 51% activities in IRES- and SINEUP activity assays, respectively. Additionally, as already mentioned, it is likely that other conserved cis-regulatory elements; TOP, m^{6}A, and uORFs sequences that are located inside or around the invSINEB2 structure **(*Figure 22* and**

***Figure 36*)** might also contribute to the 40S ribosome and probably key translation initiation factors recruitment. Therefore, just like most IRES elements, invSINEB2 requires some trans-acting factors for internal ribosome recruitment and function.

Comparing U2OS cells result with that of HEK cells, the measured IRES-activity was relatively higher ***(appendix Figure 42)***. However, similar patterns of disruption of WT IRES-activity by the Luc-IRES C1-C4 mutants were observed in U2OS cell lines. Noticeably, there was 30% loss in the WT activity by the Luc-IRES C4 mutant in U2OS cells, while 35% in HEK cells, which again confirmed the retention of some IRES-activity in the mutants. As shown in U2OS cell lines by qRT-PCR ***(appendix Figure 42)***, the observed IRES-activity in both cell lines was neither cryptic promoter activities nor alternative splicing events that could be mediated by the inserted sequences. Taken together, the data strongly suggest that the entire SL1 hairpin structure might be required for invSINEB2 mediated activities. The similarities in the WT -GUG or -GGG and -CCC mutants suggest structural rearrangement possibilities and hence maintenance of some function, though lower than the WT. Some other conserved loops in the entire structure and cis-acting elements predicted in the structure might play roles in recruiting other protein factors including ITAFs. I could hence conclude that the substituted HCV-IRES DIIId-^{263}GUUGGG^{268} WT could not recapitulate SL1 hairpin-^{74}GUUGUG^{79} mediated activities, likewise the substituted DIIId-^{263}GUUCCC^{268} mutant could not kill SL1 hairpin-^{74}GUUGUG^{79} WT activities.

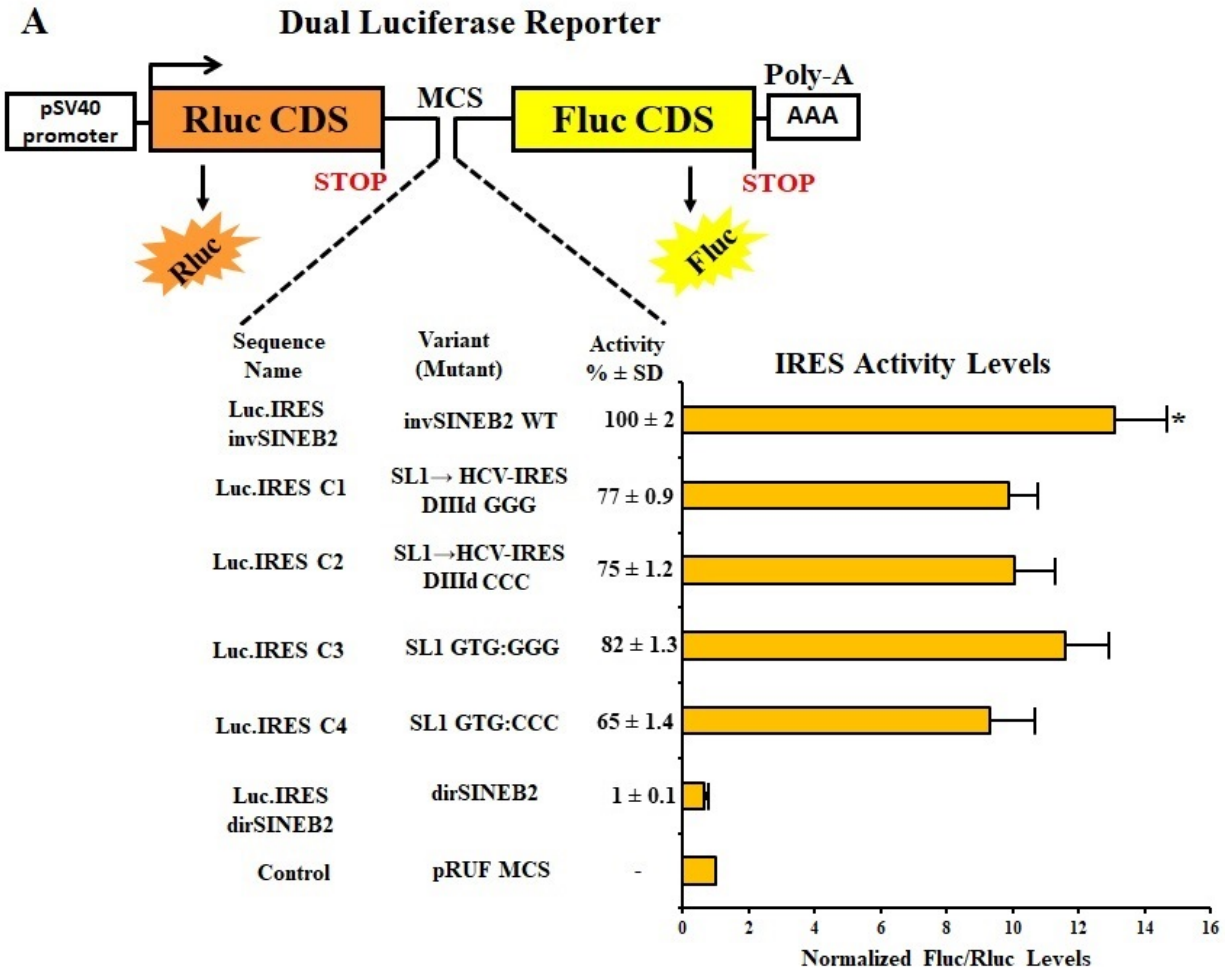


Figure 33 **invSINEB2 SL1 hairpin mutants inhibits IRES-activity in HEK 293T cell lines in vitro.** (A) Schematic representation of pRUF-dual luciferase reporter vectors: containing empty pRUF multiple cloning sites (MSC), cloned inverted- and direct-SINEB2 spacer sequences and chimera mutants; Luc.IRES C1-C4 sequences. Each inserted sequence is flanked by Renilla luciferase (Rluc CDS, brown) cDNA upstream and firefly luciferase (Fluc CDS, yellow) cDNA downstream. All the Rluc-insert-Fluc sequences are under the control of constitutive pSV40 promoter. Each construct is expected to produce a single bicistronic-transcript (Rluc-insert-Fluc) with Rluc protein translated in a cap-dependent manner, while Fluc protein could be translated in a cap-independent manner. Luc.IRES C1-C4 are the invSINEB2 chimera mutants C1-C4 sequences used as SINEUP-ED in SINEUP activity assay, figure 19A. The empty pRUF-MCS and pRUF- invSINEB2-spacer which has SL1 hairpin-^{77}GUG^{79} WT constructs were used as negative and positive controls respectively. HEK 293T cells were transiently transfected with the respective individual pRUF reporter constructs, and prepared 48hours for luciferase activities. Measured Rluc activity is due to cap-dependent translation of Rluc protein while Fluc activities are due to Fluc cap-independent protein synthesis mediated by the inserted sequences. IRES-activity was calculated as ratio of Fluc to Rluc measured activities in relative light units, and normalized against empty pRUF-MCS vector background activities. Empty pRUF activity value was set at 1. InvSINEB2 SL1 hairpin-^{77}GUG^{79} WT activity was set to 100% ±SD. Percentage IRES-activity of each construct was estimated relative to the WT activity. The plot indicates mean ± SD, and representation of N=4 independent replicas; $p > 0.05$ was considered not significant (NS).

5. DISCUSSION

The postgenomic era has changed the way the genome and transcriptome are perceived. It is known that most of the mammalian genome is transcribed, but only few genes are protein-encoding and the rest are non-coding RNAs (ncRNA), of small and long types. The non-coding transcriptome is considered as the major regulator of the mammalian genome, acting in combinatorial ways, at different levels and using several alternative pathways to fine-tune gene expression (Barrett et al., 2012; Levine and Tjian, 2003), influencing cellular processes and giving rise to phenotypic diversity and behavioral traits (Mattick, 2001).

A large group of ncRNAs are lncRNAs transcribed in sense and antisense orientations within the genome, and function at almost all levels controlling gene expression by either forming parts of complex structures or directly regulating replication, transcription, mRNA processing, stability and translational events (Garitano-Trojaola et al., 2013; Marín-Béjar and Huarte, 2015). The majority of known lncRNAs functions at the genomic levels, where they remodel chromatin to regulate gene transcription. However, a new family of natural and synthetic AS lncRNAs, named SINEUPs, were recently discovered in mouse (Carrieri et al., 2012; Zucchelli et al., 2015a) and humans (Schein et al., 2016) to post-transcriptionally activate translation, increasing protein levels of targeted sense genes.

SINEUPs, like other lncRNAs, are organized into modular structures, comprising of a BD that targets and overlaps at the 5'UTR with the sense gene, and an ED constituted by an embedded inverted SINEB2 TE that confers translational activity. By designing a BD sequence antisense to other transcripts, SINEUPs can in principle increase protein synthesis of any gene of interest. Therefore, SINEUPs may be potentially useful for applications in molecular biology experiments, protein manufacturing and RNA therapeutics. To this purpose I designed microSINEUP sequences in the range of small ncRNA that retains SINEUP activity. I also determined that poly-A tail sequences influence SINEUP activity. This work demonstrates for the first time that SINEUPs are active in *Drosophila* cells, suggesting they hijack an evolutionary conserved biological process.

Although SINEUPs are known to exhibit structural-functional relationship dependent on SINE TEs, the molecular mechanisms underlying SINEUPs upregulation of targeted mRNA translation are not fully known. This work demonstrates that SINEUPs promote cap-independent translation using IRES-like mechanisms to activate and increase targeted protein synthesis. In particular, experimental analysis of mutants suggest a model where invSINEB2 ED recruits the 40S ribosome through the SL1 hairpin

structure using an HCV-IRES like mechanism. Furthermore, I have also discovered a novel translational upregulation function for all IRES sequences in increasing targeted proteins synthesis *in trans* through an antisense BD sequence.

Moreover, this work predicts from *in silico* analysis that other cis-regulatory elements may be involved in SINEUP activity; N7-methyl guanosine (m^7G)-cap, cytoplasmic polyadenylation element and poly-A tail, terminal oligopyrimidine tracts (TOPs), N6-methyl adenosine (m^6A), upstream open reading frames (uORFs) and gamma activated inhibitor of translation (GAIT) (***Figure 17***).

The function by IRES and SINEUP molecules suggest a transcriptome-wide patterns of eukaryotic post-transcriptional gene regulation through inter-transcripts gene-networks, possibly promoted by S/AS target pairing. In cells, mRNAs are in competitions for the limited pool of eIFs, ITAFs and free translational competent ribosomes. Therefore, any component such as SINEUP EDs or IRES-containing sequences that could bring in close-proximity the translation machinery and any necessary protein factors to an mRNA through pairing will regulate protein synthesis.

5.1 SINEUPs architecture, a classical definition for lncRNAs functional dependence on TE domains organization

It has been proposed that, like proteins, lncRNAs may follow a modular organization consisting of discrete domain units that can act in combination to drive functionality by assembling diverse combinations of proteins, RNAs or DNAs (Guttman and Rinn, 2012), and the functionality by the discrete domains are due to inserted TEs (Johnson and Guigó, 2014). SINEUPs have modular organization consisting of ED and BD units that act together to drive SINEUP-mediated functionality. The modular organization of natural SINEUPs (Carrieri et al., 2012; Schein et al., 2016) are classical definition of functional dependence of some class of lncRNAs on TEs organization into functional portable domain units and coordination in their overall function. In natural SINEUP of AS UCHL1, embedded invSINEB2 TE have been shown to exclusively define AS Uchl1 ability to upregulate Uchl1 translation upon mTORC1 inhibition by rapamycin induced-stress (Carrieri et al., 2012), suggesting that subfamilies of embedded TEs are fundamental for lncRNAs function (Zucchelli et al., 2015b). Although most TEs are silenced by epigenetic control (Casa and Gabellini, 2012; Ekram et al., 2012), majority of

them are under regulated expression to mediate functionality upon diverse cellular conditions, including stress.

The embedded TEs in lncRNAs may present conserved cis-regulatory sequence and structural motifs to recruit factors necessary for a biological task. Therefore, just like proteins, SINEUP and other lncRNA structures follow modular domain organization into units mediating the overall functionality, and the domains function are driven by conserved TEs.

5.2 SINEUP functionality, an evolutionary conserved role beyond the mammalian genome

Natural SINEUPs functionality was originally discovered in the mouse genome (Carrieri et al., 2012) and subsequently, human SINEUPs were discovered (Schein et al., 2016). Considering that embedded TEs, LINEs and SINEs sequences are evolutionary different among mammals, it has been unclear whether a SINEUP could be functional in different organisms. For mammals, synthetic SINEUPs of mouse and human origins have been proved to works in cell lines from both organisms. However, SINEUPs functionality has not been tested in non-mammalian systems. Here, I tested and show for the first time that SINEUPs maintains biological activity in *Drosophila* S2 cell lines *in vitro* ***(Figure 21)***, suggesting that SINEUPs may hijack an evolutionary conserved pathway. This finding is surprising, since there are no known SINE elements in the *Drosophila* genome. As compared to over 40% of TEs in the mammalian genome (De Koning et al., 2011), the *Drosophila* genome is composed of about 20% of TEs (McCullers and Steiniger, 2017), comprising most of LTR families, followed by LINE-like families, and terminal inverted repeat (TIR) and foldback (FB) elements that are the minority families (Kaminker et al., 2002): of which at least 30% of them are actively expressed as full-length elements (McCullers and Steiniger, 2017), suggesting functionality within the *Drosophila* genome. These results may suggest that SINEUP activity is more based on structure rather than primary sequence, and therefore similarly conserved TE structures in different cell types and organisms may maintain an evolutionary conserved, structurally related function.

This is probably why IRES sequences from both cellular and viral origins are functional when substituted for invSINEB2 ED in SINEUP activity assay ***(Figure 26 Figure 27).*** Viral IRES in particular are functional in different cell types from different organisms, making for instance CrPV-IRES, supposed to

be functional only in cricket and related insects is active when used in mammalian cells, suggesting that this elements functionality is driven by structures. This data provides a basis to look for evolutionary conserved structurally related embedded TEs or other repeats that can function as "SINEUP EDs" in non-mammalian cell types such as the *Drosophila*, yeast and even the bacteria, which will provide details on all the mechanisms of SINEUP regulation of targeted proteins synthesis and overall functionality in driving cellular processes *in vivo*. Overall, this data provides functional evidence for potentially conserved SINE related TEs in flies, further suggesting that TEs may have driven the evolution of lncRNAs gene structures and functionality.

5.3 Synthetic SINEUPs optimization, the influence from the 5' and 3' ends on SINEUP activity

Although they are not part of the BD and ED sequences, the 5' and 3' ends might contain cis-acting elements such as the 5' m^7G cap-structure and 3' poly-A tail to regulate post-transcriptional processes such as RNA processing, export and stability on the SINEUP RNA. Activities of RBPs mediated by *cis*-regulatory elements on one mRNA can be transferred in trans-regulatory activities (Reveal et al., 2010). The effect can either be repression or activation in *trans* on targeted mRNA biogenesis, including translation as shown through inter-transcript interactions and ribonucleoprotein particle (RNP) formation between coding and non-coding Oskar transcripts in regulating Oskar protein synthesis by Bruno, a 3'end binding protein (Reveal et al., 2010; Macdonald et al., 2016).

In translation, mRNA 5' m^7G cap-structure and 3' poly-A tail forms close-loops through their respective RBPs; eIF4E within the eIF4F-complex and the PABP to promote cap-dependent translation (Hinnebusch, 2017; Hinnebusch and Lorsch, 2012; Komar and Hatzoglou, 2011; Sonenberg and Hinnebusch, 2009). I have showed in this work that indeed poly-A tail sequences influence SINEUP activity, where randomly added poly-A sequences are detrimental to SINEUP ability to upregulate translation, while poly-A signals placed at the 3'end of SINEUPs increase their activity (***Figure 17***), suggesting that poly-A sequences are involved in recruiting *trans*-acting RBPs like PABPs, influencing translation activities mediated by SINEUPs. When recruited by SINEUP, the PABP as a translation regulatory factor may act in *trans* through communication between SINEUPs and the targeted mRNA by inter-transcripts interactions (Macdonald et al. 2016), enhancing protein synthesis. This could be one of the reasons why the same SINEUP works better when expressed from pCS2+ plasmid as compared

pcDNA3.1- harboring SV40 and BGH poly-A signals, respectively (**Figure 15**). Thus, SV40 poly-A signal influences better SINEUP activity than BGH poly-A site, suggesting that this sequence feature is required for an optimal SINEUP activity.

Similarly, the 5'end m^7G-cap structure on SINEUPs could also influence SINEUP activity through inter-transcripts communication in *trans* (Reveal et al., 2010; Macdonald et al., 2016), where SINEUP m^7G-cap could recruit the proteins of the cap-eIF4F complex; eIF4A, eIF4E, eIF4G (Hinnebusch and Lorsch, 2012), and make them available for the translation of the sense mRNA in *trans*. Therefore, an RNP complex consisting of SINEUP-mRNAs and their respective RBPs may exist as a unique RNP of a ribonome (Keene, 2007; Mansfield and Keene, 2009; Mitchell and Parker, 2014), where for instance SINEUP 5'end m^7G-cap and 3' end poly-A tails recruits eIFs and the ribosome to mediate cap-dependent translation of the targeted mRNA under normal cellular conditions. In this regard, I propose that in addition to alternative translation pathways, SINEUPs can also enhance cap-dependent translation by using exclusively their 5'end m^7G cap-structure and 3'end poly-A tail to respectively recruit eIF4F-complex and PABPs, acting as a molecular "cargo" or an RNA chaperone in mediating cap-dependent translation of the targeted mRNA. This could be promoted by SINEUP-mRNA duplex formation by S/AS pairing mediated by the BD and further interactions by the recruited RBPs in RNP (Mitchell and Parker, 2014). As shown in ***Figure 34***, the formation of double strand RNA between SINEUP BD and 5' end of the targeted mRNA would place the SINEUP 5' m^7G-cap structure and any bound initiation complexes close to the start codon (AUGi) of the mRNA. The double strand formation from the pairing could inhibit 5'end scanning by any recruited initiation complexes on both the SINEUP and targeted mRNA, making it logical to think that any initiation complex on the SINEUP- m^7G cap structure will be put directly at the vicinity of the mRNA AUGi, increasing the efficiency of translation initiation in time and space.

The S/AS pairing by the BD itself will naturally inhibit any available mRNA 5'UTR cis-regulatory inhibitors; uORFs, secondary structures, hairpins or inhibitory protein binding sites (figure 9A) (Hinnebusch, 2014; Hinnebusch and Lorsch, 2012; Jackson et al., 2010) that can inhibit loading and scanning by initiation complexes at the 5'end on the mRNA. If this is so, then blocking part or the entire 5'UTR of any mRNA with antisense sequences that can possess 5' end caps and 3' end poly-A tails should promote translation. Indeed there are reported data of translation upregulation by short antisense-oligos in both mammalian cells (Liang et al., 2017, 2016) and prokaryotes (Hayashi et al., 2012). This could be the reason why both natural and synthetic SINEUPs BD pairings are around the AUGi of the

targeted mRNA, thus natural AS 5'end BDs have evolutional implication to probably repress 5'UTR activities on the targeted sense mRNA. It could even be that SINEUPs BD transforms a targeted mRNA into leaderless mRNA (Akulich et al., 2016), by blocking and making targeted mRNAs loose functional activities at the 5'UTR so that the organized initiation complexes by the SINEUP could easily be positioned at the mRNA initiation codon, AUGi, escaping all "time-wasting" processes of scanning and loading of initiation complexes at mRNA 5'ends, a hallmark of leaderless mRNAs and bacterial genes (Srivastava et al., 2016).

Therefore, if the BD has no role beyond specifying pairing with the targeted mRNA, then one should expect different BDs of the same SINEUP to exhibit the same effect on SINEUP activity. However, previous and current work could show that the same SINEUP with shorter and longer BDs around the mRNA AUGi have different effects on SINEUP activity, with the shorter one that has BD before the AUGi promoting better SINEUP activity than the longer one that extends beyond the AUGi of both overexpressed GFP and endogenously expressed DJ1 mRNA (**Figure *16***; Zucchelli et al., 2015a). Since translation initiation is the rate limiting step in protein synthesis (Hinnebusch and Lorsch, 2012; Jackson et al., 2010; Sonenberg and Hinnebusch, 2009), it may suggest that any enhancer activity by any trans-acting factor, a protein or RNA like SINEUP, will reduce this rate and therefore enhance translation, increasing protein levels in time and space.

However, it is still unclear if m^7G-cap structure and poly-A tail elements have roles in natural SINEUPs. For synthetic SINEUPs, both pol II and pol III mediated transcripts of full-length synthetic and miniSINEUP are functional. In fact, in the reported optimization studies, miniSINEUP-GFP expressed from human H1, a pol III promoter, was functional with optimal SINEUP activity when cells were transfected with a pDual plasmid expressing also the targeted GFP gene under a pol II CMV promoter (Figure *20*). The construct has BGH poly-A signal, and pol III is known to also produce SINE B2 transcripts that are polyadenylated (Borodulina and Kramerov, 2008), as well as pol III transcripts can be capped (Tschudi and Ullu, 2002), suggesting that the miniSINEUP-GFP from the pDual plasmid is probably polyadenylated and capped. This further suggest that cap-dependent translation can be promoted by synthetic SINEUPs in cells *in vitro*. Therefore, SINEUP activity could be influenced by both 5' and 3' end features. It will however be useful to carry out 5' and 3' RACE on both natural and synthetic SINEUPs to understand better the regulatory elements present at these regions to study their effects on SINEUP activity.

Here I would like to propose that the BD of both natural and synthetic SINEUPs may have further layers of regulations beyond just defining target specificity for SINEUPs: it contains the m^7G-cap structure that together with the 3'end poly-A tail promote cap-dependent translation of targeted mRNAs protein synthesis in an efficient way beyond what the mRNA alone could do in the absence of the BD by the SINEUP. The BD can probably block partially or entirely the 5'UTR containing potential inhibitory cis-regulatory elements of the targeted mRNA, making it "functionless and leaderless", and this may promote a bacterial-like translation initiation by mediating direct binding of initiation complexes (48S or 80S) at the AUGi, enhancing both cap-dependent and/or alternative translation, and this role could be dependent or independent of the ED. The biological implication of this reasoning is that synthetic SINEUPs will target and work better on mRNAs with long and structured inhibitory 5'UTRs.

