SINEUPs, a new class of antisense long non-coding RNAs that enhance synthesis of target proteins in cells: molecular mechanisms and applications.

Thesis submitted for the degree of
"Doctor Philosophiae"

SISSA – International School of Advanced Studies
Area of Neuroscience – *Curriculum* in Functional and Structural Genomics

CANDIDATE
Francesca Fasolo

SUPERVISOR
Prof. Stefano Gustincich

Academic Year 2015/2016

DECLARATION

The work described in this thesis was carried out at SISSA (International School of Advanced Studies) in Trieste, between November 2012 and January 2017 with the exception of:

- phage display selection and biochemical validation by ELISA-based assays, which were performed at the Laboratory of Applied Biology, Università del Piemonte Orientale, Novara, under the supervision of Prof. Claudio Santoro
- chemical footprinting experiments, which were performed at National NMR Laboratory, Ljubljana, under the supervision of Prof. Janez Plavec.

Part of the work described in this thesis is included in the following papers:

SINEUPs are modular antisense long non-coding RNAs that increase synthesis of target proteins in cells.

Zucchelli S*, Fasolo F*, Russo R, Cimatti L, Patrucco L, Takahashi H, Jones MH, Santoro C, Sblattero D, Cotella D, Persichetti F, Carninci P, Gustincich S.

*These authors contributed equally to the work

Front Cell Neurosci. 2015 May 13;9:174. doi: 10.3389/fncel.2015.00174.eCollection 2015.

Structural determinants of the SINEB2 element embedded in the long non-coding RNA activator of translation AS Uchl1

Peter Podbevšek, Francesca Fasolo, Carlotta Bon, Laura Cimatti, Sabine Reißer, Piero Carninci, Giovanni Bussi, Silvia Zucchelli1, Janez Plavec,* and Stefano Gustincich*

Manuscript submitted.

TABLE OF CONTENTS

ABSTRACT

Thanks to continuous advances in sequencing technologies, we know that a huge number of non-coding RNAs are transcribed from mammalian genomes. Of these, long non-coding RNAs (lncRNAs) represent the widest and most heterogeneous class. An increasing number of studies are unveiling lncRNA functions, supporting their active role in regulating gene expression. Regardless of lncRNAs specific functional features, their organization into discrete domains seems to represent a common denominator. Through such domains lncRNAs can recruit and coordinate the activity of multiple effectors, thus working as "flexible modular scaffolds". This model has globally driven towards the quest for regulatory elements within lncRNAs, with a special attention on functional cues deriving from RNA folding. Since transposable elements (TEs) represent 40% of nucleotides of lncRNA sequences, they have been proposed as candidate functional modules.

Carrieri and colleagues recently reported that an embedded inverted SINEB2 element acts as a functional domain in antisense (AS) Uchl1, an AS lncRNA able to increase translation of partially-overlapping protein-coding sense Uchl1 mRNA. AS Uchl1 regulatory properties depend on two RNA domains. A 5' overlapping sequence to the sense transcript is the Binding Domain (BD) and drives specificity of action. An embedded inverted SINEB2 element functions as Effector Domain (ED) conferring translational activation power. AS Uchl1 is the representative member of a new class of lncRNAs, named SINEUPs, as they rely on a SINEB2 element to UP-regulate translation. AS Uchl1 activity can be transferred to a synthetic construct by manipulating the AS sequence in the BD, suggesting the potential use of AS Uchl1-derived synthetic SINEUPs as tools to increase translation of selected targets.

This work was the first example of a specific biological function assigned to an embedded TE leading to the hypothesis that embedded TEs provide functional modules to lncRNAs.

A major limit to the application of SINEUPs is represented by the poor knowledge of the basic mechanisms underlying the biological activity of the ED. A crucial challenge becomes the identification of secondary structures that may confer characteristic protein binding properties. Protein partners would modulate SINEUPs

action and contribute to achieve specific functional outputs.

In this thesis, I focus on understanding the molecular basis of SINEUPs activity in cells and I discuss the potential applications of synthetic SINEUPs as translation enhancers.

First, I investigated the structural basis for translation activation mediated by the ED of SINEUPs. I pointed out that specific structural regions, containing a short terminal hairpin, are involved in the ability of natural and synthetic SINEUPs to increase translation of target mRNAs.