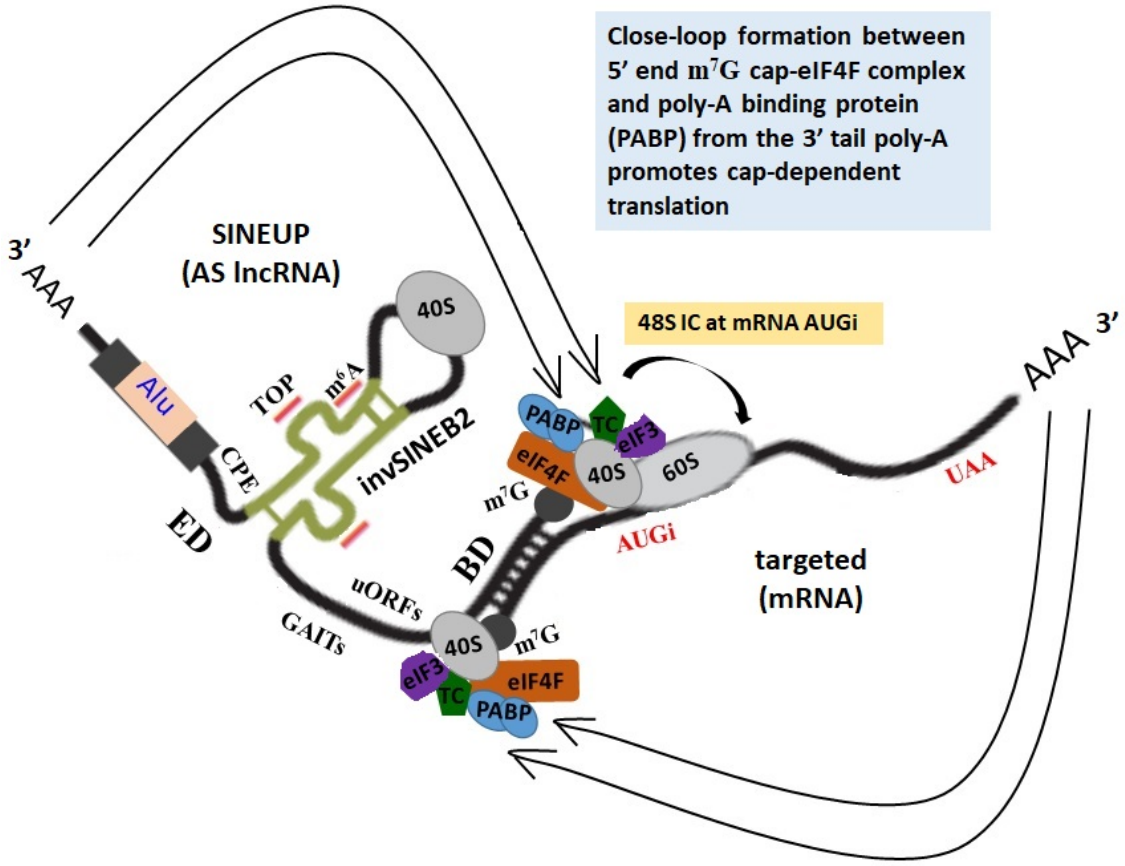


Figure 34 Cap-dependent translation possibilities mediated by SINEUP 5' and 3' ends. Scheme showing S/AS 5'head-to-head divergent pairing between SINEUP and targeted mRNA. The pairing by the BD blocks and probably inhibit the 5'UTR of the mRNA, making it "leaderless and functionless". Under normal cellular conditions when all factors are available, close-loop formation by SINEUP 5' and 3' end (shown by pair of black arrows) could organize and position the 48S initiation complex (IC) within the vicinity of the mRNA AUGi in a cap-dependent way. The recruited 48S IC can only move in 5'to 3' forward direction on the mRNA, promoting translation of the mRNA. Conversely, recruited 48S by the mRNA can also move on SINEUP, but since it is a lncRNA, protein synthesis might not be favored. This pathway might not require the invSINEB2 structure and associated cis-acting elements; TOPs, m^6A, uORFs, CPEs and GAITs present in the ED of SINEUP.

5.4 *Cis*-regulatory elements in SINEUPs: a paradigm for their role as "molecular-hubs" in driving lncRNAs evolution and function

Any transcribed RNA, including both mRNA and small ncRNA or lncRNA, will exist in a complex with other molecules including proteins and other RNAs-forming RNP complexes (Mitchell and Parker, 2014). This observation implies that regulation of protein synthesis exists hand-in-hand with modulation by both proteins and ncRNA assembling, forming complex-structures and functioning dynamically to regulate gene expression in time and space within a ribonome (Mansfield and Keene, 2009). The ability of an mRNA to interact with any protein and coding or non-coding RNAs depends on specific sequence and structural features within the mRNA itself (Mitchell and Parker, 2014). These are the so called cis-regulatory elements such as 7-methyl guanosine (7mG)-cap structure, poly-A tails, terminal oligopyrimidine tracts (TOPs), N6-methyl adenosine (m^6A), upstream open reading frames (uORFs), miRNA response elements (MRE), AU-rich elements (ARE) and internal ribosome entry sites (IRESs) that act as binding sites for trans-acting RBPs and complexes, like the ribosome. Thus, mRNAs and even lncRNAs with similar cis-regulatory motifs will compete for trans-acting factors for functional regulations in conditions like stress where there are limited pools of such factors.

However, since lncRNAs are non-protein encoding, they could act as RNA scaffolds or carriers to make available recruited RBPs and ribosome complexes to a targeted mRNA for its remodeling through exchange of the factors, enabling the regulation of such targeted mRNA. lncRNAs may possess these elements for their own cis-regulations; such as splicing, export, stability and degradation. For instance, the MRE, ARE and siRNA sites may be involved in regulating processes leading to lncRNAs degradation. However, the presence of m^7G-cap, poly-A sequences, TOPs, m^6A, uORFs and IRESs could also mean they exist for regulations in *trans*, especially for the role of TOPs, m^6A, uORFs and IRESs in cap-independent translation. The prediction of all the aforementioned cis-regulatory elements, especially poly-A, TOPs, m^6A, uORF and IRES elements ***(appendix Figure 22)*** in the SINEUP ED would therefore suggest that these cis-acting elements bind to trans-acting factors and complexes, providing them for the regulation of the translation of the targeted mRNA through specific binding by the S/AS pairing. The SINEUP BD therefore provides target-specific S/AS pairing, which could be a way to form a messenger mRNP complex (Ferrandon et al., 1997; Mitchell and Parker, 2014) tethering it to a recruited polysome for the mRNA-specific translation upregulation. In this manner, cis-regulatory elements such as m^7G-cap, poly-A tail, TOPs, m^6A, uORFs and IRESs in SINEUPs will recruit, for instance, PTB, eIF3, translation competent ribosomes/eIFs respectively to form a mRNP complex (Mitchell et al., 2014) to

promote translation upregulation, escaping cytoplasmic stress-mediated translation inhibitor complexes localized on ribosomes or within some cellular compartments.

Sequence specific complementary in RNA base-pairing is a known strategy to segregate mRNA pairs into mRNP granules comprising of several mRNAs of the same or different gene, leading to specific mRNA modulations (Ferrandon et al., 1997; Keene, 2007; Mitchell and Parker, 2014). It could be inferred that the SINEUP, targeted mRNA and their associated RBPs will co-exist in a way that benefits mRNA translation. It could be that SINEUP-mRNA pairing and association with various RBPs recruits a limited pool of eIFs and translation-competent ribosomes to form complexes under a particular stress signaling cascade, escaping the extracellular regulation. In this way, proteins become synthesized without any external effect ***(Figure 29 and Figure** 30)**.***

Therefore, this work is proposing that just like SINEUPs, lncRNAs and even mRNAs may function as "molecular-hubs" to convey important trans-acting RBPs such as canonical translation initiation factors (eIFs), co-factors like ITAFs (PTBs, PABPs), and translational competent ribosomes for the *trans*-regulation of the bound mRNA or mRNP in time and space. This will be a *trans*-effect of a *cis*-acting element, because the elements on the lncRNAs recruit the RBPs or the ribosomal complexes acting in *trans* to promote the translation of the mRNA. The ability of SINEUPs to mediate cap-independent translation **(*Figure 23* - Figure** 25**)** will mean that a SINEUP, as a "molecular-hub", is able to recruit the necessary eIFs and translation competent ribosomal complexes to promote internal ribosome entry on the mRNA for its translation, independently or in coordination with the 5'cap-structure on the mRNA. Similar to the ability of miRNA to bring Argonaut proteins to mRNA 3'UTR through miRNA response elements (MRE) for its regulation (Barret et al., 2012) the mRNA is supposed to possess a "seed sequence" at the 5'UTR to be bound and regulated by SINEUP as a free bound mRNA or in an mRNP, consisting of SINEUP-mRNA, RBPs, and ribosomal complexes. It is this same seed that is utilized in c-myc and other IRES -containing mRNAs to increase targeted proteins synthesis in *trans*.

5.5 c-myc 5'UTR cis-acting IRES elements mediates *trans*-translation upregulation of targeted protein synthesis by a possible inter-transcripts communication

The transcriptome at any point in time is dynamic, consisting of inter- and intra-specific interactions among different sets of RNAs. The formation of RNA duplexes, complexes and other structured types of RNAs are key to regulation of gene expression (Sharma et al., 2016). The dynamic interactions of snRNA-snRNA and snRNA-mRNAs leads to splicing, interactions between amino-acylated tRNAs and mRNAs mediate peptide bond formation, snoRNAs and mRNAs for their modification (Engreitz et al., 2014; Sharma et al., 2016) with functional roles known for small ncRNAs like snRNA, snoRNAs, miRNAs, siRNAs, while little is known for other interactions like lncRNA-mRNA and mRNA-mRNA. It has been considered that long range interactions involving the ends of mRNAs and their associated factors are required for translation (Macdonald et al., 2016), thus a factor bound at the 5'end of an mRNA could interact with the 3'end sequence directly or through another factor of the other mRNA, which could lead to *trans*-regulatory effects. There are also experimental evidences that functionally related mRNAs could form mRNA regulons similar to a bacterial "operon" which are coordinately regulated together post-transcriptionally from splicing to translation (Keene, 2001, 2007; Mitchell and Parker, 2014; Simone and Keene, 2013)

I discover that synthetic lncRNAs containing IRES elements, SINEUP-IRES, have SINEUP-like activity, that is they are able to increase translation of targeted mRNAs post-transcriptionally, as mediated by IRES sequences acting as cis-elements in *trans* in promoting translational activation of the targeted protein synthesis. I detected *in silico* that c-myc 5'UTR can form partial base-pairing with sequences on targeted mRNAs and upregulate their protein synthesis in a SINEUP-like fashion. This was a post-transcriptional increase in protein synthesis mediated by the IRES element in *trans.* This is a functional prove for the first time that an IRES-containing mRNA stimulates in *trans* an increase in protein synthesis of targeted mRNAs. This activity is mediated by the functional equivalent of BD in SINEUPs by pairing to antisense sequences in the 5'UTR of target mRNAs. Therefore, the transcribed c-myc mRNA forms mRNA-mRNA base pairing which could lead to recruitment of RBPs, forming simple or higher-order mRNP complexes or granules (Ferrandon et al., 1997; Keene, 2007; Mitchell and Parker, 2014), probably consisting of many c-myc mRNAs paired with single type or all the targeted mRNAs, RBPs and other complexes involved in translation.

As such this could be a mechanism by which IRES-containing mRNAs, like c-myc mRNA, are able to upregulate translation of targeted mRNAs in *trans* **(Figure 29 and *Figure 30*).** It probably forms mRNA-mRNA pairing, regulatory RNP granules with eIFs, ITAFs, the ribosome and tRNAs to be able to synthesize both c-myc and targeted proteins to form simple structures or complexes as known to exist for other mRNAs pairing in mRNPs (Ferrandon et al., 1997; Mitchell and Parker, 2014). This regulatory pathway has probably evolved as a strategy to group mRNAs based on target-specific binding into a confine compartment and fine-tuning their translation in *trans* as a way to upregulate their expression. This is again a functional proof of the concept of mRNA operons/regulons where functionally related mRNAs are coordinately regulated by multi-targeted RBPs (Keene, 2007; Mansfield and Keene, 2009; Mitchell and Parker, 2014). This may represent a way by which the c-myc coding sequence itself could be protected and expressed alongside its regulatory targets in such confined complexes, escaping external cytoplasmic regulations. Accordingly, few ribosomal complexes and eIFs could easily be recycled and reused in an optimized manner for cells when these factors become limited.

While the ability of an mRNA to activate translation in *trans* needs proof of the existence of pairing between c-myc and it targets, this observation may represent a new paradigm for how translation is regulated. This may have implications in health and disease, where this mechanism may become a drug target.

5.6 SINEUP TEs promotes cap-independent translation using HCV-IRES like internal ribosome entry mechanism

This study employed both *in silico* and experimental approaches to unravel the cap-independent translation activities mediated by SINEUPs. *in silico* studies have revealed sequence and structural motifs similarities between the main SINEUP ED, the embedded invSINEB2 TE, and IRES elements annotated in the IRESsite database ***(Table 3)***, suggesting potential similarities in mediating cap-independent translation. Translational activity studies confirmed *in silico* analysi, thus I was able to show that IRESs work as SINEUP-ED, the embedded invSINEB2 of AS Uchl1 SINEUP and vice versa. Thus, invSINEB2 and IRES elements complement each other, indicating that SINEUPs mainly rely on the invSINEB2 element to promote cap-independent translation of targeted protein synthesis by internal ribosome entry mechanisms. But since different IRES classes use different pathways to facilitate internal ribosome entry, I further studied this mechanism relying on HCV-IRES. This is because HCV-IRES does not require an

eIF4F-complex in activating translation, instead it interacts directly with few available eIFs and the 40S ribosome (Jaafar et al., 2016; Matsuda and Mauro, 2014). Furthermore, I know that upon mTORC1 inhibition by rapamycin-induced stress, eIF4F-complex mediated cap-dependent translation is inhibited, but SINEUP invSINEB2 TE of AS Uchl1 promotes translation of Uchl1 protein synthesis under eIF4F-complex inhibition (Carrieri et al., 2015), and here I show that invSINEB2 acts as an IRES in a bicistronic assay.

Our data on HCV-IRES mutant studies are consistent in terms of how HCV-IRES promotes translational activation of protein synthesis, using SINEUP activity assay (***Figure 31***). Thus, I was able to reproduce the disruption of WT activity from key known mutations within the HCV-IRES that were shown to inhibit the WT translational function in IRES-activity assay (Kieft et al., 2001; Matsuda and Mauro, 2014). Noticeably, mutations within the HCV-IRES DIIId of WT, used as SINEUP-ED, disrupted translational activation. Therefore it is tempting to speculate that IRES and invSINEB2 elements complement each other functionally by using similarly conserved structures to recruit similar molecular-partners in translational activation.

Structurally, invSINEB2 TE and HCV-IRES have similar conserved sequence motifs at the region where HCV-IRES interacts with the ribosome. Previously, chemical footprinting combined with mutant studies have identified a stable SLI hairpin structure with dynamic loop properties in invSINEB2 of AS Uchl1 to be the main structural determinant driving invSINEB2-mediated SINEUP activity (Podbevšek et al., 2018). Direct sequence comparisons show that the SL1 hairpin structure has ^{74}GUUGUG^{79} nucleotides similar to HCV-IRES DIIId loop at positions -^{263}GUUGGG^{268} with only the respective U^{78} and G^{267} nucleotides mismatch ***(Figure 32A)***, suggesting potential similarities in interaction with the same external reactive agents including the 18S rRNA helix (h) 26 site, where the HCV-IRES DIIId-^{266}GGG^{268} is known to directly bind to recruit the 40S ribosome (Kieft et al., 2001; Matsuda and Mauro, 2014). Although not with the same strength of a G nucleotide, the U nucleotide can basically form non-classical Watson-Crick base pair with C nucleotide in RNAs (Leontis et al., 2002; Varani and McClain, 2000), in particular through a *trans* sugar-edge/Hoogsten base pair configuration (Leontis and Westhof, 2001) in hairpins and stem loop structures that are relevant for RNA functions.

The isomeric form of Uridine (U) nucleotide, pseudouridine is known to occur in tRNAs, snoRNAs, snRNAs, and well-conserved functionally important regions of ribosomal RNAs (Schwartz et al., 2014), where it offers RNA stability when it occurs in helical stems and de-stability in helical loops, promoting

RNA-RNA interactions through base pairing with A, G, C and U nucleotides and also, it promotes RNA-protein interactions (Kierzek et al., 2014). Therefore, the U^{78} nucleotide in invSINEB2 SL1 hairpin-$^{77}GUG^{79}$ structure may probably pair to similar bases used by the G^{267} nucleotide in HCV-IRES DIIId-$^{266}GGG^{268}$ structure at the known 18S rRNA h26 site of the 40S ribosome. However, there could be differences with regards to how each nucleotide stabilizes the 40S ribosome upon mediating complementary base-pairing, shaping the overall 40S ribosome binding dynamism, interaction with other protein complexes involved in translation initiation, and positioning on the targeted mRNA for the protein synthesis. HCV-IRES DIIId-$^{266}GGG^{268}$ WT but not a DIIId-$^{266}CCC^{268}$ mutant is known to directly interact with the 18S-rRNA $^{1113}AUUCCCA^{1119}$ h26 site by direct complementary base pairing involving the DIIId-$^{266}GGG^{268}$ and 18S rRNA $^{1116}CCC^{1118}$ to mediate HCV-IRES translational activity (Malygin et al., 2013; Matsuda and Mauro, 2014).

I hypothesized and tested experimentally the structural and functional complementation of invSINEB2 SL1 hairpin -$^{77}GUG^{79}$ sequences with HCV-IRES. To this purpose, I generated various chimera mutations between the two structures to form C1, C2, C5, C6 and C7 mutants as well as single and triple nucleotides substitutions mutations to generate C3 and C4 mutants, respectively within the invSINEB2 SL1 hairpin ***(Figure 32A).*** I employed both SINEUP- and IRES-activity assays to experimentally test our hypothesis. The overall results obtained for all the chimera mutants using both assays were lower than the invSINEB2 SL1 hairpin WT translational activities. This is probably because the entire SL1 hairpin-$^{74}UUGUGA^{80}$ WT loop region might be involved in interactions with the 40S ribosome or the other components of the initiation complex. This does not seem to be the case for the HCV-IRES, because DIIId-$^{264}UUGGG^{268}$ WT and mutants C5 and C6, except C7, have similar translational activities as compared to the SL1 hairpin WT ***(Figure 32B)***. This indicates that conserved structures with triplet -GUG and -GGG loops are solely used in interaction with the ribosome when in HCV-IRES, but not in invSINEB2-containing sequences. Thus, the preceding $^{74}UU^{75}$ and A^{80} nucleotides are probably not involved in the interactions with the 40S ribosome in SL1 hairpin and DIIId- structures in HCV-IRES containing sequences. If so, then the activity of mutant C7 is expected to be much lower than the observed 63% activity, because it has the SLI hairpin-$^{75}UUCCCA^{80}$ substituted for DIIId-$^{264}UUGGG^{268}$, where the -CCC is expected to kill translational activity in a similar way as HCV-IRES DIIId-$^{264}UUCCC^{268}$ M9 mutant ***(Figure 31B)***. The 63% activity by C7 would however suggest that in the presence of triplet -CCC nucleotides, the neighbor nucleotides $^{74}UU^{75}$ and A^{80} within the SL1 hairpin-$^{75}UUCCCA^{80}$ loop might be active and functional, leading to the observed C7 mutant activity. If so, this would then be a

functional proof for the suggestion from chemical footprinting data that nucleotides U^{76} and A^{80} within the conserved invSINEB2 SLI hairpin-^{75}UUCCCA^{80} are prone to external reactive agents and might be involved in the overall SL1 hairpin-mediated functions (Podbevšek et al., 2018). Thus, nucleotides U^{76} and A^{80} might have been involved in base pairing interactions with 40S ribosome recruitment and therefore the observed C7 mutant translational activity.

Interestingly, potential base-pairing within the hypothesised 18S-rRNA ^{1113}AUUCCCA^{1119} h26 site by the C7 could only be possible by reverse complementary base-pairing as shown in ***Figure 35B***, since, unlike HCV-IRES DIIId- WT, direct complementary base pairing by the SL1 hairpin-^{75}UUCCCA^{80} seems unfavorable, because only nucleotide U^{76} and sometimes ^{77}C but neither -^{78}CC^{79} nor A^{80} nucleotides could pair with potential U^{1115} or C^{1116} nucleotides within the h26 site, as shown in ***Figure 35B.*** This is possibly the reason why HCV-IRES DIIId-^{264}UUCCC^{268} M9 mutant is not active, showing only 2% and 13% of the DIIId-^{264}UUGGG^{268} WT translational activities in IRES-activity assay (Kieft et al., 1999) and SINEUP activity assay ***(Figure 31B),*** respectively. This notion is in fact supported by C7 direct opposite Chimera C2 mutant with SL1 hairpin substituted by HCV-IRES DIIId-^{264}UUCCC^{268}, which also recorded 63 and 75% translational activities, respectively in SINEUP activity and IRES-activity assay. The 63% activity in SINEUP activity assay for both C2 and C7 could be a confirmation that the loop sequences within invSINEB2 conserved structures; SL1 hairpin-^{75}UUCCCA^{80} and HCV-IRES DIIId-^{264}UUCCC^{268} mutant structures made use of one or the two -UU nucleotides preceding the -CCC triplet nucleotides in making interactions with the 40S ribosome by reverse complementary base-pairing at the hypothesized 18S rRNA h26 site, leading to the observed activities.

For better characterization and interpretation of the chimera mutant C2 and C7 results, I relied on the triplet-^{77}CCC^{79} direct substitution for the conserved SL1 hairpin-^{77}GUG^{79} WT in C4 mutant, using mutant C3 with one U^{78} to G^{78} nucleotide substitution as additional positive control when compared to the SL1 hairpin-^{77}GUG^{79} WT. The retention of 51 and 65% of WT translational activities by the mutant C4 respectively in SINEUP and IRES-activity assay in HEK cells confirmed that the mode of interaction between the invSINEB2 SL1 hairpin structure and the 18S-rRNA h26 is not exactly the same as that used by the HCV-IRES DIIId WT structure. While the 18S-rRNA h26 site is probably the same, any potential interaction should be by reverse complementary base-pairing involving reactive nucleotides U^{76} and A^{80} (Podbevšek et al., 2018), conferring translational activities by the C4 mutant. While it is much lower, the activity was similar to the chimera mutants C2 and C7. Instead, invSINEB2 SL1 hairpin-^{77}GUG^{79} WT structure could potentially interact by both direct and reverse complementary base pairing as depicted

(Figure 35A). Just like HCV-IRES DIIId-^{266}GGG^{268} WT, direct complementary base-pairing will potentially involve the terminal -^{77}GUG^{79} loop through two main probable interactions ***(Figure 35A1).***

However, the speculated reverse complementary base-pairing could be the dominant mode of interaction used by the SL1 hairpin-^{75}UUGUGA^{80} WT, since reverse complimentary base-paring would favor the ability of the predicted reactive nucleotides U^{76} and A^{80} to interact by base-pairing at the 18S rRNA h26 ^{1114}UUCCCA^{1119} site. Any probable interactions by nucleotides U^{76} and A^{80} in the SL1 hairpin-^{75}UUGUGA^{80} WT will increase the number of base-pairing from 2 in direct to up to 4 in reverse complementary base-pairing ***(Figure 35A)***, offering extra base-pairs that could lead to structural conformations within the 40S ribosome, which could promote the 40S ribosome positioning on the sense mRNA initiator AUGi site. The reactive 18S rRNA nucleotides ^{1114}UUCCCA^{1119} were revealed by chemical footprinting studies to be the conserved reactive nucleotides in the 18S-rRNA h26 site (Malygin et al., 2013; Matsuda and Mauro, 2014). The ^{1114}UUCCCA^{1119} of the 18S rRNA and C4-^{75}UUCCCA^{80} mutants could only react through reverse complementary base pairing involving all UU and A nucleotides in several combinations that will include two of the C nucleotides as indicated ***(Figure 35B)***. The A^{80} nucleotide could also offer better stability by forming canonical Watson-Crick base-pairing with U nucleotides in the ^{1114}UUCCCA^{1119} of the 18S rRNA, which can only occur by reverse complementary base-pairing as shown ***(Figure 35B).***

Moreover, this reverse mode of complementary base-pairing could be promoted though S/AS interaction between the two structures, which seems to favor the ability of any recruited pre-initiation complex to be positioned and moved in 5' to 3' direction in decoding the targeted sense mRNA. This could biologically promote SINEUPs-mediated activation of the sense mRNA translation. The ribosome as a biological machine has probably evolved to have only 5' to 3' activity. The direct complementary mode of base-paring may probably lead to scanning on the SINEUP RNA by the preinitiation complex instead of moving on the sense mRNA, which would probably not yield the SINEUP-mediated task. However, in the case of driving cap-independent translation of Fluc protein synthesis in IRES-activity assay, direct complementary base-pairing will dominate ***(Figure 35A1***), since in this context, the invSINEB2 WT is placed upstream a Fluc mRNA ***(Figure 33).*** For HCV-IRES-mediated translation of HCV proteins, direct complementary base-pairing could be the only scenario, because the ribosome has only one directional 5' to 3' movement during translation. In the context of its being used as an ED in SINEUPs, both complementary base-parings could be activated, while the reverse pairing could be lowly favored because the DIIId triple -GGG Watson-click base-pairing with h26 triplet-CCC should be able to offer

the needed stability, explaining why in SINEUP activity assay, the HCV-IRES with DIIId -CCC M9 had only 13% of WT activity ***(Figure 31B)***. Therefore, in the context of SINEUPs, S/AS interaction between SL1 hairpin and the 18S rRNA h26 could easily position the pre-initiation complex at the decoding center on the sense -mRNA in 5'-to- 3' directional movement, leading to SINEUP-mediated activity.

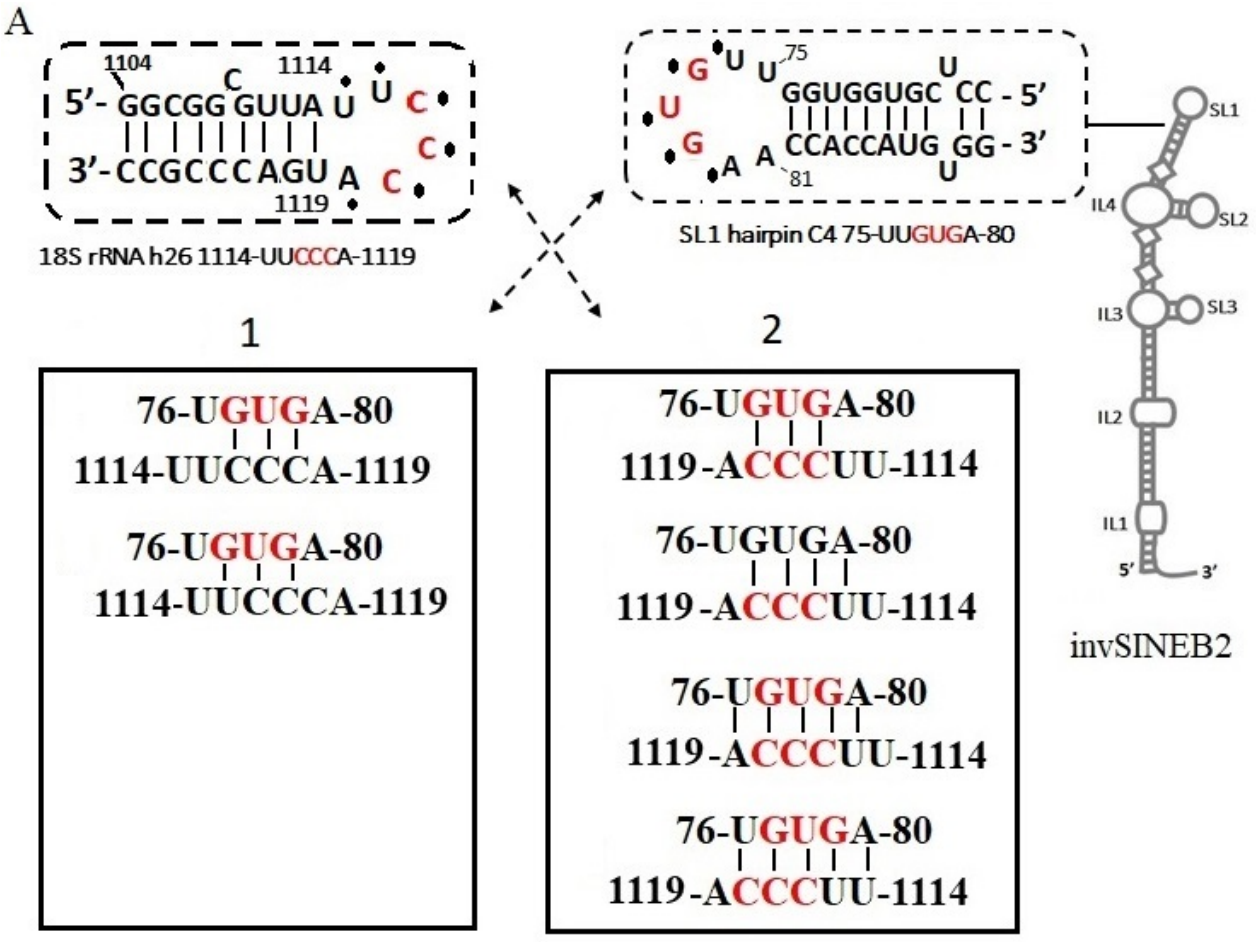
5.6.1 Nucleotides conservation in SL1 hairpin apical loop-structure may dictate ribosome interactions

Overall, it seems that the conserved nucleotide U^{78} in the SL1 hairpin-^{76}UGUGA^{80} may play a crucial role in uniquely dictating this suggested mode of reverse complementary base-pairing in S/AS interaction with the ribosome, which is not possible in the HCV-IRES DIIId-^{264}UUGGG^{268} because it has ^{267}G instead of the ^{78}U nucleotide. Thus, nucleotide U^{78} may cause this kind of speculated redundancy by forming *trans* sugar-edge/Hoogsten base-pairing (Leontis and Westhof, 2001) at the 18S-rRNA ^{1114}UUCCCA^{1119} h26 site, and hence promoting better function. It could also be modified and isomerized to pseudouridine (Kierzek et al., 2014), promoting RNA-RNA interactions through base pairing with A, G, C and U nucleotides and RNA-protein interactions. This could also be one of the reasons why mutant C3 with SL1 hairpin-^{77}GGG^{79} failed to retain the full function of the SL1 hairpin-^{77}GUG^{79} WT in both IRES-activity and SINEUP activity assays ***(Figure 32B and Figure 33A)***. The nucleotide G would be expected to offer better stability due to its ability to form canonical Watson-Crick base-pairing with C during interactions at 18S rRNA h26 site. However, it lost 20% of the WT activity, probably because it lost other modes of functionality, which is promoted by the conserved U^{78} in the hairpin-^{77}GUG^{79} WT structure. Alternatively, it could be that G^{78} canonical Watson-Crick base-pairing in the C3 mutant relative to U^{78} *trans* Sugar-edge/Hoogsten base-pairing in SL1 hairpin WT stabilizes the bond during interactions with 40S ribosome to an extent that it reduces the efficiency by which the ribosome is released from the invSINEB2 element and loaded to the sense-mRNA, leading to less translational activity than the WT.

The U^{78} mediated pairing may be more flexible and dynamic with regard to how any recruited 40S or 43S PIC is released for the mRNA translation, leading to better activities in the WT than the mutants. Moreover, the U^{78} nucleotide could be the isomer, pseudouridine which is known to de-stabilize RNA hairpin loops promoting RNA-RNA structures interaction, and hence it could offer better and dynamic interactions with 18S rRNA h26, leading to better recruitment of the 40S ribosome by the -GUG WT

than -GGG mutant C3. All this again confirms our thinking from all the chimera mutants' activities that the entire SL1 hairpin structure seemed to be very important in mediating translational activities. It probably recruits the 40S ribosome at the 18S rRNA h26 site by reverse rather than direct complementary base-pairing, which is dictated by the U^{78} nucleotide interactions, and supported by the reactive nucleotides U^{76} and A^{80} within the dynamic SL1 hairpin structure.

In addition, other potential cis-regulatory elements or structures within the invSINEB2 secondary structure or surrounding sequences in Δ5'-AS Uchl1 backbone might also play novel roles in the 40S ribosome recruitment or assembly of the overall initiation complex. Taken together, the mutant data indicate that invSINEB2 through the conserved terminal SL1 hairpin-^{76}UGUGA^{80} structure interacts at the18S rRNA h26 ^{1114}UUCCCA^{1119} site by reverse instead of direct complementary base-pairing that is known for the HCV-IRES DIIId structure in recruiting the 40S ribosome.



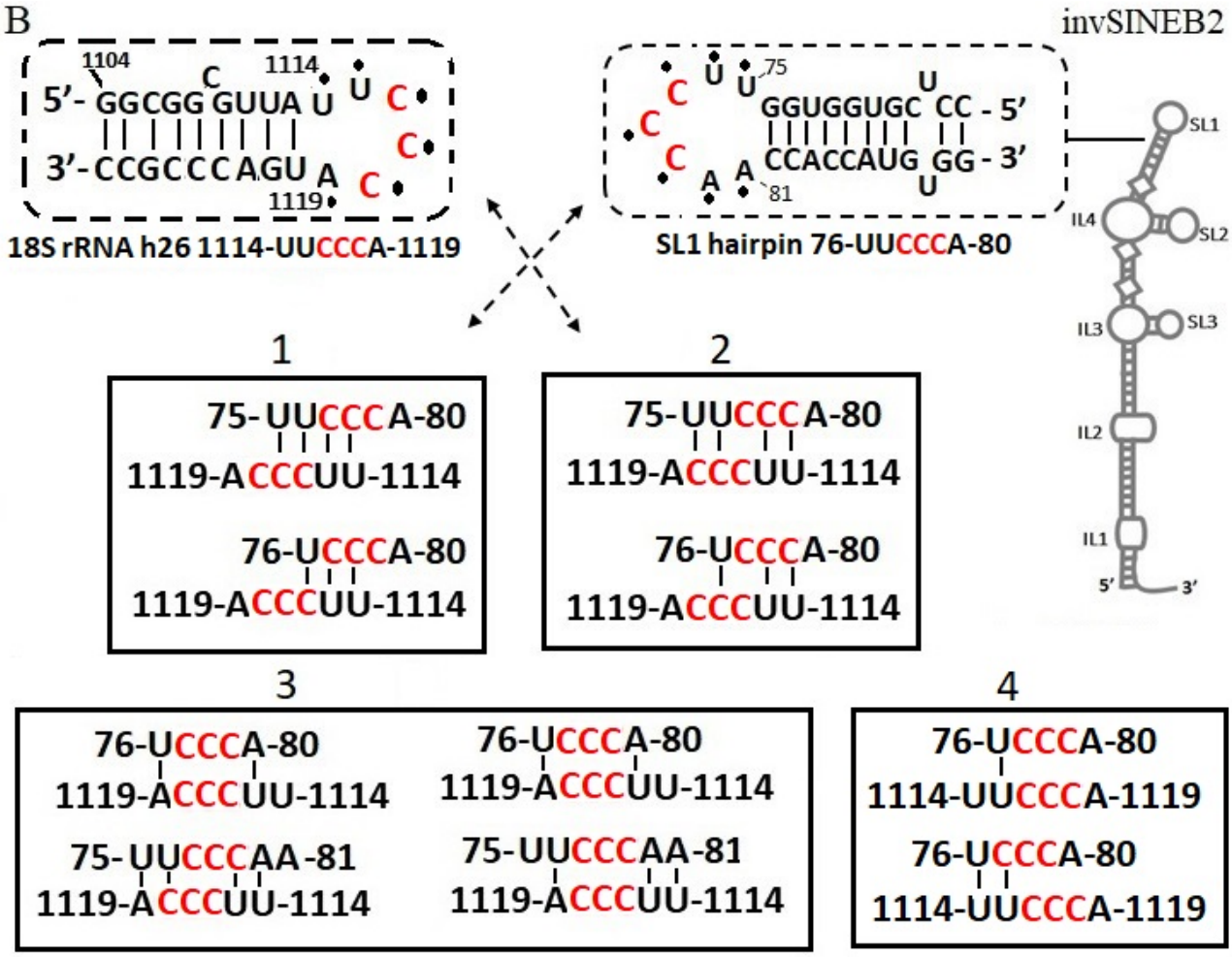


Figure 35 **Reverse complementary base pairing possibilities between invSINEB2 SL1 hairpin and 18S rRNA helix 26 structures.** Schematic representation of invSINEB2 secondary structure. Highlighted with dotted box are SL1 hairpin structure (top, right) and portion of the 18S rRNA helix 26 structure (top, left). The loop region of each highlighted structure shows the potential reactive nucleotides that are marked with dots as revealed by chemical footprinting studies. The main reactive nucleotides are depicted in red letters. SL1 hairpin triplet -^{77}GUG^{79} was hypothesized to bind to the same 18S rRNA helix 26 site as HCV-IRES DIIId-^{266}GGG^{268} structure. (A) The SL1 hairpin-^{74}UUGUGAA^{81} WT with the conserved triplet -^{76}GUG^{78} have two complementary base pairing options at the helix 26 site in the 18S rRNA; direct complementary base pairing with two possible interactions (as depicted in box 1) that is similar to the direct interactions use by HCV-IRES DIIId-^{266}GGG^{268} loop, and reverse complementary base pairing with four possible interactions (as depicted in box 2). This possible interactions are probably dictated by the U^{78} nucleotide, which is absent in the conserved HCV-IRES DIIId-^{266}GGG^{268} loop. (B) The SL1 hairpin-^{74}UUCCCAA^{81} mutant C4 with the conserved triplet -^{76}CCC^{78} will potentially inhibit direct complementary base pairing at the 18S rRNA ^{1114}UUCCCA^{1119} helix 26 site. Only reverse complementary base pairing could possibly occur by eight possible interactions that are categories into 3 groups (box 1, 2 and 3). Categories 1, 2, and 3 respectively represents reverse complementary interactions without any potential gap, one and two/more potential gaps, as shown. Direct complementary base pairing seemed unfavorable as categorized in group 4. U^{75} and A^{80}A^{81} in SL1 hairpin were predicted by chemical footprinting to be

reactive with external agents, suggesting these potential alternative interactions at the helix 26 site in the 18S rRNA

5.6.2 *In Silico* predictions suggest multiple cap-independent ribosome recruitment pathways may mediate SINEUPs translational-activities

The fact that invSINEB2 of AS Uchl1 SINEUP mediates cap-independent translation of Fluc protein synthesis ***(Figure 23Figure*** 25 ***Figure 33)*** in IRES-activity assay, and Uchl1 protein synthesis upon eIF4F-cap complex inactivation by rapamycin does not necessary mean that other potential alternative cap-independent pathways could not exist in AS Uchl1 RNA and other SINEUPs. The ternary complex assembly for the 43S PIC formation is another pathway that is strongly regulated and suggested to be the most rate-limiting step controlling translation initiation upon amino acid starvation, viral infections, etc (Jackson et al., 2010; Sonenberg and Hinnebusch, 2009). Here, I show that invSINEB2 through it conserved SL1 hairpin structure use potentially HCV-IRES like mechanism to recruit the ribosome in promoting eIF4-complex cap-independent pathway. However, based on *in silico* predictions and the mutant data, I propose that there are alternative cap-independent ribosome recruitment pathways which are present in the currently known SINEUP sequences and those yet to be discovered. This is because, SINEUP probably evolved to harbor together with the "IRES-like" TE elements, other cis-regulatory elements that either assist the SINE elements or uniquely regulate translation competent ribosomes to activate translation of targeted mRNAs.

It is reasonable to hypothesize that the surrounding sequences have roles in overall SINEUP translational function. This argument is supported by *in silico* prediction of potential active cis-regulatory motifs such as cytoplasmic polyadenylation elements (CPE) and 3' end poly-A tail elements, terminal oligopoly pyrimidine (TOP) tracts, upstream open reading frames (uORFs), gamma activation of inhibitor of Translation (GAIT), and N6-methyl adenosine (m^6A) conserved in the long ED sequences of SINEUP of AS Uchl1 ***(***

Figure 36) and other SINEUPs, which could also provide additional support in recruiting the ribosome and assembly of the entire translation machinery. For instance, GAITs (Fox, 2015) and uORFs (Hinnebusch, 2005) can cause the release of translation-competent ribosomes prematurely from a transcript, and these complexes could be made available for the synthesis of the sense mRNA due to

S/AS pairing and inter-transcripts communication mediated by recruited RBPs. Therefore, the predicted presence of GAITs and uORFs in SINEUPs suggest that any recruited complexes on the SINEUP RNA are released from it and made available when in complex with the mRNA for translation enhancement. Also, depending on the location within the transcripts, TOPs, m^6A, poly-A tails and CPE motifs strongly define which translation pathways are activated. They can recruit ribosomal complexes or trans-acting factors, acting as ITAFs for IRES sequences in mRNAs, and in this case the invSINEB2 element which then used these recruited complexes for upregulating the sense-mRNA translation in *trans*

Therefore, just like most IRES elements, invSINEB2 requires some trans-acting factors for internal ribosome recruitment and function. This seems to be the case for invSINEB2 of AS Uchl1, since our initial data on IRES-activity involving just invSINEB2 elements recorded very low cap-independent Fluc protein synthesis ***(Figure 23),*** while there was optimal activity when an invSINEB2 sequence with upstream and downstream surrounding sequences containing TOPs, uORFs and m^6A-motifs from AS Uchl1 was used ***(Figure 21 and Figure 23).*** These elements are probably regulating in *trans* (as already explained above) the ability of invSINEB2 SL1 hairpin to contact 40S ribosome by recruiting trans-acting factors that recognize and make conformational changes within the SINEUP ED structures, leading to activity. In particular, I propose the predicted 5'UTR m^6A motif in SINEUP of AS Uchl1 ***(appendix***

Figure 36) as a potential source for eIF3, YTHD2 and METTL3 (Coots et al., 2017; Meyer et al., 2015; Zhou et al., 2015), which uniquely promotes alternative cap-independent translation to act in combinatory way with invSINEB2 to mediate SINEUP translational activities. It will be interesting to disrupt these motifs one at a time and all together through deletion or point mutations and measure translational activities; but as notified, the measured IRES-activity was significantly reduced when TOPs, uORFs and m^6A-motifs upstream invSINEB2 were initially not included in the sequence between Rluc and Fluc bicistronic reporter in the IRES-activity measurement.

A recent study has classified IRES into 2 types; "local" and "global" IRESs where unlike the "global" IRESs which require RNA structures, the "local" IRESs function by very short sequence motifs with complementarity to the 18S rRNA active regions; nucleotides 812 to 1233 including helix 23 and 26 of the 40S ribosome (Weingarten-Gabbay et al., 2016). Most of these short stretch sequences, called "local" IRESs, are polyuridine rich sequence motifs, which are abundantly present and some are classified as TOP motifs in Δ5' AS Uchl1 backbone ***(***

Figure 36) and human SINEUPs of R12A-AS1 and ITFG2 containing FRAM and MIR1b TEs respectively. This suggests that, apart from the WT SL1 hairpin structure of invSINEB2, there might be alternative ways for ribosome recruitment by SINEUPs. Notable, "local IRES" sequence motifs, such as UUCCUUU from poliovirus type 2 and AUUCCUA from simian virus 40 capsid protein 1 (SV40 VP1), have complementarity at the 18S rRNA active regions (Weingarten-Gabbay et al., 2016) and are like the C4 mutant, SL1 hairpin -^{75}UUCCCA^{80} loop sequence, providing further information on why the C4 mutant could retain some activity. I also reported SINEUP activity for the mouse Gtx IRES in human cells when used as ED in SINEUP backbone ***(Figure 26).*** The Gtx IRES has just nine nucleotide (9nt) short sequence (Hu et al., 1999), a "local" IRES-type that directly interact with the 18S rRNA in a Shine-Dalgarno-like mRNA-rRNA complementary interactions. Mouse and human have highly conserved 18S rRNA structures at the "active regions" with, for instance, just one nucleotide outside the reactive apical loop regions in the human h26 that is not present in mouse (Matsuda and Mauro, 2014), making mouse mRNAs to contact the human 18S rRNA h26 and vice versa.