Next, I identified protein partners modulating the activity of SINEUPs in cells. I found that AS Uchl1 interacts with the interleukin enhancer-binding factor 3 (ILF3) and that the presence of the inverted SINEB2 favors binding *in vivo*. In particular, I demonstrated that the AS Uchl1-embedded TEs, inverted SINEB2 and Alu, direct AS Uchl1 localization to ILF3-containing complexes, thus contributing to AS Uchl1 bias towards nuclear localization. I thus suggest that nuclear retention could represent a possible mechanism regulating SINEUP activity.

I also validated the scalability of synthetic SINEUPs as tools to increase protein synthesis of targets of choice. I showed that SINEUP technology can be adapted to a broader number of targets, with interesting potential applications in different fields, from biotechnology to therapy. SINEUPs function in an array of cell lines and can be efficiently directed toward N-terminally tagged proteins. Their biological activity is retained in a miniaturized version within the range of small RNAs length. Their modular structure can be exploited to successfully design synthetic SINEUPs against selected endogenous targets, supporting their efficacy as tools to modulate gene expression *in vitro* and *in vivo*. Hence, I propose SINEUPs as versatile tools to enhance translation of mRNAs of choice.

ABBREVIATIONS

ADAR	adenosine deaminase acting on RNA
AS	antisense
BD	binding domain
bp	base pair
cDNA	complementary DNA
ceRNA	competing endogenous RNA
ciRNA	circular RNA
ds	double-stranded
dsRBD	double-stranded RNA-binding domain
dsRBM	double-stranded RNA-binding motif
ED	effector domain
ENCODE	Encyclopedia of DNA Elements
eRNA	enhancer RNA
FANTOM	Functional Annotation of the Mammalian Genome
FBS	fetal bovine serum
FCS	fetal calf serum
HRP	horseradish peroxidase
IF	immunofluorescence
IP	immunoprecipitation
k	kilo
lincRNA	long intergenic non-coding RNA
LINE	long interspersed elements
lncRNA	long non-coding RNA
LTR	long term repeat
miRNA	micro RNA
mRNA	messenger RNA
MS	mass spectrometry
ncRNA	non-coding RNA
NGS	next generation sequencing
nt	nucleotide
ORF	open reading frame
PD	Parkinson's disease
qRT-PCR	quantitative real time PCR

RBP	RNA-binding protein
RIDome	RNA-interacting domainome
RNA Pol	RNA polymerase
RNP	ribonucleoprotein
RNRE	ribonuclear retention element
rRNA	ribosomal RNA
RT	retrotranscriptase
S	sense
SINE	short interspersed element
siRNA	short-interfering RNA
snoRNA	small nucleolar RNA
ss	single-stranded
TE	transposable elements
tRNA	transfer RNA
Uchl1	Ubiquitin carboxyl-terminal esterase L1
WB	western blot

INTRODUCTION

The central "dogma" of biology states the unidirectional flow of information from gene sequence to protein product (Crick, 1958). According to this vision, the genetic information is encoded in DNA, transcribed to form messenger (m) RNA molecules which are eventually translated into proteins (Jacob and Monod, 1961). Thanks to genomic studies, we know that many exceptions to such "dogma" exist. Namely, the presence of a large amount of RNAs that do not template protein synthesis, but which unexpectedly exert essential regulatory functions in cells. Some of these non-coding RNAs (ncRNAs) had already been spotted in the "pre-genomic era" (Salditt-Georgieff et al., 1981; Weinberg and Penman, 1968) and were later explained by mRNA splicing and RNA genes with infrastructural and regulatory roles (i.e. ribosomal RNA -rRNA-, transfer RNA -tRNA-, RNAse P, SRP-7S). However, for many others, decades were to pass before getting to their isolation and understanding of their function. Pioneer genetic studies in the early nineties unveiled the presence of large regulatory ncRNAs involved in imprinting and other cellular processes (Brown et al., 1992; Huarte and Marín-Béjar, 2015). Almost in parallel, small ncRNAs (microRNAs -miRNAs- and interfering RNAs) were discovered to modulate gene expression by interfering with mRNAs transcription or translation (Hamilton and Baulcombe, 1999; Lagos-Quintana et al., 2001; Lee et al., 1993). The development of full genome sequencing techniques made it possible to survey the transcriptomes of many organisms at an unprecedented degree. In this context, large genomic projects, such as FANTOM (Carninci et al., 2005; FANTOM Consortium and the RIKEN PMI and CLST (DGT) et al., 2014) and ENCODE (ENCODE Project Consortium, 2004; Derrien et al., 2012) marked the beginning of the so called "post-genomic era". These studies provided the scientific world with the astonishing knowledge that the majority (70-80%) of the mammalian genome is transcribed, but that only a tiny part (1-2%) of the transcriptionally active regions correspond to protein-coding genes. Indeed, pervasive transcription produces a vast repertoire of ncRNAs of all sizes and shapes, including small ncRNAs, long non-coding RNAs (lncRNAs) and RNAs of Transposable Elements (TEs). Globally, ncRNAs have been shown to play crucial roles in generating a previously underestimated complexity in