Another potential source of 40S ribosomal complexes recruitment is a second IRES-like element called "IRES-2" predicted within the Δ5' AS Uchl1 backbone that is located downstream to the invSINEB2 sequence containing the "IRES-1" element ***(***

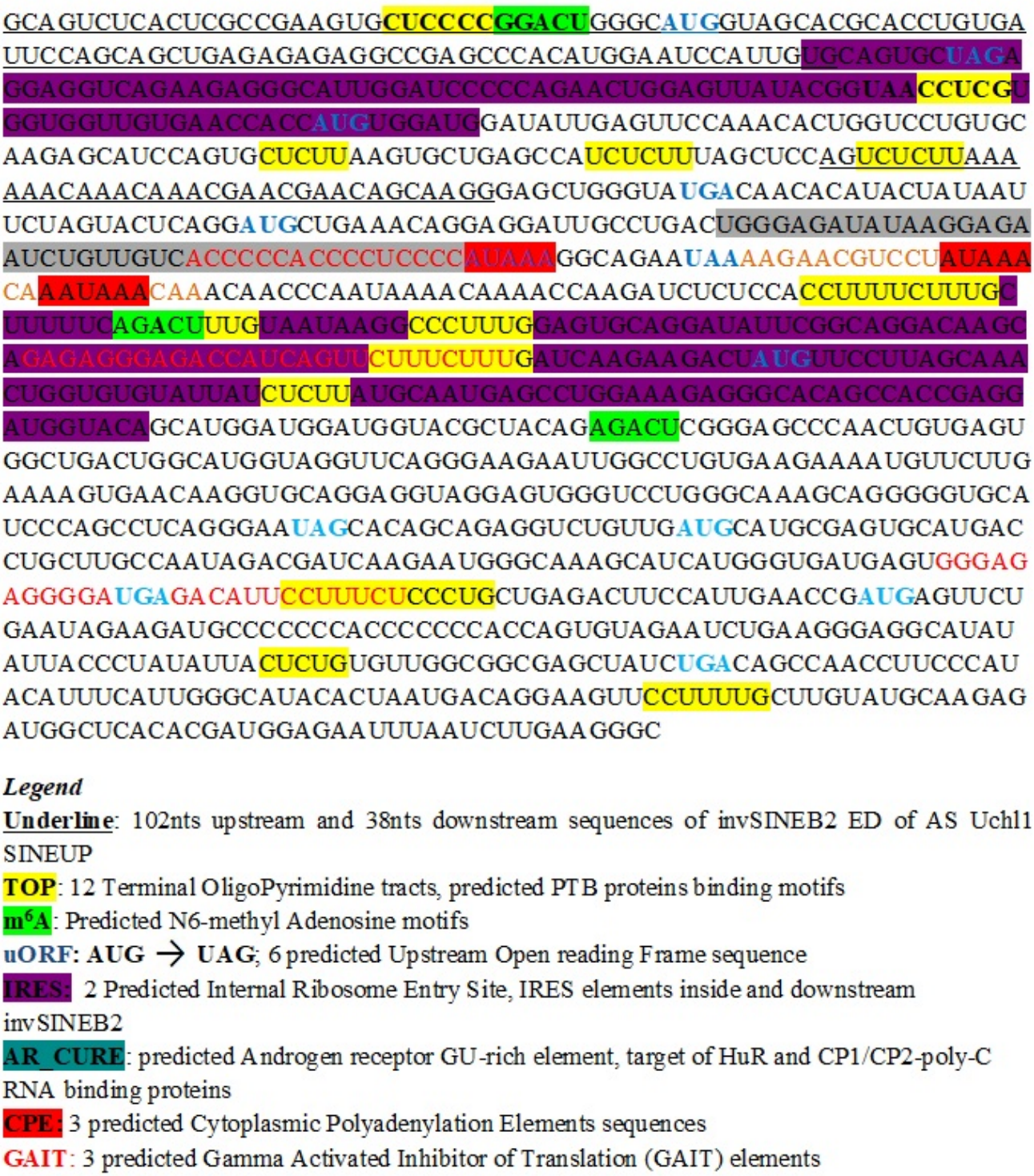
Figure 36, violent legend). Its IRES-activity is yet to be tested experimentally. Due to the presence of other predicted cis-regulatory elements, TOPs, uORFs and m^{6}A-motifs in that sequence, I suggest that the predicted "IRES-2" has IRES-activity, capable of recruiting initiation competent ribosomes for translational activities. Since different IRESs use different pathways in internally recruiting the ribosome, "IRES-2" may be evolutionary conserved to promote alternative cap-independent translation in certain stress when "IRES-1" is inactive. It could even be that both pathways act in synergy when activated to promote cap-independent translation in full-length SINEUPs.

Overall, the recruitment of the translation initiation complexes by SINEUP sequences may also depend on GAITs, poly-A tail, TOP, uORF and m^{6}A cis-regulatory elements interaction with trans-acting factors for the synthesis of targeted proteins. Experimental prove of involvement of "IRES-2", TOP, uORFs and m^{6}A in addition to cytoplasmic poly-A sequences will overall provide deep insights into the full SINEUP-mediated mechanisms, and how lncRNAs like SINEUPs may act as "molecular-hubs" harboring cis-regulatory elements to recruit eIFs, ITAFs, ribosomal complexes through S/AS pairing,

mobilization and inter-transcript communications to promote *trans*-translation upregulation of targeted protein synthesis, driving cellular processes and shaping evolution.

6. APPENDIX

6.1 Supplementary figures



Legend

Underline: 102nts upstream and 38nts downstream sequences of invSINEB2 ED of AS Uchl1 SINEUP

TOP: 12 Terminal OligoPyrimidine tracts, predicted PTB proteins binding motifs

m^{6}A: Predicted N6-methyl Adenosine motifs

uORF: AUG → UAG; 6 predicted Upstream Open reading Frame sequence

IRES: 2 Predicted Internal Ribosome Entry Site, IRES elements inside and downstream invSINEB2

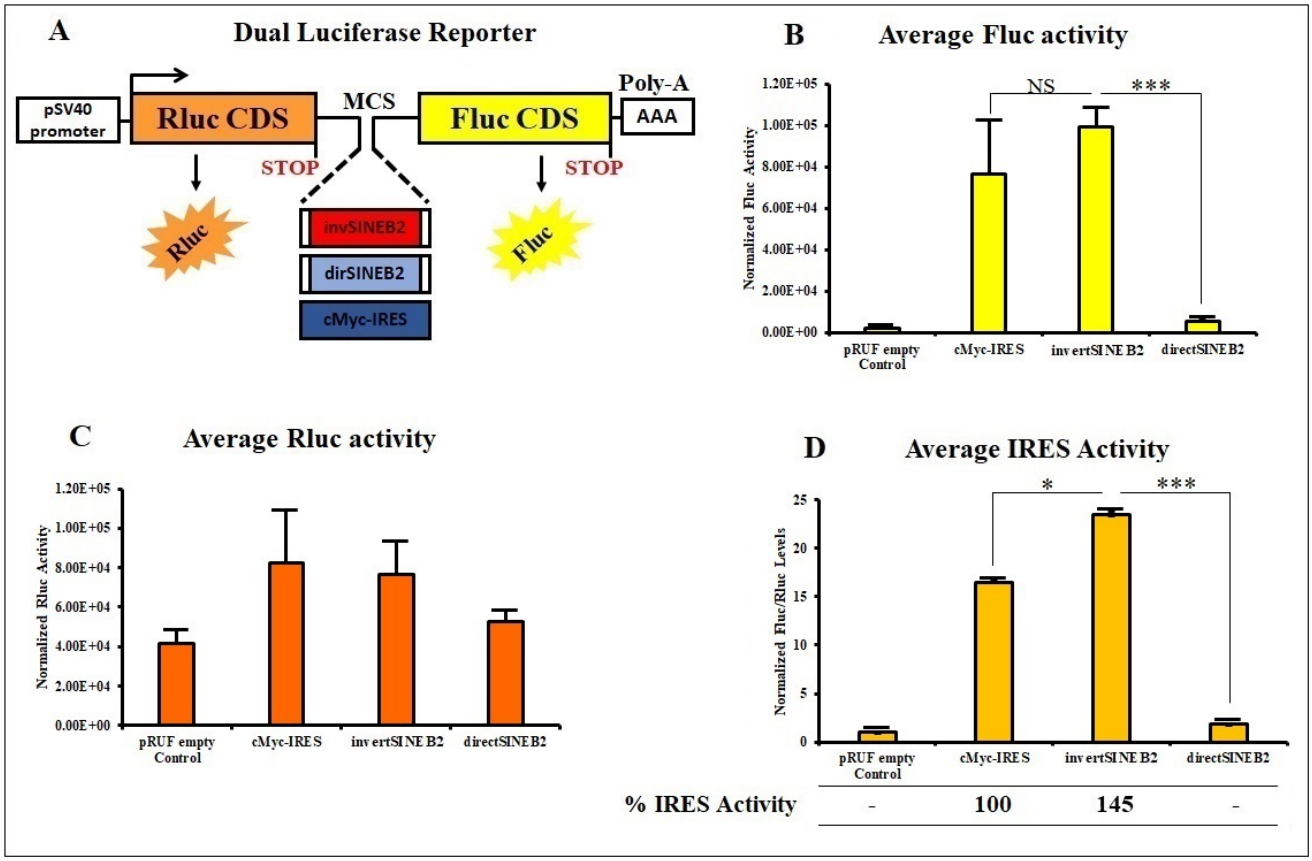
AR_CURE: predicted Androgen receptor GU-rich element, target of HuR and CP1/CP2-poly-C RNA binding proteins

CPE: 3 predicted Cytoplasmic Polyadenylation Elements sequences

GAIT: 3 predicted Gamma Activated Inhibitor of Translation (GAIT) elements

Figure 36 Predicted cis-regulatory elements in SINEUP-ED of AS Uchl1. Snapshot of SINEB2 ED (1,213nts) with underlined upstream (102nts) and downstream (38nts) spacer sequences around invSINEB2, main SINEUP-ED AS Uchl1 with IRES-activity. These spacer sequence when combined with the 167nt invSINEB2 ED, recorded optimal IRES-activity. Highlighted are predicted cis-regulatory elements within the full-length SINEUP-ED sequence. In yellow are predicted TOP sequences which are targets of PTB proteins, green are predicted m^{6}A motifs which can act as CITEs to promoted cap-independent translation activation. In violet is the predicted IRES sequences inside

and downstream invSINEB2 that could mediate translational activation. AUG-TUG, highlights the predicted start and stop codons of an upstream open reading frame (uORF) sequence that could also recruit translation competent ribosomes. In grey highlight is a predicted AR_CURE that is target of HUR CP1/CP2-polyC RNA binding proteins while in red highlights are cytoplasmic polyadenylation elements (CPE) elements downstream invSINEB2. Red letter coded sequence represent predicted gamma activated inhibitor of translation (GAIT) elements. uORF, TOPs and IRES, AR-CURE, CPE and GAIT motifs were predicted with (Chang et al., 2013); http://regrna.mbc.nctu.edu.tw/index1.php) while m^6A motif was predicted with (Zhou et al., 2016; http://www.cuilab.cn/sramp/#predSRAMP).



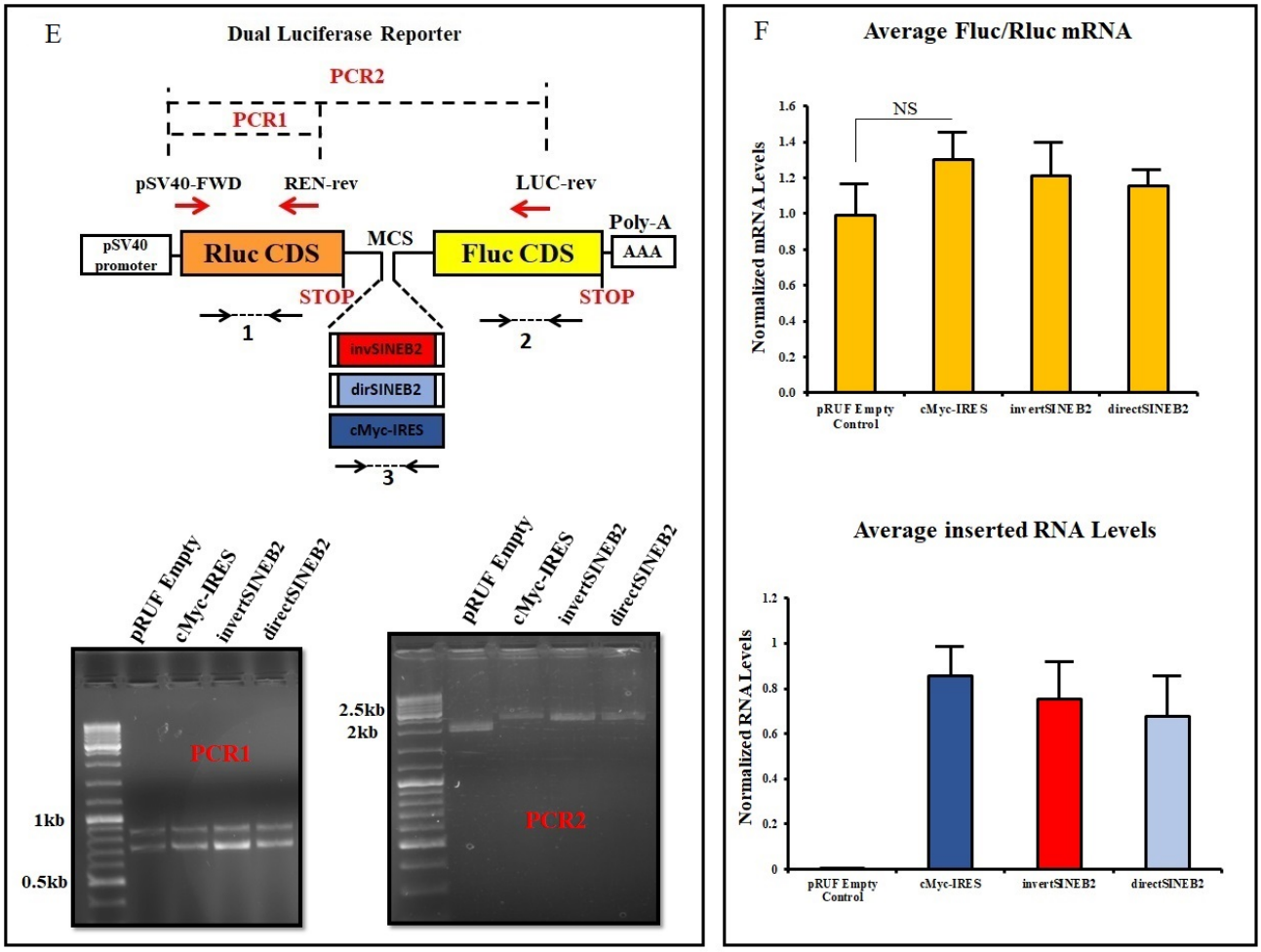
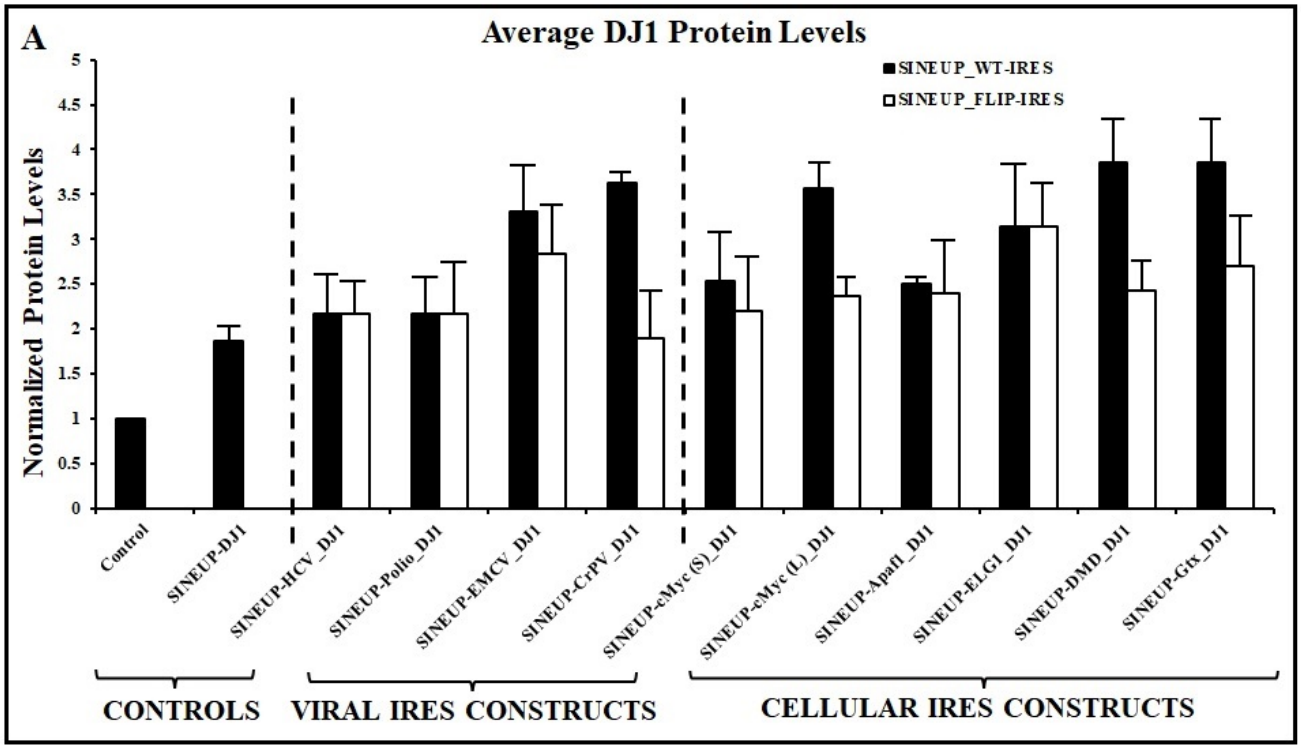
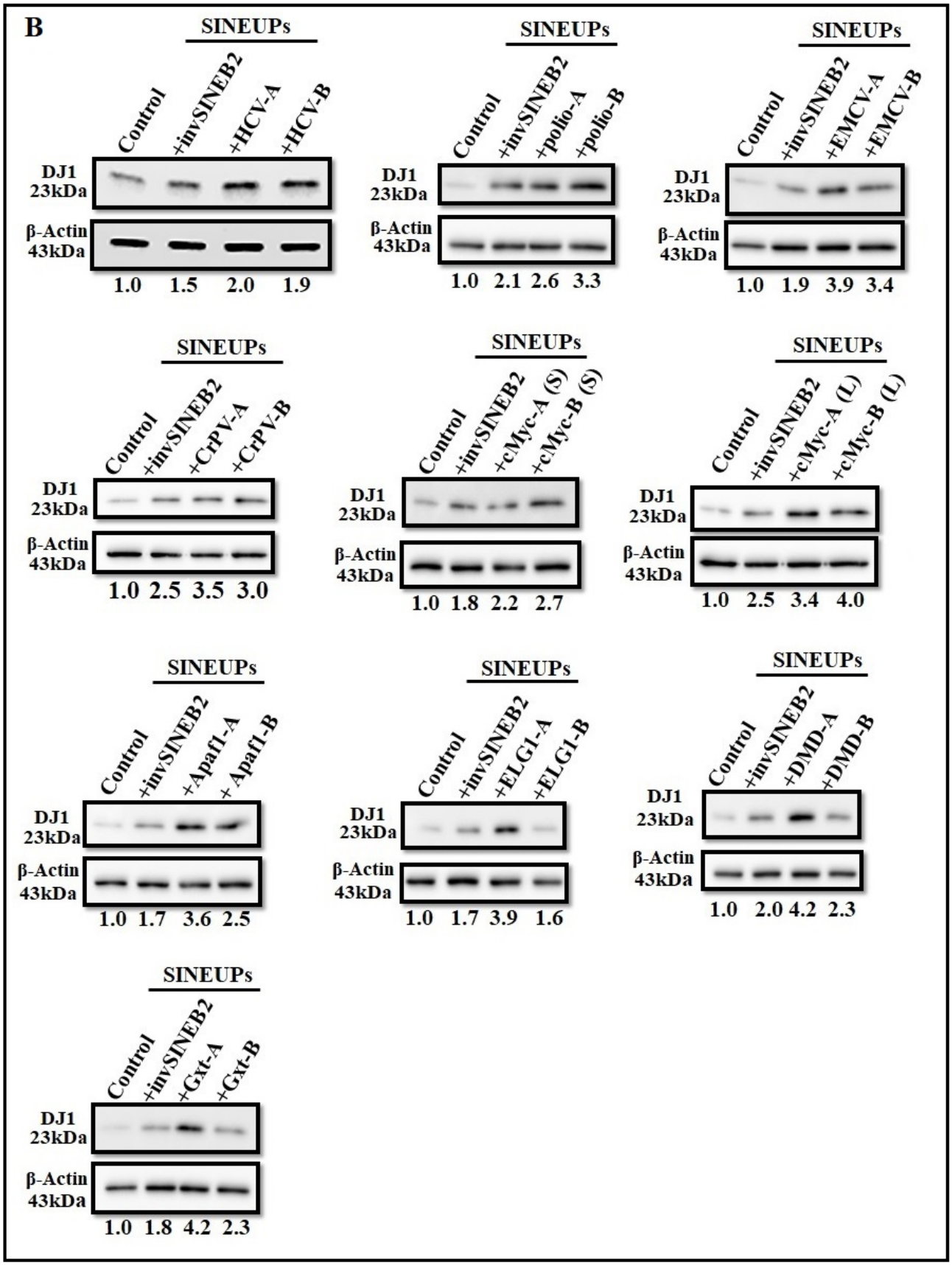


Figure 37 InvSINEB2 with few surrounding sequences in AS Uchl1 has optimal IRES activity in U2OS cells. (A) Schematic representation of pRUF-dual luciferase reporter vectors: containing empty pRUF multiple cloning sites (MCS), cloned inverted (red) and direct SINEB2 (light blue) spacer sequences and c-myc-IRES (blue) sequences flanked by Renilla luciferase (Rluc CDS, brown) cDNA upstream and firefly luciferase (Fluc CDS, yellow) cDNA downstream, all under the control of constitutive pSV40 promoter. Each construct produces a single bicistronic-transcript with Rluc protein translated in a cap-dependent manner, while Fluc protein could be translated in a cap-independent manner. The empty pRUF-MCS and pRUF- c-myc-IRES constructs were used as negative and positive controls, respectively. (B, C). U2OS cells were transiently transfected with the respective constructs, and prepared 48hours for luciferase activities. RNA was also prepared for qRT-PCR to check for alternative promoter and splicing events. Results are shown as average of Rluc (B) and Fluc (C) measured activities in relative light units. (C) Measured Fluc activities are due to Fluc cap-independent protein synthesis induced by the inv- and dir-SINEB2 and c-myc-IRES inserted sequences. (D) IRES-activity was calculated as ratio of Fluc (C) to Rluc (B) measured activities, and normalized against empty pRUF-MCS vector background activities. Empty pRUF activity value was set at 1. c-myc-IRES activity was set to 100%. Percentage IRES-activity were estimated relative to the measured activity of the c-myc-IRES construct. (E) Scheme of qRT-PCR primers design.

Two series of normal PCR 1 and 2, using a pSV40 forward primer that anneal downstream the transcription start site in the pRUF vector. Combining independently with a reverse primer 1, REN-rev and 2, LUC-rev resulted in PCR 1 and 2 products respectively shown in lower panel E. The DNA amplicons are of the expected size, indicating that both reporters were translated from a single bicistronic transcript, annulling

doubts of alternative splicing events. The visible double bands on PCR1 image are short and long amplicons made from process and un-processed transcripts, which arise due to the presence of an intronic sequence downstream the pSV40 promoter of the pRUF vector backbone. (F) qPCR was used to quantify the schematic regions marked by dotted arrows 1, 2 and 3 within Rluc, Fluc and the inserted sequences respectively. The mRNA quantities of Fluc relative to Rluc as normalized by the empty pRUF shows no sign of significantly increase Fluc mRNA, indicating that Fluc proteins were made by cap-independent translation process and hence absence of cryptic promoter events. Error bars show standard deviation (SD) from four biological replicates. All plots indicate mean ± standard deviation, and representation of N=4 independent duplicate-replicas; $p > 0.05$ was considered not significant (NS) $^{*}p < 0.05$ and $^{***}p < 0.001$ were considered significant.


A
Average DJ1 Protein Levels
SINEUP_WT-IRES
SINEUP_FLIP-IRES
Normalized Protein Levels
5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
Control
SINEUP-DJ1
SINEUP-HCV_DJ1
SINEUP-Polio_DJ1
SINEUP-EMCV_DJ1
SINEUP-CrPV_DJ1
SINEUP-cMyc (S)_DJ1
SINEUP-cMyc (L)_DJ1
SINEUP-Apaf1_DJ1
SINEUP-ELG1_DJ1
SINEUP-DMD_DJ1
SINEUP-Gtx_DJ1
CONTROLS
VIRAL IRES CONSTRUCTS
CELLULAR IRES CONSTRUCTS


B
SINEUPs
Control
+invSINEB2
+HCV-A
+HCV-B
DJ1 23kDa
β-Actin 43kDa
1.0 1.5 2.0 1.9
SINEUPs
Control
+invSINEB2
+polio-A
+polio-B
DJ1 23kDa
β-Actin 43kDa
1.0 2.1 2.6 3.3
SINEUPs
Control
+invSINEB2
+EMCV-A
+EMCV-B
DJ1 23kDa
β-Actin 43kDa
1.0 1.9 3.9 3.4
SINEUPs
Control
+invSINEB2
+CrPV-A
+CrPV-B
DJ1 23kDa
β-Actin 43kDa
1.0 2.5 3.5 3.0
SINEUPs
Control
+invSINEB2
+cMyc-A (S)
+cMyc-B (S)
DJ1 23kDa
β-Actin 43kDa
1.0 1.8 2.2 2.7
SINEUPs
Control
+invSINEB2
+cMyc-A (L)
+cMyc-B (L)
DJ1 23kDa
β-Actin 43kDa
1.0 2.5 3.4 4.0
SINEUPs
Control
+invSINEB2
+Apaf1-A
+Apaf1-B
DJ1 23kDa
β-Actin 43kDa
1.0 1.7 3.6 2.5
SINEUPs
Control
+invSINEB2
+ELG1-A
+ELG1-B
DJ1 23kDa
β-Actin 43kDa
1.0 1.7 3.9 1.6
SINEUPs
Control
+invSINEB2
+DMD-A
+DMD-B
DJ1 23kDa
β-Actin 43kDa
1.0 2.0 4.2 2.3
SINEUPs
Control
+invSINEB2
+Gxt-A
+Gxt-B
DJ1 23kDa
β-Actin 43kDa
1.0 1.8 4.2 2.3

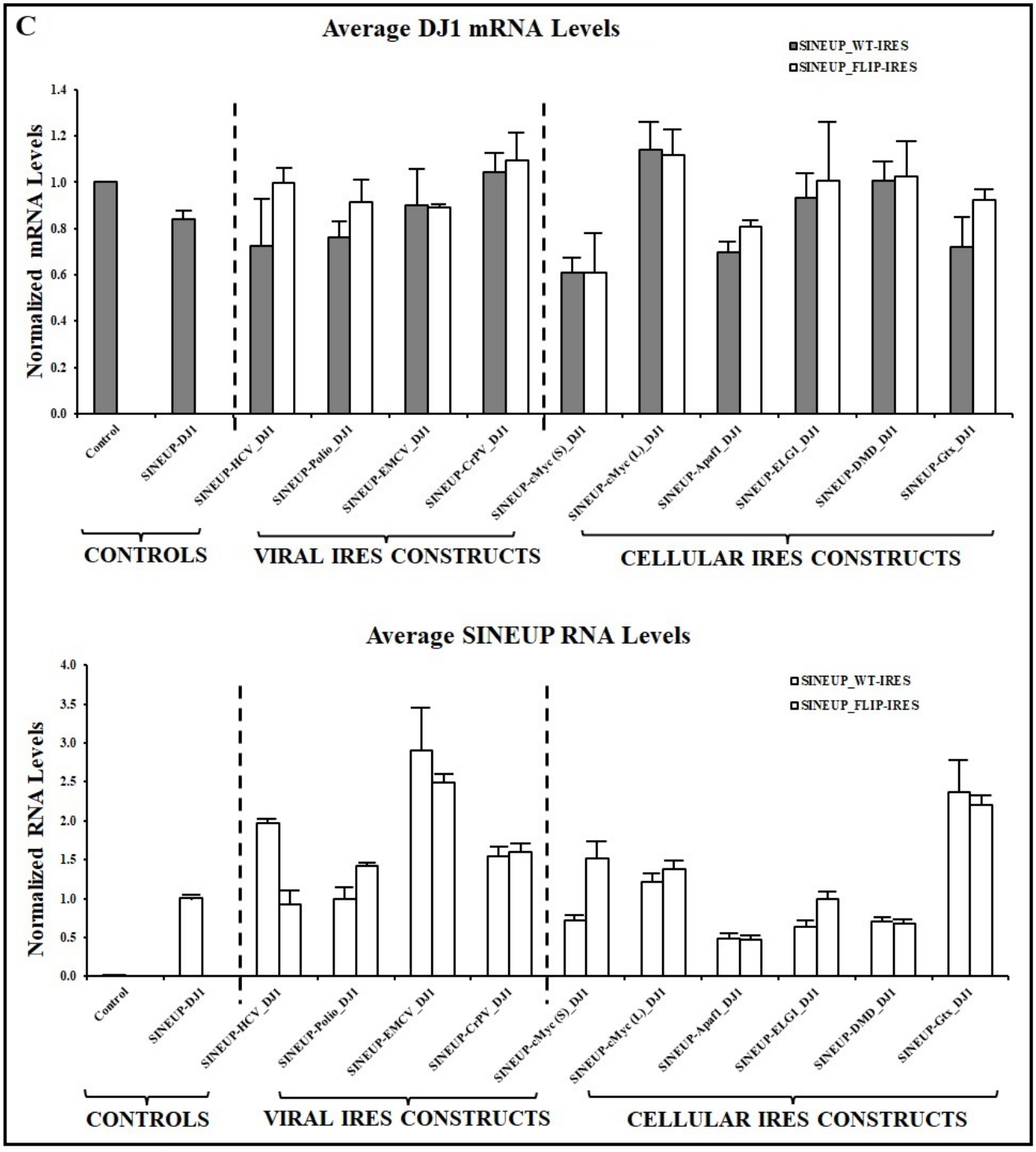


Figure 38 Synthetic SINEUPs of viral and cellular IRESs increase DJ1 proteins in HepG2 cells. SINEUP with substituted cellular and viral IRES sequences as EDs targeting DJ1 mRNA are as describe before. **(A and B)** The SINEUP constructs were transiently expressed here in HepG2 cell lines and SINEUP-activities measured 48 hours post-transfection by WB analysis, using anti-DJ1 antibody to detect and quantify DJ1 proteins as fold change of no SINEUP-empty pCS2+ vector. β-Actin was used as normalized loading controls. Mean DJ1 fold change plot of at least 5 independent replicates are shown (black bars for WT ED constructs and white bars for flip IRES ED constructs,

panel A).. (B) Represented western blot analysis showing DJ1 protein fold change band intensities, normalized to β-Actin loading control. (D) RT-qPCR was used to quantify DJ1 mRNA (grey bars for the SINEUP with WT EDs and white bar for SINEUP with flip IRES EDS, upper panel) and compared with no SINEUP-empty vector control, which shows no statistical significant levels (mean ± S.D, $p<0.05$, sample t.Test vs empty pCS2+ plasmid control). This indicates post-transcriptional upregulation of DJ1 protein synthesis that was mediated by the SINEUP constructs transfected to HepG2 cell lines. Synthetic SINEUP RNA sequences (white plot, lower panel B) were also detected and measured by RT-qPCR with primers designed to the 3'tail of Δ5'-ASUchl1 backbone sequence.

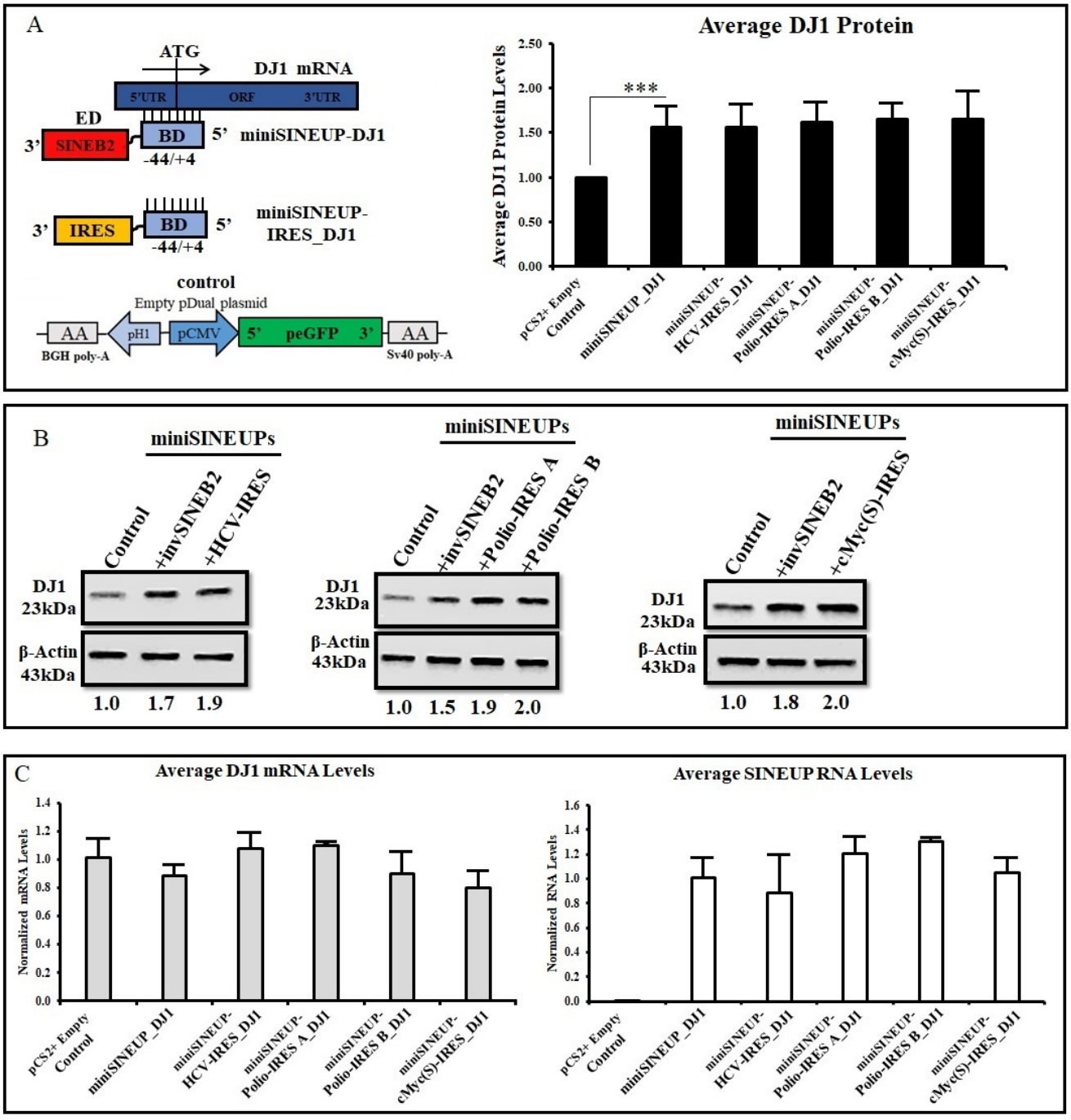


Figure 39 **Viral and Cellular IRESs are portable SINEUP EDs to increase synthesis of DJ1 protein when expressed in a pDual-peGFP plasmid. (A)** Structural organization of miniSINEUPs with 44 nucleotide BD, designed to target endogenous expressed DJ1 mRNA (blue scheme) but not peGFP expressed from the same pDual-peGFP plasmid construct. MiniSINEUP-DJ1 with invSINEB2 ED (Red scheme) and no SINEUP-empty pDual-peGFP construct were used as experimental positive and negative controls respectively. MiniSINEUP-IRES_DJ1 IRES EDs are also depicted (Orange bar). DJ1 protein increasing ability of each unique IRES sequence as portable SINEUP-ED was tested in HEK 293T cells upon 48hours transient transfection. Mean DJ1-protein intensity plots (black bars, panel A right) for each miniSINEUP/IRES activities quantified from at least five independent replicas

with WB analysis (B) as normalized to fold change of empty pDual-peGFP control is indicated. β-Actin was used as WB loading control. (D) RT-qPCR analysis as described before showed stable DJ1 mRNA (grey bar, left), indicating post-transcription effects. Both invSINEB2 and each IRES EDs RNA (white bar, right) could also be detected and measured by RT-qPCR with respective primers (Supplementary table S2). Data indicate mean ± SD of at least 5 independent biological replicas; $^{***}p < 0.001$ was considered significant.

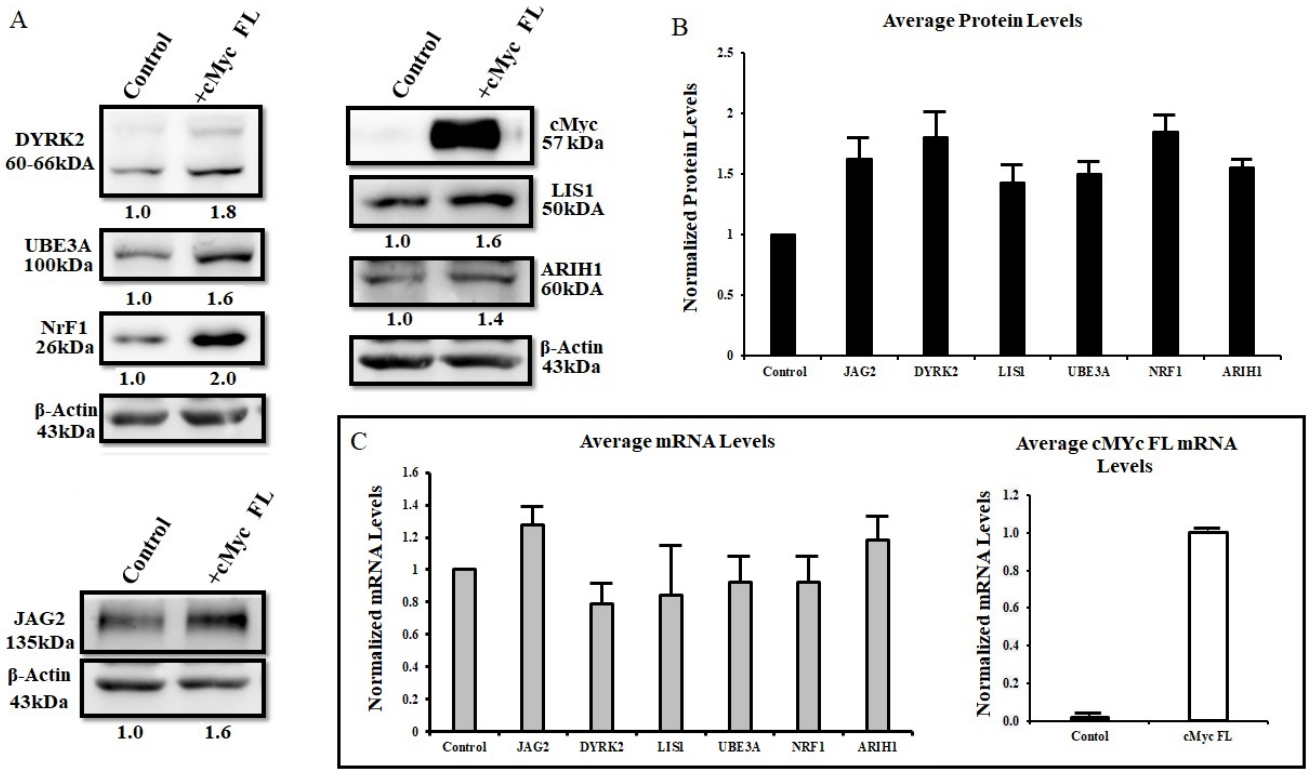


Figure 40 IRES-containing full-length c-myc cDNA has SINEUP-like function in *trans* in HEK cells. c-myc full-length transcript cDNA was used to transiently transfect U2OS cells. Transfected cells were harvested 48 hours and processed for detecting SINEUP-like activity on targeted genes, as performed in HEK cells (figure 15). Control cells were transfected with an empty pcDNA3.1- plasmid used in expressing c-myc Full length cDNA. (A) Western blotting of targeted JAG2, DYRK2, UBE3A, LIS1, NRF1 and ARIH1 proteins were detected using respective antibodies. The quantification was normalized to β-Actin that was used as loading control. (B) Average targeted protein fold change quantification normalized to the empty plasmid control. Empty pcDNA3.1- plasmid transfection was arbitrarily set to 1. (C) Expression of targeted mRNA (grey plot, left) and c-myc full-length constructs RNA (white plot, left) were monitored by qRT-PCR with respective specific primers. All plots indicate mean ± standard deviation, and representation of N=3 independent replicates $^{*}p < 0.05$, $^{**}p < 0.01$, NS; not significant

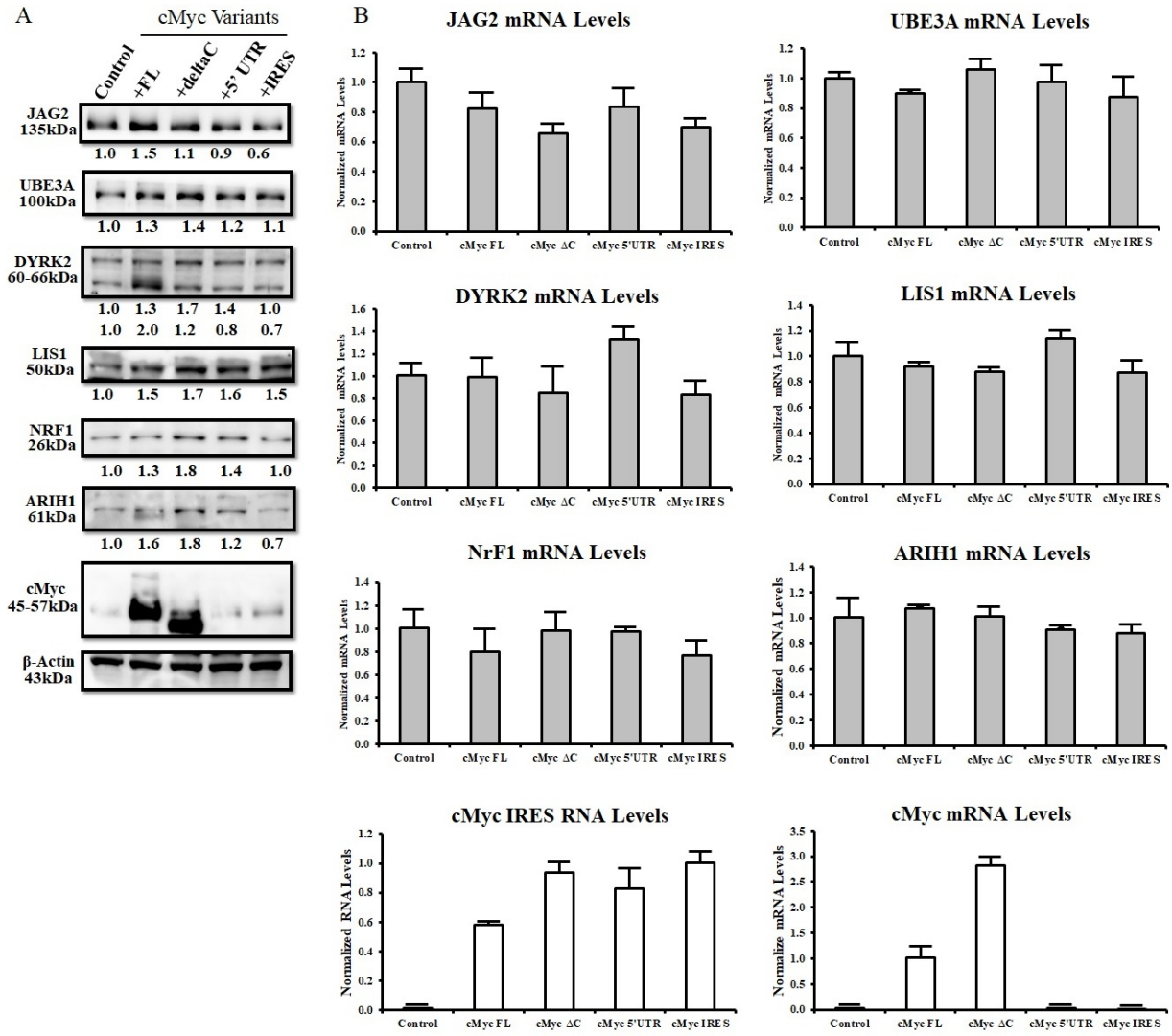
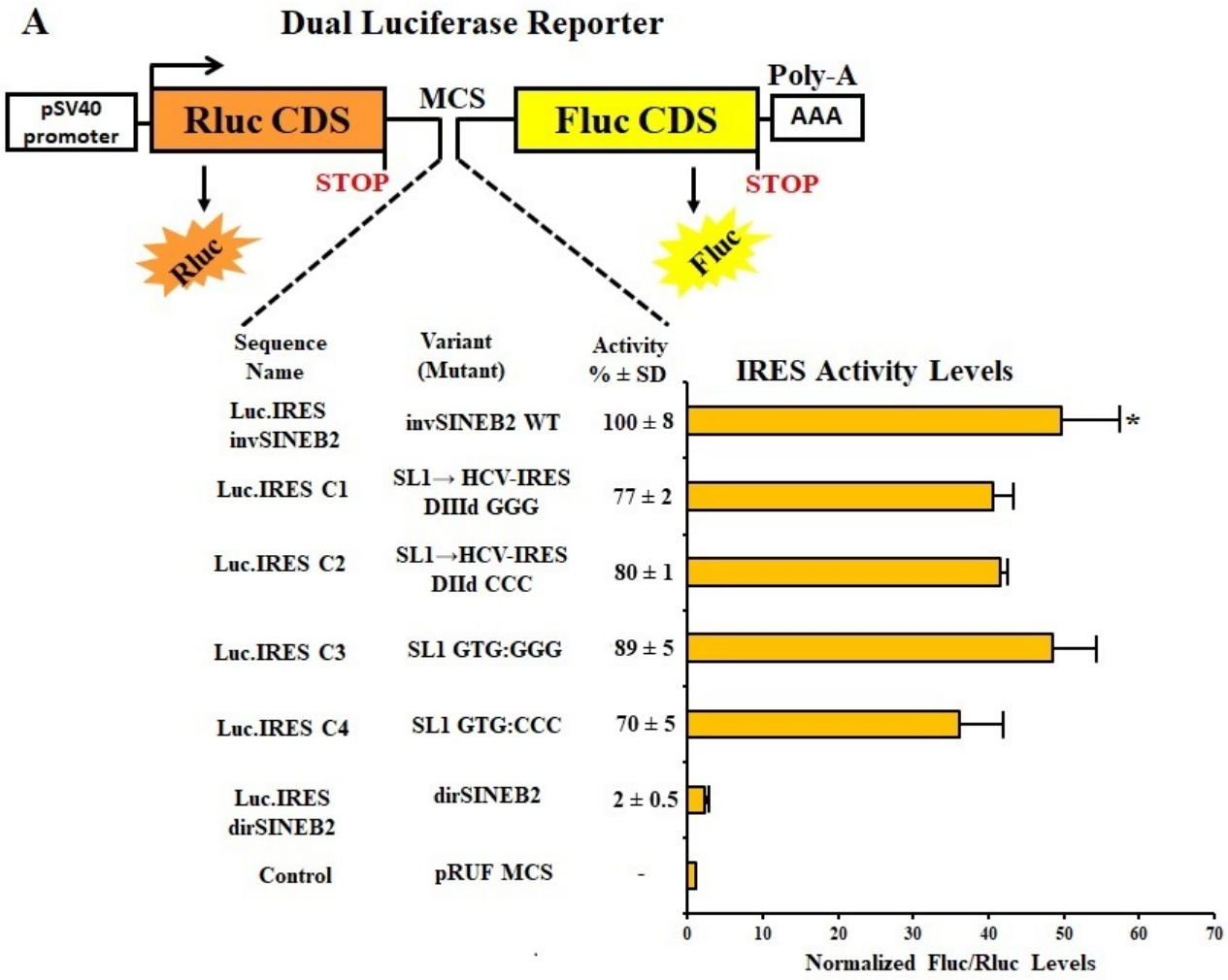