gene regulatory networks. In this context, the past decade has seen a strong interplay between method development, exploration and discovery about RNA biology.

1 LncRNAs

Among ncRNAs, long (or large) non-coding RNAs represent the widest and most heterogeneous class. By definition, lncRNAs are transcripts exceeding 200 nucleotides (nt) in length, with no significant protein coding capacity. According to LNCipedia v3.1, the human genome counts more than 90000 lncRNAs, with their number increasing almost in a daily basis (Iyer et al., 2015; Volders et al., 2015).

The majority of lncRNAs share some basic features of mRNAs: they are transcribed by RNA polymerase II, they undergo splicing, present 5′ caps and are polyadenylated (Guttman et al., 2009). However, what allows the cells to distinguish them from mRNAs, except for the lack of coding potential, is still an open question.

Computational approaches combined with gain/loss-of-function studies have contributed to reveal a close association between lncRNAs and protein-coding transcripts expression patterns, thus providing precious cues on lncRNAs function (Guttman et al., 2009). It is nowadays extensively acknowledged that lncRNAs contribute to gene expression regulation through an array of different mechanisms that depend on their anatomical properties, their subcellular localization and the interactions with other molecular elements. Examples of lncRNAs intervening in transcriptional and translational regulation, cellular trafficking, nuclear organization and compartmentalization have been shown (Huarte and Marín-Béjar, 2015). Moreover, lncRNAs have been reported to be involved in normal organism development as well as in disease (Fatica and Bozzoni, 2013; Chen et al., 2016; Sánchez and Huarte, 2013).

Although the list of lncRNAs involved in arguably important biochemical, cellular, and developmental activities is steadily growing, our knowledge of the mechanisms regulating their biological activity is still limited.

1.1 Classification of lncRNAs

LncRNAs represent the most heterogeneous class of ncRNAs, including an array of transcripts differing in size, anatomical properties, sequence content, subcellular localization and biological function. Accordingly, different criteria can be adopted for their classification.

Anatomical properties

A first classification of lncRNAs can be carried out on the basis of the genomic location relative to their neighbouring protein-coding genes. LncRNAs can thus be intergenic (referred to as long intergenic non-coding RNAs or lincRNAs) if they do not overlap with any other gene or, alternatively, may overlap to genes in exonic, intronic or fully overlapping configuration. In particular, antisense (AS) lncRNAs are transcribed from the opposite DNA strand overlapping with exons of a protein-coding gene.

Subcellular localization

Conversely to mRNAs, which are predominantly localized in the cytoplasm, lncRNAs may be assigned to different subcellular compartments and, accordingly, exert specific functions. Hence, it is possible to distinguish nuclear and cytoplasmic lncRNAs.