Figure 41 c-myc 5' UTR has novel role in increasing other proteins synthesis post-transcriptionally in cell lines in vitro. The c-myc full-length transcript and mutants cDNA described in figure 16 were used to transiently transfect HEK 293T cells. Transfected cells were harvested 48 hours and processed for detecting SINEUP-like activity in trans on targeted genes. Control HEK 293T cells were transfected with an empty pcDNA3.1- plasmid. (A) Western blotting of targeted JAG2, DYRK2, UBE3A, LIS1, NRF1 and ARIH1 proteins were detected using respective antibodies. The quantification was normalized to β-Actin that was used as loading control. (B) Expression of targeted mRNAs (grey bar plots, upper panel) and c-myc full-length and IRES RNA (white bar plots, lower panel) were monitored by qRT-PCR with respective specific primers. All plots indicate mean ± standard deviation, and representation of N=5 independent replicates, NS; not significant


A
Dual Luciferase Reporter
pSV40 promoter
Rluc CDS
MCS
Fluc CDS
Poly-A
AAA
STOP
STOP
Rluc
Fluc
Sequence Name
Variant (Mutant)
Activity % ± SD
IRES Activity Levels
Luc.IRES invSINEB2
invSINEB2 WT
100 ± 8
*
Luc.IRES C1
SL1→ HCV-IRES DIIId GGG
77 ± 2
Luc.IRES C2
SL1→HCV-IRES DIId CCC
80 ± 1
Luc.IRES C3
SL1 GTG:GGG
89 ± 5
Luc.IRES C4
SL1 GTG:CCC
70 ± 5
Luc.IRES dirSINEB2
dirSINEB2
2 ± 0.5
Control
pRUF MCS
-
0
10
20
30
40
50
60
70
Normalized Fluc/Rluc Levels

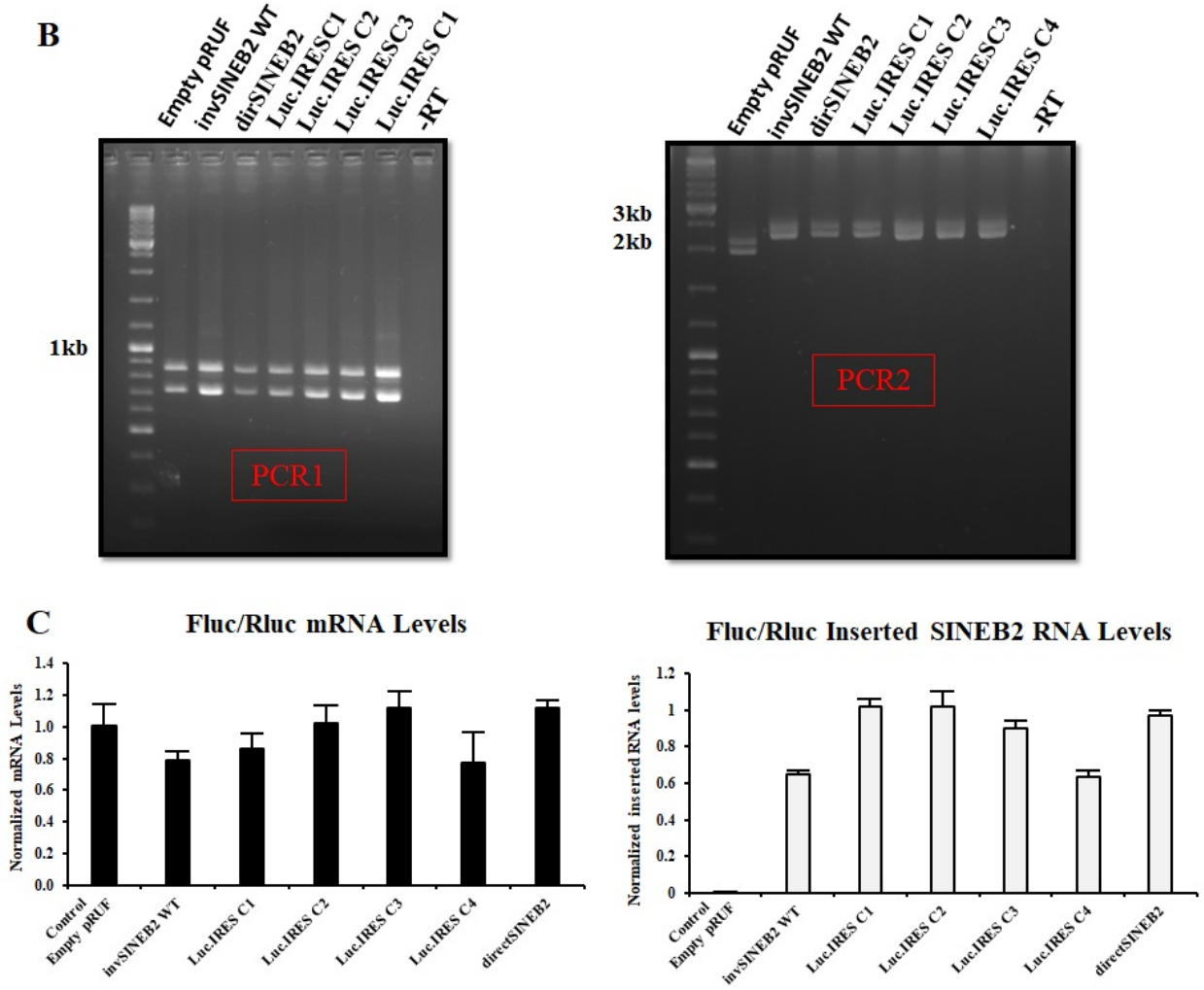


Figure 42 invSINEB2 SL1 hairpin mutants inhibits IRES-activity in U2OS cell lines. (A) Schematic representation of pRUF-dual luciferase reporter vectors: containing empty pRUF multiple cloning sites (MSC), cloned inverted- and direct-SINEB2 spacer sequences and chimera mutants, Luc.IRES C1-C4 sequences flanked by Renilla luciferase coding sequence (Rluc CDS, brown) cDNA upstream and firefly luciferase coding sequence (Fluc CDS, yellow) cDNA downstream and are all under the control of constitutive pSV40 promoter. Each construct produces a single bicistronic-transcript with Rluc protein translated in a cap-dependent manner, while Fluc protein could be translated in a cap-independent manner. Luc.IRES C1-C4 are the invSINEB2 chimera mutants C1-C4 sequences used as SINEUP-EDs in SINEUP activity assay. The empty pRUF-MCS and pRUF-invSINEB2-spacer WT constructs were used as negative and positive controls respectively. U2OS cells were transiently transfected with the respective individual pRUF reporter constructs, and prepared 48hours for luciferase activities. RNA was also prepared for qRT-PCR to check for alternative promoter and splicing events that could be mediated by any of the inserted sequences. (A) Results are shown as mean relative IRES-activity (Fluc/Rluc) levels. Measured Rluc-activity is due to cap-dependent translation of Rluc protein while Fluc activities are due to Fluc cap-independent protein synthesis mediated by the inserted sequences. IRES-activity was calculated as ratio of Fluc to Rluc

measured activities in relative light units, and normalized against empty pRUF-MCS vector background activities. Empty pRUF activity value was set at 1. InvSINEB2 WT activity was set to 100% ±SD. Percentage IRES activity were estimated relative to the WT activity. (B) Electropherogram result of two series of PCR 1 and 2, performed using a pSV40 forward primer that anneal downstream the transcription start site in the pRUF vector, combining independently with a reverse primer 1 in Rluc and primer 2 in Fluc as described previously (figure 12 and 22). The DNA amplicons are of the expected size, indicating that both reporters were translated from a single bicistronic transcript, annulling possibilities of alternative splicing events. The visible double bands on PCR 1 and 2 images are short and long amplicons that represent process and unprocessed transcripts, which arise due to the presence of an intronic sequence downstream the pSV40 promoter of the pRUF vector backbone. (C) qPCR was used to quantify Rluc and Fluc mRNAs, and the inserted RNA sequences . The mRNA quantities of Fluc relative to Rluc as normalized by the empty pRUF shows no sign of significantly increased Fluc mRNA, indicating that Fluc proteins were made by cap-independent translation process (A) and this suggest absence of cryptic promoter events. Error bars show standard deviation (SD) from four biological replicates. All plots indicate mean ± SD, and representation of N=4 independent duplicate-replicas; $p > 0.05$ was considered not significant (NS)

6.2 Supplementary tables

Table 1: Sequence of primers used in the miniSINEUP-DJ, miniSINEUP-IRES-DJ1, miniSINEUP_HCV-IRES-GFP and pRUF-SINEB2 (direct and inverted) clones

Name of SINEUP	Forward Primer 5'=>3'	Reverse Primers 5'=>3'
miniSINEUP-DJ1	ATATACTCGAGCCATTTTTATGTTATATG	GCGCAAGCTTGGAGCTAAAGAGATGGCTCAG
miniSINEUP_HCV-IRES001ADJ1	"	GCGCAAGCTTGTTACGTTTGGTTTTTCTTTGAGG
miniSINEUP_Polio-IRES002ADJ1	"	GCGCAAGCTTGGCCAATCCAATTCGCTTTATG
miniSINEUP_Polio-IRES002BDJ1	"	GCGCAAGCTTATGAGTCTGGACATCCCTCAC
miniSINEUP_Myc-IRES003ADJ1	"	GCGCAAGCTTAGAGTCGCGTCCTTGCTCGG
miniSINEUP_HCV-IRES001A-GFP	ATATAGAATTCGCCCTTGCCAGCCC	GCGCAAGCTTGTTACGTTTGGTTTTTCTTTGAGG
SINEUP-DJ1BDΔED	FWD1: AGTCTCTTAAAAAACAAACAAACG FWD2: ATATACTCGAGCCATTTTTATGTTATATGTTTACAAGCCCCACACCAGGCTGAAAAGTCTCTTAAAAAACAAACAAACG	GCGCTCTAGAAAGCTTGGTACCGAG
pRUF-invertSINEB2	ATATAGAATTCGCAGTCTCACTCGCCGAAG	GCGCCTCGAGCCTTGCTGTTCGTTCGTTTG

pRUF-directSINEB2	ATATAGAATTCCCTTGCTGTTCGTTCGTTTG	GCGCCTCGAGGCAGTCTCACTCGCCGAAG

Underline = non-homology tails containing restriction endonuclease sites

7. BIBLIOGRAPHY

Aagaard, L., and Rossi, J.J. (2007). RNAi therapeutics: principles, prospects and challenges. Adv. Drug Deliv. Rev. *59*, 75–86.

Aitken, C.E., and Lorsch, J.R. (2012). A mechanistic overview of translation initiation in eukaryotes. Nat. Struct. Mol. Biol. *19*, 568–576.

Akulich, K.A., Andreev, D.E., Terenin, I.M., Smirnova, V.V., Anisimova, A.S., Makeeva, D.S., Arkhipova, V.I., Stolboushkina, E.A., Garber, M.B., Prokofjeva, M.M., et al. (2016). Four translation initiation pathways employed by the leaderless mRNA in eukaryotes. Sci. Rep. *6*.

Ali, I.K., McKendrick, L., Morley, S.J., and Jackson, R.J. (2001). Activity of the Hepatitis A Virus IRES Requires Association between the Cap-Binding Translation Initiation Factor (eIF4E) and eIF4G. J. Virol. *75*, 7854–7863.

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.

Amit, M., Donyo, M., Hollander, D., Goren, A., Kim, E., Gelfman, S., Lev-Maor, G., Burstein, D., Schwartz, S., Postolsky, B., et al. (2012). Differential GC content between exons and introns establishes distinct strategies of splice-site recognition. Cell Rep. *1*, 543–556.

Aw, J.G.A., Shen, Y., Wilm, A., Sun, M., Lim, X.N., Boon, K.-L., Tapsin, S., Chan, Y.-S., Tan, C.-P., Sim, A.Y.L., et al. (2016). In Vivo Mapping of Eukaryotic RNA Interactomes Reveals Principles of Higher-Order Organization and Regulation. Mol. Cell *62*, 603–617.

Baird, S.D., Lewis, S.M., Turcotte, M., and Holcik, M. (2007). A search for structurally similar cellular internal ribosome entry sites. Nucleic Acids Res. *35*, 4664–4677.

Banerjee, A.K. (1980). 5'-terminal cap structure in eucaryotic messenger ribonucleic acids. Microbiol. Rev. *44*, 175–205.

Barreau, C., Paillard, L., and Osborne, H.B. (2005). AU-rich elements and associated factors: are there unifying principles? Nucleic Acids Res. *33*, 7138–7150.

Barrett, L.W., Fletcher, S., and Wilton, S.D. (2012). Regulation of eukaryotic gene expression by the untranslated gene regions and other non-coding elements. Cell. Mol. Life Sci. CMLS *69*, 3613–3634.

Barría, M.I., González, A., Vera-Otarola, J., León, U., Vollrath, V., Marsac, D., Monasterio, O., Pérez-Acle, T., Soza, A., and López-Lastra, M. (2009). Analysis of natural variants of the hepatitis C virus internal ribosome entry site reveals that primary sequence plays a key role in cap-independent translation. Nucleic Acids Res. *37*, 957–971.

Bartel, D.P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. Cell *116*, 281–297.

Bisio, A., Nasti, S., Jordan, J.J., Gargiulo, S., Pastorino, L., Provenzani, A., Quattrone, A., Queirolo, P., Bianchi-Scarrà, G., Ghiorzo, P., et al. (2010). Functional analysis of CDKN2A/p16INK4a 5'-UTR variants predisposing to melanoma. Hum. Mol. Genet. *19*, 1479–1491.

Bisio, A., Latorre, E., Andreotti, V., Bressac-de Paillerets, B., Harland, M., Scarra, G.B., Ghiorzo, P., Spitale, R.C., Provenzani, A., and Inga, A. (2015). The 5'-untranslated region of p16INK4a melanoma tumor suppressor acts as a cellular IRES, controlling mRNA translation under hypoxia through YBX1 binding. Oncotarget *6*, 39980–39994.

Borodulina, O.R., and Kramerov, D.A. (2008). Transcripts synthesized by RNA polymerase III can be polyadenylated in an AAUAAA-dependent manner. RNA *14*, 1865–1873.

Calvo, S.E., Pagliarini, D.J., and Mootha, V.K. (2009). Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. Proc. Natl. Acad. Sci. U. S. A. *106*, 7507–7512.

Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M.C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C., et al. (2005). The transcriptional landscape of the mammalian genome. Science *309*, 1559–1563.

Carrieri, C., Cimatti, L., Biagioli, M., Beugnet, A., Zucchelli, S., Fedele, S., Pesce, E., Ferrer, I., Collavin, L., Santoro, C., et al. (2012). Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat. Nature *491*, 454–457.

Carrieri, C., Forrest, A.R.R., Santoro, C., Persichetti, F., Carninci, P., Zucchelli, S., and Gustincich, S. (2015). Expression analysis of the long non-coding RNA antisense to Uchl1 (AS Uchl1) during dopaminergic cells' differentiation in vitro and in neurochemical models of Parkinson's disease. Front. Cell. Neurosci. *9*, 114.

Casa, V., and Gabellini, D. (2012). A repetitive elements perspective in Polycomb epigenetics. Front. Genet. *3*.

Chang, T.-H., Huang, H.-Y., Hsu, J.B.-K., Weng, S.-L., Horng, J.-T., and Huang, H.-D. (2013). An enhanced computational platform for investigating the roles of regulatory RNA and for identifying functional RNA motifs. BMC Bioinformatics *14 Suppl 2*, S4.

Chen, J.-J. (2014). Translational Control by Heme-Regulated eIF2α Kinase during Erythropoiesis. Curr. Opin. Hematol. *21*, 172–178.

Chen, J., Sun, M., Kent, W.J., Huang, X., Xie, H., Wang, W., Zhou, G., Shi, R.Z., and Rowley, J.D. (2004). Over 20% of human transcripts might form sense–antisense pairs. Nucleic Acids Res. *32*, 4812–4820.

Cheng, Y., Ma, Z., Kim, B.-H., Wu, W., Cayting, P., Boyle, A.P., Sundaram, V., Xing, X., Dogan, N., Li, J., et al. (2014). Principles of regulatory information conservation between mouse and human. Nature *515*, 371–375.

Chong, R.H.E., Gonzalez-Gonzalez, E., Lara, M.F., Speaker, T.J., Contag, C.H., Kaspar, R.L., Coulman, S.A., Hargest, R., and Birchall, J.C. (2013). Gene silencing following siRNA delivery to skin via coated steel microneedles: In vitro and in vivo proof-of-concept. J. Control. Release Off. J. Control. Release Soc. *166*, 211–219.

Chu, C., Qu, K., Zhong, F., Artandi, S.E., and Chang, H.Y. (2011). Genomic maps of lincRNA occupancy reveal principles of RNA-chromatin interactions. Mol. Cell *44*, 667–678.

Churakov, G., Sadasivuni, M.K., Rosenbloom, K.R., Huchon, D., Brosius, J., and Schmitz, J. (2010). Rodent evolution: back to the root. Mol. Biol. Evol. *27*, 1315–1326.

Coots, R.A., Liu, X.-M., Mao, Y., Dong, L., Zhou, J., Wan, J., Zhang, X., and Qian, S.-B. (2017). m6A Facilitates eIF4F-Independent mRNA Translation. Mol. Cell *68*, 504-514.e7.

Costantino, D.A., Pfingsten, J.S., Rambo, R.P., and Kieft, J.S. (2008). tRNA–mRNA mimicry drives translation initiation from a viral IRES. Nat. Struct. Mol. Biol. *15*, 57–64.

Damgaard, C.K., and Lykke-Andersen, J. (2011). Translational coregulation of 5'TOP mRNAs by TIA-1 and TIAR. Genes Dev. *25*, 2057–2068.

Deng, Y., Wang, C.C., Choy, K.W., Du, Q., Chen, J., Wang, Q., Li, L., Chung, T.K.H., and Tang, T. (2014). Therapeutic potentials of gene silencing by RNA interference: principles, challenges, and new strategies. Gene *538*, 217–227.

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

Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., Guernec, G., Martin, D., Merkel, A., Knowles, D.G., et al. (2012). The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res. *22*, 1775–1789.

Dinger, M.E., Pang, K.C., Mercer, T.R., and Mattick, J.S. (2008). Differentiating protein-coding and noncoding RNA: challenges and ambiguities. PLoS Comput. Biol. *4*, e1000176.

Djebali, S., Davis, C.A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F., et al. (2012). Landscape of transcription in human cells. Nature *489*, 101–108.

Dmitriev, S.E., Terenin, I.M., Andreev, D.E., Ivanov, P.A., Dunaevsky, J.E., Merrick, W.C., and Shatsky, I.N. (2010). GTP-independent tRNA Delivery to the Ribosomal P-site by a Novel Eukaryotic Translation Factor. J. Biol. Chem. *285*, 26779–26787.

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.

Dragon, F., Lemay, V., and Trahan, C. (2006). snoRNAs: Biogenesis, Structure and Function. p.

Ekram, M.B., Kang, K., Kim, H., and Kim, J. (2012). Retrotransposons as a major source of epigenetic variations in the mammalian genome. Epigenetics *7*, 370–382.

Elbarbary, R.A., Lucas, B.A., and Maquat, L.E. (2016). Retrotransposons as regulators of gene expression. Science *351*, aac7247.

Elisaphenko, E.A., Kolesnikov, N.N., Shevchenko, A.I., Rogozin, I.B., Nesterova, T.B., Brockdorff, N., and Zakian, S.M. (2008). A dual origin of the Xist gene from a protein-coding gene and a set of transposable elements. PloS One *3*, e2521.

ENCODE Project Consortium, Birney, E., Stamatoyannopoulos, J.A., Dutta, A., Guigó, R., Gingeras, T.R., Margulies, E.H., Weng, Z., Snyder, M., Dermitzakis, E.T., et al. (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature *447*, 799–816.

Engreitz, J.M., Sirokman, K., McDonel, P., Shishkin, A.A., Surka, C., Russell, P., Grossman, S.R., Chow, A.Y., Guttman, M., and Lander, E.S. (2014). RNA-RNA interactions enable specific targeting of noncoding RNAs to nascent Pre-mRNAs and chromatin sites. Cell *159*, 188–199.

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

Faghihi, M.A., Modarresi, F., Khalil, A.M., Wood, D.E., Sahagan, B.G., Morgan, T.E., Finch, C.E., St Laurent, G., Kenny, P.J., and Wahlestedt, C. (2008). Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of beta-secretase. Nat. Med. *14*, 723–730.

Ferrandon, D., Koch, I., Westhof, E., and Nüsslein-Volhard, C. (1997). RNA-RNA interaction is required for the formation of specific bicoid mRNA 3' UTR-STAUFEN ribonucleoprotein particles. EMBO J. *16*, 1751–1758.

Filbin, M.E., and Kieft, J.S. (2009). Toward a structural understanding of IRES RNA function. Curr. Opin. Struct. Biol. *19*, 267–276.

Filbin, M.E., Vollmar, B.S., Shi, D., Gonen, T., and Kieft, J.S. (2013). HCV IRES manipulates the ribosome to promote the switch from translation initiation to elongation. Nat. Struct. Mol. Biol. *20*, 150–158.