A considerable portion of well characterized lncRNAs is nuclear and exert functional roles within the nucleus, in *cis* or *trans*. Nuclear lncRNAs can regulate transcription (Vance and Ponting, 2014), provide structural components of specific subnuclear compartments (Fox and Lamond, 2010) and contribute to shape chromatin organization (Hacisuleyman et al., 2014). Some lncRNAs reside instead in the cytoplasm, where they contribute to post-transcriptional gene expression regulation by "sponging" miRNAs, by sequestering specific proteins or by modulating translation (Carrieri et al., 2012; Cesana et al., 2011; Liu et al., 2015; Tichon et al., 2016). Interestingly, examples of lncRNAs shuttling across different cellular districts in response to specific stimuli have been reported as well (Carrieri et al., 2012; Kino et al., 2010). Although a link between lncRNAs localization and function is supported by different reports (Chen, 2016), the specific mechanisms determining

lncRNAs subcellular fate are still obscure. A major issue towards a comprehensive understanding of lncRNAs regulatory activity concerns the code of rules dictating localization. In this context, a number of studies have focused on high-throughput characterization of lncRNAs localization patterns by combining computational approaches with high-resolution single-molecule imaging techniques (Cabili et al., 2011, 2015; Djebali et al., 2012). Although lncRNAs seem to be predominantly nuclear, recent reports showed that cytoplasmic lncRNAs are actually more abundant than expected and, almost paradoxically, can physically interact with ribosomes (van Heesch et al., 2014; Ingolia et al., 2011). In particular, Carlevaro-Fita and colleagues pointed out some specific characteristics of ribosome-interacting lncRNAs, such as capping, 5'UTR length and depletion of retrotransposons, stressing that intrinsic features of lncRNAs seem to affect their localization (Carlevaro-Fita et al., 2016). Indeed, the presence or absence of specific sequence and structure elements within lncRNAs have been shown to have a say in determining lncRNAs subcellular fate (Chen, 2016). For example, nuclear retention signals can be stored within lncRNAs primary sequence. Often intron-containing or intron-derived transcripts show a bias towards nuclear localization, such as circular intronic RNAs (ciRNAs) (Zhang et al., 2013), which control transcription of parent protein coding genes *in cis*, and lncRNAs with snoRNA ends (sno-lncRNAs), which accumulates at their transcription and processing sites within the nucleoplasm (Yin et al., 2012). Occasionally, some divergently transcribed lncRNAs as Khps1 may form a DNA-RNA triplex with double-stranded DNA thus being trapped closed to their gene loci (Postepska-Igielska et al., 2015). Furthermore, many are the examples of lncRNAs with structural features affecting subcellular localization, which result in restricting their action to specific compartments. Highly-structured domains can be identified in a subset of nuclear lncRNAs which have been or might be implicated in localization, as shown for Kaposi's sarcoma-associated herpesvirus PAN lncRNA's, which contains a U-rich internal loop determining its nuclear retention (Mitton-Fry et al., 2010). Notably, lncRNAs structures can mediate the association with molecular partners, either chromosomes or protein complexes, dictating their subcellular fate. Possible mechanisms determining lncRNAs nuclear retention are summarized in **Figure 1**.