Fox, P.L. (2015). Discovery and investigation of the GAIT translational control system. RNA N. Y. N *21*, 615–618.

Fraser, C.S. (2015). Quantitative studies of mRNA recruitment to the eukaryotic ribosome. Biochimie *114*, 58–71.

Fraser, C.S., and Doudna, J.A. (2007). Structural and mechanistic insights into hepatitis C viral translation initiation. Nat. Rev. Microbiol. *5*, 29–38.

Garitano-Trojaola, A., Agirre, X., Prósper, F., and Fortes, P. (2013). Long non-coding RNAs in haematological malignancies. Int. J. Mol. Sci. *14*, 15386–15422.

Gebauer, F., and Hentze, M.W. (2016). Genetics. IRES unplugged. Science *351*, 228.

Gebauer, F., Preiss, T., and Hentze, M.W. (2012). From Cis-Regulatory Elements to Complex RNPs and Back. Cold Spring Harb. Perspect. Biol. *4*.

Gingras, A.C., Gygi, S.P., Raught, B., Polakiewicz, R.D., Abraham, R.T., Hoekstra, M.F., Aebersold, R., and Sonenberg, N. (1999). Regulation of 4E-BP1 phosphorylation: a novel two-step mechanism. Genes Dev. *13*, 1422–1437.

Girard, A., Sachidanandam, R., Hannon, G.J., and Carmell, M.A. (2006). A germline-specific class of small RNAs binds mammalian Piwi proteins. Nature *442*, 199–202.

Gogolevsky, K.P., Vassetzky, N.S., and Kramerov, D.A. (2009). 5S rRNA-derived and tRNA-derived SINEs in fruit bats. Genomics *93*, 494–500.

Gong, C., and Maquat, L.E. (2011). "Alu"strious long ncRNAs and their role in shortening mRNA half-lives. Cell Cycle Georget. Tex *10*, 1882–1883.

Gong, C., Tang, Y., and Maquat, L.E. (2013). mRNA–mRNA duplexes that auto-elicit Staufen1-mediated mRNA decay. Nat. Struct. Mol. Biol. *20*, 1214–1220.

Goodenbour, J.M., and Pan, T. (2006). Diversity of tRNA genes in eukaryotes. Nucleic Acids Res. *34*, 6137–6146.

Gradi, A., Svitkin, Y.V., Imataka, H., and Sonenberg, N. (1998). Proteolysis of human eukaryotic translation initiation factor eIF4GII, but not eIF4GI, coincides with the shutoff of host protein synthesis after poliovirus infection. Proc. Natl. Acad. Sci. *95*, 11089–11094.

Guénet, J.L. (2005). The mouse genome. Genome Res. *15*, 1729–1740.

Gupta, R.A., Shah, N., Wang, K.C., Kim, J., Horlings, H.M., Wong, D.J., Tsai, M.-C., Hung, T., Argani, P., Rinn, J.L., et al. (2010). Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature *464*, 1071–1076.

Guttman, M., and Rinn, J.L. (2012). Modular regulatory principles of large non-coding RNAs. Nature *482*, 339–346.

Harvey, R.F., Smith, T.S., Mulroney, T., Queiroz, R.M.L., Pizzinga, M., Dezi, V., Villenueva, E., Ramakrishna, M., Lilley, K.S., and Willis, A.E. (2018). Trans- acting translational regulatory RNA binding proteins. Wiley Interdiscip. Rev. RNA *9*.

Hayashi, G., Hong, C., Hagihara, M., and Nakatani, K. (2012). Activation of prokaryotic translation by antisense oligonucleotides binding to coding region of mRNA. Biochem. Biophys. Res. Commun. *429*, 105–110.

Hayashizaki, Y., and Carninci, P. (2006). Genome Network and FANTOM3: Assessing the Complexity of the Transcriptome. PLoS Genet. *2*.

Hellen, C.U.T. (2009). IRES-induced conformational changes in the ribosome and the mechanism of translation initiation by internal ribosomal entry. Biochim. Biophys. Acta *1789*, 558–570.

Hellen, C.U., and Sarnow, P. (2001). Internal ribosome entry sites in eukaryotic mRNA molecules. Genes Dev. *15*, 1593–1612.

Hershey, J.W.B., Sonenberg, N., and Mathews, M.B. (2012). Principles of translational control: an overview. Cold Spring Harb. Perspect. Biol. *4*.

Hinnebusch, A.G. (2005). Translational regulation of GCN4 and the general amino acid control of yeast. Annu. Rev. Microbiol. *59*, 407–450.

Hinnebusch, A.G. (2011). Molecular mechanism of scanning and start codon selection in eukaryotes. Microbiol. Mol. Biol. Rev. MMBR *75*, 434–467, first page of table of contents.

Hinnebusch, A.G. (2014). The scanning mechanism of eukaryotic translation initiation. Annu. Rev. Biochem. *83*, 779–812.

Hinnebusch, A.G. (2017). Structural Insights into the Mechanism of Scanning and Start Codon Recognition in Eukaryotic Translation Initiation. Trends Biochem. Sci. *42*, 589–611.

Hinnebusch, A.G., and Lorsch, J.R. (2012). The mechanism of eukaryotic translation initiation: new insights and challenges. Cold Spring Harb. Perspect. Biol. *4*.

Holcik, M., and Sonenberg, N. (2005). Translational control in stress and apoptosis. Nat. Rev. Mol. Cell Biol. *6*, 318–327.

Hu, M.C.-Y., Tranque, P., Edelman, G.M., and Mauro, V.P. (1999). rRNA-complementarity in the 5′ untranslated region of mRNA specifying the Gtx homeodomain protein: Evidence that base- pairing to 18S rRNA affects translational efficiency. Proc. Natl. Acad. Sci. *96*, 1339–1344.

IIott, N.E., Heward, J.A., Roux, B., Tsitsiou, E., Fenwick, P.S., Lenzi, L., Goodhead, I., Hertz-Fowler, C., Heger, A., Hall, N., et al. (2014). Long non-coding RNAs and enhancer RNAs regulate the lipopolysaccharide-induced inflammatory response in human monocytes. Nat. Commun. *5*.

Ingolia, N.T., Ghaemmaghami, S., Newman, J.R.S., and Weissman, J.S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science *324*, 218–223.

Ingolia, N.T., Lareau, L.F., and Weissman, J.S. (2011). Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. Cell *147*, 789–802.

Jaafar, Z.A., Oguro, A., Nakamura, Y., and Kieft, J.S. (2016). Translation initiation by the hepatitis C virus IRES requires eIF1A and ribosomal complex remodeling. ELife *5*.

Jackson, R.J., Hellen, C.U.T., and Pestova, T.V. (2010). The mechanism of eukaryotic translation initiation and principles of its regulation. Nat. Rev. Mol. Cell Biol. *11*, 113–127.

Jan, E., and Sarnow, P. (2002). Factorless ribosome assembly on the internal ribosome entry site of cricket paralysis virus. J. Mol. Biol. *324*, 889–902.

Jan, E., Thompson, S.R., Wilson, J.E., Pestova, T.V., Hellen, C.U., and Sarnow, P. (2001). Initiator Met-tRNA-independent translation mediated by an internal ribosome entry site element in cricket paralysis virus-like insect viruses. Cold Spring Harb. Symp. Quant. Biol. *66*, 285–292.

Jeon, Y., and Lee, J.T. (2011). YY1 tethers Xist RNA to the inactive X nucleation center. Cell *146*, 119–133.

Ji, H., Fraser, C.S., Yu, Y., Leary, J., and Doudna, J.A. (2004). Coordinated assembly of human translation initiation complexes by the hepatitis C virus internal ribosome entry site RNA. Proc. Natl. Acad. Sci. U. S. A. *101*, 16990–16995.

Jia, G., Fu, Y., Zhao, X., Dai, Q., Zheng, G., Yang, Y., Yi, C., Lindahl, T., Pan, T., Yang, Y.-G., et al. (2011). N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. Nat. Chem. Biol. *7*, 885–887.

Johnson, R., and Guigó, R. (2014). The RIDL hypothesis: transposable elements as functional domains of long noncoding RNAs. RNA N. Y. N *20*, 959–976.

Jubin, R., Vantuno, N.E., Kieft, J.S., Murray, M.G., Doudna, J.A., Lau, J.Y., and Baroudy, B.M. (2000). Hepatitis C virus internal ribosome entry site (IRES) stem loop IIId contains a phylogenetically conserved GGG triplet essential for translation and IRES folding. J. Virol. *74*, 10430–10437.

Jung, H., Gkogkas, C.G., Sonenberg, N., and Holt, C.E. (2014). Remote control of gene function by local translation. Cell *157*, 26–40.

Kafasla, P., Morgner, N., Robinson, C.V., and Jackson, R.J. (2010). Polypyrimidine tract-binding protein stimulates the poliovirus IRES by modulating eIF4G binding. EMBO J. *29*, 3710–3722.

Kaminker, J.S., Bergman, C.M., Kronmiller, B., Carlson, J., Svirskas, R., Patel, S., Frise, E., Wheeler, D.A., Lewis, S.E., Rubin, G.M., et al. (2002). The transposable elements of the Drosophila melanogaster euchromatin: a genomics perspective. Genome Biol. *3*, research0084.1-84.2.

Kapranov, P., Drenkow, J., Cheng, J., Long, J., Helt, G., Dike, S., and Gingeras, T.R. (2005). Examples of the complex architecture of the human transcriptome revealed by RACE and high-density tiling arrays. Genome Res. *15*, 987–997.

Kapranov, P., Cheng, J., Dike, S., Nix, D.A., Duttagupta, R., Willingham, A.T., Stadler, P.F., Hertel, J., Hackermüller, J., Hofacker, I.L., et al. (2007). RNA maps reveal new RNA classes and a possible function for pervasive transcription. Science *316*, 1484–1488.

Katayama, S., Tomaru, Y., Kasukawa, T., Waki, K., Nakanishi, M., Nakamura, M., Nishida, H., Yap, C.C., Suzuki, M., Kawai, J., et al. (2005). Antisense transcription in the mammalian transcriptome. Science *309*, 1564–1566.

Keene, J.D. (2001). Ribonucleoprotein infrastructure regulating the flow of genetic information between the genome and the proteome. Proc. Natl. Acad. Sci. U. S. A. *98*, 7018–7024.

Keene, J.D. (2007). RNA regulons: coordination of post-transcriptional events. Nat. Rev. Genet. *8*, 533–543.

Keene, J.D., and Lager, P.J. (2005). Post-transcriptional operons and regulons co-ordinating gene expression. Chromosome Res. Int. J. Mol. Supramol. Evol. Asp. Chromosome Biol. *13*, 327–337.

Kelley, D., and Rinn, J. (2012). Transposable elements reveal a stem cell-specific class of long noncoding RNAs. Genome Biol. *13*, R107.

Kerr, C.H., Ma, Z.W., Jang, C.J., Thompson, S.R., and Jan, E. (2016). Molecular analysis of the factorless internal ribosome entry site in Cricket Paralysis virus infection. Sci. Rep. *6*.

Khalil, A.M., Guttman, M., Huarte, M., Garber, M., Raj, A., Rivea Morales, D., Thomas, K., Presser, A., Bernstein, B.E., van Oudenaarden, A., et al. (2009). Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. Proc. Natl. Acad. Sci. U. S. A. *106*, 11667–11672.

Khawaja, A., Vopalensky, V., and Pospisek, M. (2015). Understanding the potential of hepatitis C virus internal ribosome entry site domains to modulate translation initiation via their structure and function. Wiley Interdiscip. Rev. RNA *6*, 211–224.

Kieft, J.S., Zhou, K., Jubin, R., Murray, M.G., Lau, J.Y., and Doudna, J.A. (1999). The hepatitis C virus internal ribosome entry site adopts an ion-dependent tertiary fold. J. Mol. Biol. *292*, 513–529.

Kieft, J.S., Grech, A., Adams, P., and Doudna, J.A. (2001). Mechanisms of internal ribosome entry in translation initiation. Cold Spring Harb. Symp. Quant. Biol. *66*, 277–283.

Kierzek, E., Malgowska, M., Lisowiec, J., Turner, D.H., Gdaniec, Z., and Kierzek, R. (2014). The contribution of pseudouridine to stabilities and structure of RNAs. Nucleic Acids Res. *42*, 3492–3501.

Kim, V.N. (2005). MicroRNA biogenesis: coordinated cropping and dicing. Nat. Rev. Mol. Cell Biol. *6*, 376–385.

King, H.A., Cobbold, L.C., and Willis, A.E. (2010). The role of IRES trans-acting factors in regulating translation initiation. Biochem. Soc. Trans. *38*, 1581–1586.

Komar, A.A., and Hatzoglou, M. (2005). Internal ribosome entry sites in cellular mRNAs: mystery of their existence. J. Biol. Chem. *280*, 23425–23428.

Komar, A.A., and Hatzoglou, M. (2011). Cellular IRES-mediated translation. Cell Cycle *10*, 229–240.

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

Kopp, F., and Mendell, J.T. (2018). Functional Classification and Experimental Dissection of Long Noncoding RNAs. Cell *172*, 393–407.

Kramerov, D.A., and Vassetzky, N.S. (2005). Short retroposons in eukaryotic genomes. Int. Rev. Cytol. *247*, 165–221.

Kramerov, D.A., and Vassetzky, N.S. (2011a). Origin and evolution of SINEs in eukaryotic genomes. Heredity *107*, 487–495.

Kramerov, D.A., and Vassetzky, N.S. (2011b). SINEs. Wiley Interdiscip. Rev. RNA *2*, 772–786.

Krishnamoorthy, T., Pavitt, G.D., Zhang, F., Dever, T.E., and Hinnebusch, A.G. (2001). Tight binding of the phosphorylated alpha subunit of initiation factor 2 (eIF2alpha) to the regulatory subunits of guanine nucleotide exchange factor eIF2B is required for inhibition of translation initiation. Mol. Cell. Biol. *21*, 5018–5030.

Kudla, G., Lipinski, L., Caffin, F., Helwak, A., and Zylicz, M. (2006). High guanine and cytosine content increases mRNA levels in mammalian cells. PLoS Biol. *4*, e180.

Lachance, P.E.D., Miron, M., Raught, B., Sonenberg, N., and Lasko, P. (2002). Phosphorylation of eukaryotic translation initiation factor 4E is critical for growth. Mol. Cell. Biol. *22*, 1656–1663.

Lahr, R.M., Fonseca, B.D., Ciotti, G.E., Al-Ashtal, H.A., Jia, J.-J., Niklaus, M.R., Blagden, S.P., Alain, T., and Berman, A.J. (2017). La-related protein 1 (LARP1) binds the mRNA cap, blocking eIF4F assembly on TOP mRNAs. ELife *6*.

Landry, D.M., Hertz, M.I., and Thompson, S.R. (2009). RPS25 is essential for translation initiation by the Dicistroviridae and hepatitis C viral IRESs. Genes Dev. *23*, 2753–2764.

Lapidot, M., and Pilpel, Y. (2006). Genome-wide natural antisense transcription: coupling its regulation to its different regulatory mechanisms. EMBO Rep. *7*, 1216–1222.

Leontis, N.B., and Westhof, E. (2001). Geometric nomenclature and classification of RNA base pairs. RNA N. Y. N *7*, 499–512.

Leontis, N.B., Stombaugh, J., and Westhof, E. (2002). The non-Watson-Crick base pairs and their associated isostericity matrices. Nucleic Acids Res. *30*, 3497–3531.

Leppek, K., Das, R., and Barna, M. (2018). Functional 5’ UTR mRNA structures in eukaryotic translation regulation and how to find them. Nat. Rev. Mol. Cell Biol. *19*, 158–174.

Levine, M., and Tjian, R. (2003). Transcription regulation and animal diversity. Nature *424*, 147–151.

Li, P.W.-L., Li, J., Timmerman, S.L., Krushel, L.A., and Martin, S.L. (2006). The dicistronic RNA from the mouse LINE-1 retrotransposon contains an internal ribosome entry site upstream of each ORF: implications for retrotransposition. Nucleic Acids Res. *34*, 853–864.

Li, R., Zhu, H., and Luo, Y. (2016). Understanding the Functions of Long Non-Coding RNAs through Their Higher-Order Structures. Int. J. Mol. Sci. *17*.

Li, T., Spearow, J., Rubin, C.M., and Schmid, C.W. (1999). Physiological stresses increase mouse short interspersed element (SINE) RNA expression in vivo. Gene *239*, 367–372.

Liang, D., and Wilusz, J.E. (2014). Short intronic repeat sequences facilitate circular RNA production. Genes Dev. *28*, 2233–2247.

Liang, X., Sun, H., Shen, W., Wang, S., Yao, J., Migawa, M.T., Bui, H.-H., Damle, S.S., Riney, S., Graham, M.J., et al. (2017). Antisense oligonucleotides targeting translation inhibitory elements in 5′ UTRs can selectively increase protein levels. Nucleic Acids Res. *45*, 9528–9546.

Liang, X.-H., Shen, W., Sun, H., Migawa, M.T., Vickers, T.A., and Crooke, S.T. (2016). Translation efficiency of mRNAs is increased by antisense oligonucleotides targeting upstream open reading frames. Nat. Biotechnol. *34*, 875–880.

Lin, J.-Y., Chen, T.-C., Weng, K.-F., Chang, S.-C., Chen, L.-L., and Shih, S.-R. (2009). Viral and host proteins involved in picornavirus life cycle. J. Biomed. Sci. *16*, 103.

Lin, S., Zhang, L., Luo, W., and Zhang, X. (2015). Characteristics of Antisense Transcript Promoters and the Regulation of Their Activity. Int. J. Mol. Sci. *17*.

Lin, S., Choe, J., Du, P., Triboulet, R., and Gregory, R.I. (2016). The m(6)A Methyltransferase METTL3 Promotes Translation in Human Cancer Cells. Mol. Cell *62*, 335–345.

Liu, B., and Qian, S.-B. (2014). Translational reprogramming in cellular stress response. Wiley Interdiscip. Rev. RNA *5*, 301–315.

Liu, J., Yue, Y., Han, D., Wang, X., Fu, Y., Zhang, L., Jia, G., Yu, M., Lu, Z., Deng, X., et al. (2014). A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. Nat. Chem. Biol. *10*, 93–95.

Lu, Z., Zhang, Q.C., Lee, B., Flynn, R.A., Smith, M.A., Robinson, J.T., Davidovich, C., Gooding, A.R., Goodrich, K.J., Mattick, J.S., et al. (2016). RNA Duplex Map in Living Cells Reveals Higher-Order Transcriptome Structure. Cell *165*, 1267–1279.

Lukavsky, P.J. (2009). Structure and function of HCV IRES domains. Virus Res. *139*, 166–171.

Macdonald, P.M., Kanke, M., and Kenny, A. Community effects in regulation of translation. ELife *5*.

Mailliot, J., and Martin, F. (2018). Viral internal ribosomal entry sites: four classes for one goal. Wiley Interdiscip. Rev. RNA *9*.

Makałowski, W. (2000). Genomic scrap yard: how genomes utilize all that junk. Gene *259*, 61–67.

Malygin, A.A., Kossinova, O.A., Shatsky, I.N., and Karpova, G.G. (2013). HCV IRES interacts with the 18S rRNA to activate the 40S ribosome for subsequent steps of translation initiation. Nucleic Acids Res. *41*, 8706–8714.

Mansfield, K.D., and Keene, J.D. (2009). The ribonome: a dominant force in co-ordinating gene expression. Biol. Cell *101*, 169–181.

Manzella, J.M., and Blackshear, P.J. (1990). Regulation of rat ornithine decarboxylase mRNA translation by its 5'-untranslated region. J. Biol. Chem. *265*, 11817–11822.

Marín-Béjar, O., and Huarte, M. (2015). RNA pulldown protocol for in vitro detection and identification of RNA-associated proteins. Methods Mol. Biol. Clifton NJ *1206*, 87–95.

Masaki, T., Arend, K.C., Li, Y., Yamane, D., McGivern, D.R., Kato, T., Wakita, T., Moorman, N.J., and Lemon, S.M. (2015). miR-122 Stimulates Hepatitis C Virus RNA Synthesis by Altering the Balance of Viral RNAs Engaged in Replication Versus Translation. Cell Host Microbe *17*, 217–228.

Matsuda, D., and Mauro, V.P. (2014). Base pairing between hepatitis C virus RNA and 18S rRNA is required for IRES-dependent translation initiation in vivo. Proc. Natl. Acad. Sci. U. S. A. *111*, 15385–15389.

Mattick, J.S. (2001). Non-coding RNAs: the architects of eukaryotic complexity. EMBO Rep. *2*, 986–991.

Matveev, V., and Okada, N. (2009). Retroposons of salmonoid fishes (Actinopterygii: Salmonoidei) and their evolution. Gene *434*, 16–28.

Matylla-Kulinska, K., Tafer, H., Weiss, A., and Schroeder, R. (2014). Functional repeat-derived RNAs often originate from retrotransposon-propagated ncRNAs. Wiley Interdiscip. Rev. RNA *5*, 591–600.

Mayr, C. (2016). Evolution and Biological Roles of Alternative 3'UTRs. Trends Cell Biol. *26*, 227–237.

Mayr, C. (2017). Regulation by 3'-Untranslated Regions. Annu. Rev. Genet. *51*, 171–194.

McCarthy, E.M., and McDonald, J.F. (2004). Long terminal repeat retrotransposons of Mus musculus. Genome Biol. *5*, R14.

McCullers, T.J., and Steiniger, M. (2017). Transposable elements in Drosophila. Mob. Genet. Elem. *7*, 1–18.

Meng, Z., Jackson, N.L., Shcherbakov, O.D., Choi, H., and Blume, S.W. (2010). The human IGF1R IRES likely operates through a Shine-Dalgarno-like interaction with the G961 loop (E-site) of the 18S rRNA and is kinetically modulated by a naturally-polymorphic polyU loop. J. Cell. Biochem. *110*, 531–544.

Merrick, W.C. (2015). eIF4F: a retrospective. J. Biol. Chem. *290*, 24091–24099.

Meyer, K.D., and Jaffrey, S.R. (2014). The dynamic epitranscriptome: N6-methyladenosine and gene expression control. Nat. Rev. Mol. Cell Biol. *15*, 313–326.

Meyer, K.D., Saletore, Y., Zumbo, P., Elemento, O., Mason, C.E., and Jaffrey, S.R. (2012). Comprehensive Analysis of mRNA Methylation Reveals Enrichment in 3' UTRs and Near Stop Codons. Cell *149*, 1635–1646.

Meyer, K.D., Patil, D.P., Zhou, J., Zinoviev, A., Skabkin, M.A., Elemento, O., Pestova, T.V., Qian, S.-B., and Jaffrey, S.R. (2015). 5' UTR m(6)A Promotes Cap-Independent Translation. Cell *163*, 999–1010.

Michelini, F., Pitchiaya, S., Vitelli, V., Sharma, S., Gioia, U., Pessina, F., Cabrini, M., Wang, Y., Capozzo, I., Iannelli, F., et al. (2017). Damage-induced lncRNAs control the DNA damage response through interaction with DDRNAs at individual double-strand breaks. Nat. Cell Biol. *19*, 1400–1411.

Mignone, F., Gissi, C., Liuni, S., and Pesole, G. (2002). Untranslated regions of mRNAs. Genome Biol. *3*, REVIEWS0004.

Miller, D.M., Thomas, S.D., Islam, A., Muench, D., and Sedoris, K. (2012). c-Myc and Cancer Metabolism. Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res. *18*, 5546–5553.

Miller, W.A., Wang, Z., and Treder, K. (2007). The amazing diversity of cap-independent translation elements in the 3′-untranslated regions of plant viral RNAs. Biochem. Soc. Trans. *35*, 1629–1633.

Mitchell, S.F., and Parker, R. (2014). Principles and properties of eukaryotic mRNPs. Mol. Cell *54*, 547–558.

Mokrejs, M., Vopálenský, V., Kolenaty, O., Masek, T., Feketová, Z., Sekyrová, P., Skaloudová, B., Kríz, V., and Pospísek, M. (2006). IRESite: the database of experimentally verified IRES structures (www.iresite.org). Nucleic Acids Res. *34*, D125-130.

Moteki, S., and Price, D. (2002). Functional Coupling of Capping and Transcription of mRNA. Mol. Cell *10*, 599–609.

Mouse Genome Sequencing Consortium, Waterston, R.H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J.F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., et al. (2002). Initial sequencing and comparative analysis of the mouse genome. Nature *420*, 520–562.

Nishihara, H., Smit, A.F.A., and Okada, N. (2006). Functional noncoding sequences derived from SINEs in the mammalian genome. Genome Res. *16*, 864–874.

Novikova, I.V., Hennelly, S.P., Tung, C.-S., and Sanbonmatsu, K.Y. (2013a). Rise of the RNA machines: exploring the structure of long non-coding RNAs. J. Mol. Biol. *425*, 3731–3746.

Novikova, I.V., Hennelly, S.P., and Sanbonmatsu, K.Y. (2013b). Tackling structures of long noncoding RNAs. Int. J. Mol. Sci. *14*, 23672–23684.

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

Otsuka, Y., Kedersha, N.L., and Schoenberg, D.R. (2009). Identification of a cytoplasmic complex that adds a cap onto 5'-monophosphate RNA. Mol. Cell. Biol. *29*, 2155–2167.

Paek, K.Y., Hong, K.Y., Ryu, I., Park, S.M., Keum, S.J., Kwon, O.S., and Jang, S.K. (2015). Translation initiation mediated by RNA looping. Proc. Natl. Acad. Sci. *112*, 1041–1046.

Patrucco, L., Chiesa, A., Soluri, M.F., Fasolo, F., Takahashi, H., Carninci, P., Zucchelli, S., Santoro, C., Gustincich, S., Sblattero, D., et al. (2015). Engineering mammalian cell factories with SINEUP noncoding RNAs to improve translation of secreted proteins. Gene *569*, 287–293.

Patursky-Polischuk, I., Stolovich-Rain, M., Hausner-Hanochi, M., Kasir, J., Cybulski, N., Avruch, J., Rüegg, M.A., Hall, M.N., and Meyuhas, O. (2009). The TSC-mTOR pathway mediates translational activation of TOP mRNAs by insulin largely in a raptor- or rictor-independent manner. Mol. Cell. Biol. *29*, 640–649.

Pelechano, V., and Steinmetz, L.M. (2013). Gene regulation by antisense transcription. Nat. Rev. Genet. *14*, 880–893.

Pellizzoni, L., Cardinali, B., Lin-Marq, N., Mercanti, D., and Pierandrei-Amaldi, P. (1996). AXenopus laevisHomologue of the La Autoantigen Binds the Pyrimidine Tract of the 5′ UTR of Ribosomal Protein mRNAsin Vitro: Implication of a Protein Factor in Complex Formation. J. Mol. Biol. *259*, 904–915.

Pestova, T.V., Lomakin, I.B., and Hellen, C.U.T. (2004). Position of the CrPV IRES on the 40S subunit and factor dependence of IRES/80S ribosome assembly. EMBO Rep. *5*, 906–913.

Petrov, A., Grosely, R., Chen, J., O'Leary, S.E., and Puglisi, J.D. (2016). Multiple Parallel Pathways of Translation Initiation on the CrPV IRES. Mol. Cell *62*, 92–103.

Pfingsten, J.S., Costantino, D.A., and Kieft, J.S. (2007). Conservation and diversity among the three-dimensional folds of the Dicistroviridae intergenic region IRESes. J. Mol. Biol. *370*, 856–869.

Pichon, X., Wilson, L.A., Stoneley, M., Bastide, A., King, H.A., Somers, J., and Willis, A.E. (2012). RNA Binding Protein/RNA Element Interactions and the Control of Translation. Curr. Protein Pept. Sci. *13*, 294–304.

Podbevšek, P., Fasolo, F., Bon, C., Cimatti, L., Reißer, S., Carninci, P., Bussi, G., Zucchelli, S., Plavec, J., and Gustincich, S. (2018). Structural determinants of the SINE B2 element embedded in the long non-coding RNA activator of translation AS Uchl1. Sci. Rep. *8*, 3189.

Ponicsan, S.L., Kugel, J.F., and Goodrich, J.A. (2010). Genomic gems: SINE RNAs regulate mRNA production. Curr. Opin. Genet. Dev. *20*, 149–155.

Proudfoot, N.J. (2001). Genetic dangers in poly(A) signals. EMBO Rep. *2*, 891–892.

Raghavan, A., Ogilvie, R.L., Reilly, C., Abelson, M.L., Raghavan, S., Vasdewani, J., Krathwohl, M., and Bohjanen, P.R. (2002). Genome-wide analysis of mRNA decay in resting and activated primary human T lymphocytes. Nucleic Acids Res. *30*, 5529–5538.

Ramanathan, A., Robb, G.B., and Chan, S.-H. (2016). mRNA capping: biological functions and applications. Nucleic Acids Res. *44*, 7511–7526.

Reveal, B., Yan, N., Snee, M.J., Pai, C.-I., Gim, Y., and Macdonald, P.M. (2010). BREs mediate both repression and activation of oskar mRNA translation and act in trans. Dev. Cell *18*, 496–502.

Roux, P.P., and Topisirovic, I. (2018). Signaling Pathways Involved in the Regulation of mRNA Translation. Mol. Cell. Biol. *38*.

Rudin, C.M., and Thompson, C.B. (2001). Transcriptional activation of short interspersed elements by DNA-damaging agents. Genes. Chromosomes Cancer *30*, 64–71.

Sajjanar, B., Deb, R., Raina, S.K., Pawar, S., Brahmane, M.P., Nirmale, A.V., Kurade, N.P., Manjunathareddy, G.B., Bal, S.K., and Singh, N.P. (2017). Untranslated regions (UTRs) orchestrate translation reprogramming in cellular stress responses. J. Therm. Biol. *65*, 69–75.

Salmena, L., Poliseno, L., Tay, Y., Kats, L., and Pandolfi, P.P. (2011). A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? Cell *146*, 353–358.

Sánchez, Y., and Huarte, M. (2013). Long non-coding RNAs: challenges for diagnosis and therapies. Nucleic Acid Ther. *23*, 15–20.

Sawicka, K., Bushell, M., Spriggs, K.A., and Willis, A.E. (2008). Polypyrimidine-tract-binding protein: a multifunctional RNA-binding protein. Biochem. Soc. Trans. *36*, 641–647.

Schein, A., Zucchelli, S., Kauppinen, S., Gustincich, S., and Carninci, P. (2016). Identification of antisense long noncoding RNAs that function as SINEUPs in human cells. Sci. Rep. *6*, 33605.

Schulz, F., Yutin, N., Ivanova, N.N., Ortega, D.R., Lee, T.K., Vierheilig, J., Daims, H., Horn, M., Wagner, M., Jensen, G.J., et al. (2017). Giant viruses with an expanded complement of translation system components. Science *356*, 82–85.

Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2011). Global quantification of mammalian gene expression control. Nature *473*, 337–342.

Schwartz, S., Bernstein, D.A., Mumbach, M.R., Jovanovic, M., Herbst, R.H., León-Ricardo, B.X., Engreitz, J.M., Guttman, M., Satija, R., Lander, E.S., et al. (2014). Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. Cell *159*, 148–162.

Scott, M.S., and Ono, M. (2011). From snoRNA to miRNA: Dual function regulatory non-coding RNAs. Biochimie *93*, 1987–1992.

Sela, N., Mersch, B., Gal-Mark, N., Lev-Maor, G., Hotz-Wagenblatt, A., and Ast, G. (2007). Comparative analysis of transposed element insertion within human and mouse genomes reveals Alu's unique role in shaping the human transcriptome. Genome Biol. *8*, R127.

Sharma, E., Sterne-Weiler, T., O'Hanlon, D., and Blencowe, B.J. (2016). Global Mapping of Human RNA-RNA Interactions. Mol. Cell *62*, 618–626.

Shiraki, T., Kondo, S., Katayama, S., Waki, K., Kasukawa, T., Kawaji, H., Kodzius, R., Watahiki, A., Nakamura, M., Arakawa, T., et al. (2003). Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. Proc. Natl. Acad. Sci. U. S. A. *100*, 15776–15781.

Simone, L.E., and Keene, J.D. (2013). Mechanisms Coordinating ELAV/Hu mRNA Regulons. Curr. Opin. Genet. Dev. *23*, 35–43.

Sinvani, H., Haimov, O., Svitkin, Y., Sonenberg, N., Tamarkin-Ben-Harush, A., Viollet, B., and Dikstein, R. (2015). Translational tolerance of mitochondrial genes to metabolic energy stress involves TISU and eIF1-eIF4GI cooperation in start codon selection. Cell Metab. *21*, 479–492.

Siomi, M.C., Sato, K., Pezic, D., and Aravin, A.A. (2011). PIWI-interacting small RNAs: the vanguard of genome defence. Nat. Rev. Mol. Cell Biol. *12*, 246–258.

Somers, J., Pöyry, T., and Willis, A.E. (2013). A perspective on mammalian upstream open reading frame function. Int. J. Biochem. Cell Biol. *45*, 1690–1700.

Sonenberg, N., and Hinnebusch, A.G. (2009). Regulation of translation initiation in eukaryotes: mechanisms and biological targets. Cell *136*, 731–745.

de Souza, F.S.J., Franchini, L.F., and Rubinstein, M. (2013). Exaptation of Transposable Elements into Novel Cis-Regulatory Elements: Is the Evidence Always Strong? Mol. Biol. Evol. *30*, 1239–1251.

Spriggs, K.A., Cobbold, L.C., Jopling, C.L., Cooper, R.E., Wilson, L.A., Stoneley, M., Coldwell, M.J., Poncet, D., Shen, Y.-C., Morley, S.J., et al. (2009). Canonical Initiation Factor Requirements of the Myc Family of Internal Ribosome Entry Segments. Mol. Cell. Biol. *29*, 1565–1574.

Spriggs, K.A., Bushell, M., and Willis, A.E. (2010). Translational regulation of gene expression during conditions of cell stress. Mol. Cell *40*, 228–237.

Srivastava, A., Gogoi, P., Deka, B., Goswami, S., and Kanaujia, S.P. (2016). In silico analysis of 5'-UTRs highlights the prevalence of Shine-Dalgarno and leaderless-dependent mechanisms of translation initiation in bacteria and archaea, respectively. J. Theor. Biol. *402*, 54–61.

Stergachis, A.B., Neph, S., Sandstrom, R., Haugen, E., Reynolds, A.P., Zhang, M., Byron, R., Canfield, T., Stelhing-Sun, S., Lee, K., et al. (2014). Conservation of trans-acting networks during mammalian regulatory evolution. Nature *515*, 365–370.

Sun, C., Shepard, D.B., Chong, R.A., López Arriaza, J., Hall, K., Castoe, T.A., Feschotte, C., Pollock, D.D., and Mueller, R.L. (2012). LTR Retrotransposons Contribute to Genomic Gigantism in Plethodontid Salamanders. Genome Biol. Evol. *4*, 168–183.

Terenin, I.M., Andreev, D.E., Dmitriev, S.E., and Shatsky, I.N. (2013). A novel mechanism of eukaryotic translation initiation that is neither m7G-cap-, nor IRES-dependent. Nucleic Acids Res. *41*, 1807–1816.

Thompson, S.R. (2012). So you want to know if your message has an IRES? Wiley Interdiscip. Rev. RNA *3*, 697–705.

Truniger, V., Miras, M., and Aranda, M.A. (2017). Structural and Functional Diversity of Plant Virus 3′-Cap-Independent Translation Enhancers (3′-CITEs). Front. Plant Sci. *8*.

Tsai, M.-C., Manor, O., Wan, Y., Mosammaparast, N., Wang, J.K., Lan, F., Shi, Y., Segal, E., and Chang, H.Y. (2010). Long noncoding RNA as modular scaffold of histone modification complexes. Science *329*, 689–693.

Tschudi, C., and Ullu, E. (2018). Unconventional Rules of Small Nuclear RNA Transcription and Cap Modification in Trypanosomatids. Gene Expr. *10*, 3–16.

Tufarelli, C., Stanley, J.A.S., Garrick, D., Sharpe, J.A., Ayyub, H., Wood, W.G., and Higgs, D.R. (2003). Transcription of antisense RNA leading to gene silencing and methylation as a novel cause of human genetic disease. Nat. Genet. *34*, 157–165.

Uszczynska-Ratajczak, B., Lagarde, J., Frankish, A., Guigó, R., and Johnson, R. (2018). Towards a complete map of the human long non-coding RNA transcriptome. Nat. Rev. Genet. *19*, 535–548.

Van Eden, M.E., Byrd, M.P., Sherrill, K.W., and Lloyd, R.E. (2004). Demonstrating internal ribosome entry sites in eukaryotic mRNAs using stringent RNA test procedures. RNA N. Y. N *10*, 720–730.

Varani, G., and McClain, W.H. (2000). The G x U wobble base pair. A fundamental building block of RNA structure crucial to RNA function in diverse biological systems. EMBO Rep. *1*, 18–23.

Vassetzky, N.S., and Kramerov, D.A. (2013). SINEBase: a database and tool for SINE analysis. Nucleic Acids Res. *41*, D83-89.

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.

Villalba, A., Coll, O., and Gebauer, F. (2011). Cytoplasmic polyadenylation and translational control. Curr. Opin. Genet. Dev. *21*, 452–457.

Volders, P.J., Verheggen, K., Menschaert, G., Vandepoele, K., Martens, L., Vandesompele, J., and Mestdagh, P. (2015). An update on LNCipedia: a database for annotated human lncRNA sequences. Nucleic Acids Res. *43*, 4363–4364.

Walter, M. Transposon regulation upon dynamic loss of DNA methylation. 266.

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

Wang, K.C., and Chang, H.Y. (2011). Molecular mechanisms of long noncoding RNAs. Mol. Cell *43*, 904–914.

Wang, X., Zhao, B.S., Roundtree, I.A., Lu, Z., Han, D., Ma, H., Weng, X., Chen, K., Shi, H., and He, C. (2015). N6-methyladenosine Modulates Messenger RNA Translation Efficiency. Cell *161*, 1388–1399.

Wassarman, D.A., and Steitz, J.A. (1992). Interactions of small nuclear RNA's with precursor messenger RNA during in vitro splicing. Science *257*, 1918–1925.

Weingarten-Gabbay, S., and Segal, E. (2016). Toward a systematic understanding of translational regulatory elements in human and viruses. RNA Biol. *13*, 927–933.

Weingarten-Gabbay, S., Elias-Kirma, S., Nir, R., Gritsenko, A.A., Stern-Ginossar, N., Yakhini, Z., Weinberger, A., and Segal, E. (2016). Comparative genetics. Systematic discovery of cap-independent translation sequences in human and viral genomes. Science *351*.

Werner, A. (2013). Biological functions of natural antisense transcripts. BMC Biol. *11*, 31.

Wicker, T., Sabot, F., Hua-Van, A., L Bennetzen, J., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., et al. (2007). A unified classification system for eukaryotic transposbale elments.

Wight, M., and Werner, A. (2013). The functions of natural antisense transcripts. Essays Biochem. *54*, 91–101.

Willingham, A.T., Orth, A.P., Batalov, S., Peters, E.C., Wen, B.G., Aza-Blanc, P., Hogenesch, J.B., and Schultz, P.G. (2005). A strategy for probing the function of noncoding RNAs finds a repressor of NFAT. Science *309*, 1570–1573.

Wilson, J.E., Pestova, T.V., Hellen, C.U., and Sarnow, P. (2000a). Initiation of protein synthesis from the A site of the ribosome. Cell *102*, 511–520.

Wilson, J.E., Pestova, T.V., Hellen, C.U., and Sarnow, P. (2000b). Initiation of protein synthesis from the A site of the ribosome. Cell *102*, 511–520.

Wutz, A. (2011a). Gene silencing in X-chromosome inactivation: advances in understanding facultative heterochromatin formation. Nat. Rev. Genet. *12*, 542–553.

Wutz, A. (2011b). RNA-mediated silencing mechanisms in mammalian cells. Prog. Mol. Biol. Transl. Sci. *101*, 351–376.

Xie, X., Lu, J., Kulbokas, E.J., Golub, T.R., Mootha, V., Lindblad-Toh, K., Lander, E.S., and Kellis, M. (2005). Systematic discovery of regulatory motifs in human promoters and 3′ UTRs by comparison of several mammals. Nature *434*, 338–345.

Xie, X., Mikkelsen, T.S., Gnirke, A., Lindblad-Toh, K., Kellis, M., and Lander, E.S. (2007). Systematic discovery of regulatory motifs in conserved regions of the human genome, including thousands of CTCF insulator sites. Proc. Natl. Acad. Sci. *104*, 7145–7150.

Xue, S., and Barna, M. (2015). Cis-regulatory RNA elements that regulate specialized ribosome activity. RNA Biol. *12*, 1083–1087.

Yan, P., Luo, S., Lu, J.Y., and Shen, X. (2017). Cis- and trans-acting lncRNAs in pluripotency and reprogramming. Curr. Opin. Genet. Dev. *46*, 170–178.

Yao, Y., Jin, S., Long, H., Yu, Y., Zhang, Z., Cheng, G., Xu, C., Ding, Y., Guan, Q., Li, N., et al. (2015). RNAe: an effective method for targeted protein translation enhancement by artificial non-coding RNA with SINEB2 repeat. Nucleic Acids Res. *43*, e58.

Yue, F., Cheng, Y., Breschi, A., Vierstra, J., Wu, W., Ryba, T., Sandstrom, R., Ma, Z., Davis, C., Pope, B.D., et al. (2014). A Comparative Encyclopedia of DNA Elements in the Mouse Genome. Nature *515*, 355–364.

Yustein, J.T., Liu, Y.-C., Gao, P., Jie, C., Le, A., Vuica-Ross, M., Chng, W.J., Eberhart, C.G., Bergsagel, P.L., and Dang, C.V. (2010). Induction of ectopic Myc target gene JAG2 augments hypoxic growth and tumorigenesis in a human B-cell model. Proc. Natl. Acad. Sci. U. S. A. *107*, 3534–3539.

Zheng, G., Dahl, J.A., Niu, Y., Fedorcsak, P., Huang, C.-M., Li, C.J., Vågbø, C.B., Shi, Y., Wang, W.-L., Song, S.-H., et al. (2013). ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. Mol. Cell *49*, 18–29.

Zhou, J., Wan, J., Gao, X., Zhang, X., Jaffrey, S.R., and Qian, S.-B. (2015). Dynamic m(6)A mRNA methylation directs translational control of heat shock response. Nature *526*, 591–594.

Zhou, J., Wan, J., Shu, X.E., Mao, Y., Liu, X.-M., Yuan, X., Zhang, X., Hess, M.E., Brüning, J.C., and Qian, S.-B. (2018). N6-Methyladenosine Guides mRNA Alternative Translation during Integrated Stress Response. Mol. Cell *69*, 636-647.e7.

Zhou, Y., Zeng, P., Li, Y.-H., Zhang, Z., and Cui, Q. (2016). SRAMP: prediction of mammalian N6-methyladenosine (m6A) sites based on sequence-derived features. Nucleic Acids Res. *44*, e91.

Zucchelli, S., Fasolo, F., Russo, R., Cimatti, L., Patrucco, L., Takahashi, H., Jones, M.H., Santoro, C., Sblattero, D., Cotella, D., et al. (2015a). SINEUPs are modular antisense long non-coding RNAs that increase synthesis of target proteins in cells. Front. Cell. Neurosci. *9*, 174.

Zucchelli, S., Cotella, D., Takahashi, H., Carrieri, C., Cimatti, L., Fasolo, F., Jones, M.H., Sblattero, D., Sanges, R., Santoro, C., et al. (2015b). SINEUPs: A new class of natural and synthetic antisense long non-coding RNAs that activate translation. RNA Biol. *12*, 771–779.

8. ACKNOLEDGEMENTS

First of all, I am grateful to the Almighty God for the protection to this far. I am thankful to my PhD studies advisor, Prof. Stefano Gustincich for granting me the opportunity to join his lab at SISSA and to develop my scientific career and ambitions. A special thank you also goes to Dr. Silvia Zucchelli for the supports during my entire scientific studies. Besides my supervisor, I would like to also thank the Joint Molecular Biology (JuMBO) director, Prof. Giuseppe Legname and all faculty members for their support and shared thoughts on my project during the annual progress report discussions. Also, thank you to my thesis committee for taking their time to go through my written thesis and sharing their thoughts on the presentation.

My sincere thank you also goes to Dr. Marta Codrich for her guidance in the first year of my scientific studies at SISSA, I am grateful for that. Also, a special thank you also goes to Dr. Marta Maurutto for her unflinching support in offering me ride in her car from SISSA to ICGEB and supporting me take measurements. A thank you also goes to Dr. Luca Braga at ICGEB for reserving instruments for taking my experimental measurements.

My deepest thank you also goes to former and current laboratory-mates (Dr. Sara Finaurini, Dr. Maria Bertuzzi, Dr. Aurora Maurizio, Dr. Lavinia Floreani, Dr. Alice Urzì, Dr. Francesca Fasolo, Dr. Carlotta Bon, Chiara Santulli, Gabriele Leoni, Giovanni Spirito, Federico Ansaloni, Bianca Pierattini) for their enormous and constant support throughout my entire studies. Indeed, they made all my days spent in SISSA amazing. I am also grateful to all SISSA lab technicians and administrative stuffs for their unflinching support towards my studies in SISSA; thank you.

Last but not the least, I am grateful to my parents for giving me life, and together with other family members and friends for their constant love, support and encouragements in all things to this far; I am grateful to you all.