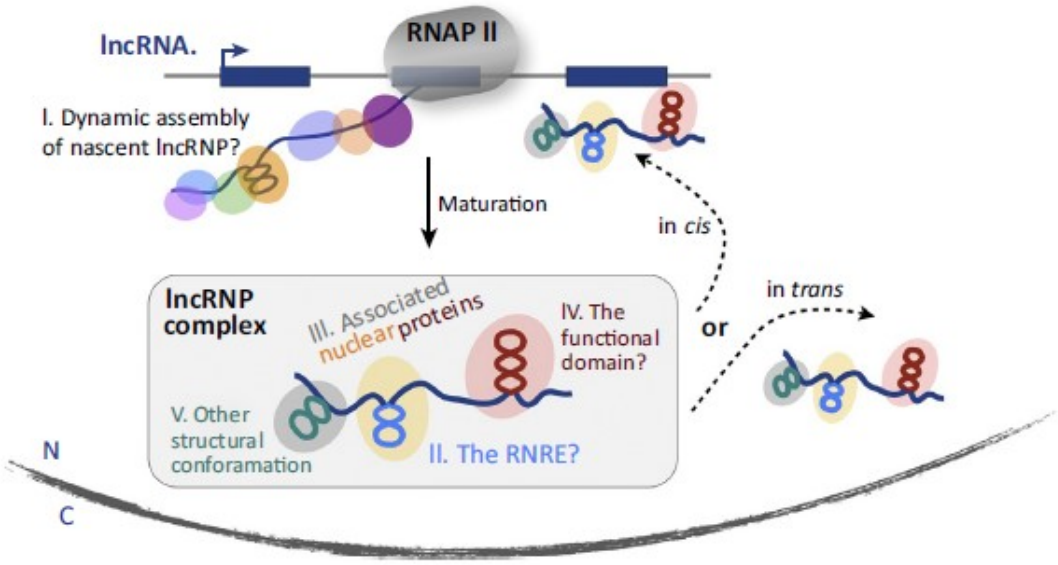


Figure 1. Mechanisms of lncRNAs nuclear retention. Multiple factors can be involved in lncRNAs nuclear retention. Nuclear retention signals can be deposited during the assembly of nascent lncRNPs (I). Upon lncRNP maturation, ribonuclear retention elements (RNRE; II) contained in lncRNAs may associate with nuclear matrix proteins (III) to constrain lncRNAs in the nucleus. Functional domains (IV) of lncRNAs may recruit chromatin modifying complexes or transcription factors or other structural conformation (V) of lncRNAs that may subsequently facilitate their nuclear localization *in cis* or *in trans.* Abbreviation: RNAP=RNA polymerase II. Modified from Chen, 2016.

Mechanism of action

Globally, lncRNAs can exert their regulatory actions by acting in *cis* or *trans.* Interestingly, despite displaying poor conservation within their primary sequences, there are several similarities in lncRNAs modes of action, arguing for a role of higher-order structures in determining function.

Wang and Chang (Wang and Chang, 2011a) recently proposed a classification based on the molecular functions that lncRNAs can execute, distinguishing four main archetypes: signal, decoy, guide and scaffold (**Figure 2**). Of notice, a single lncRNA can match several archetypes, which are not to be intended as mutually exclusive.

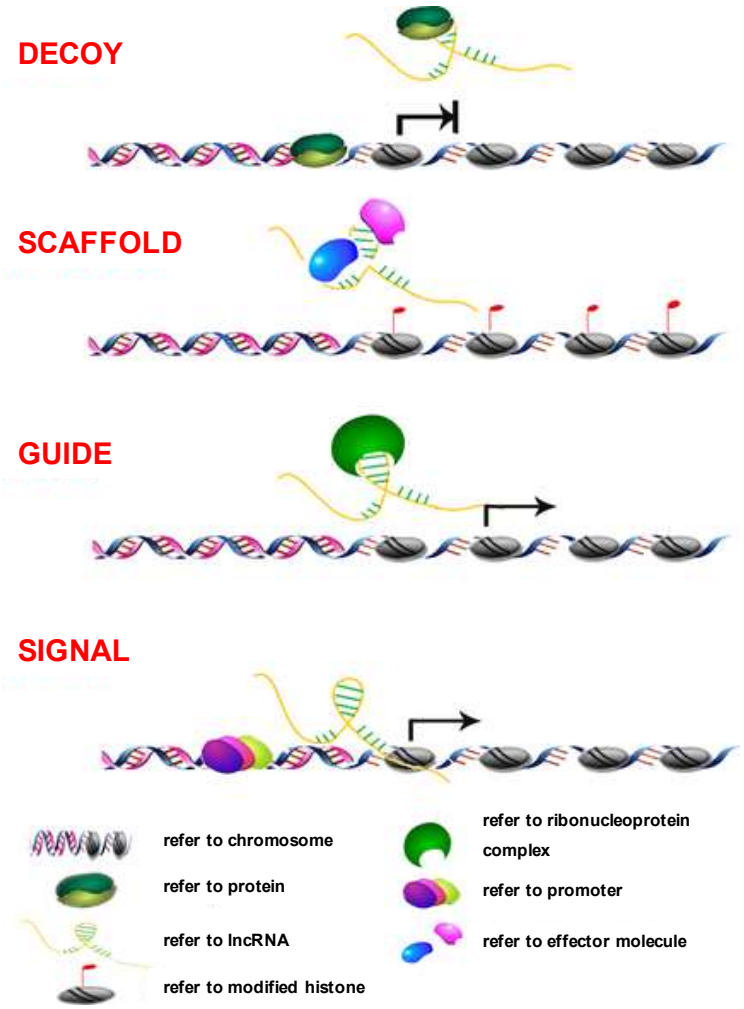


Figure 2. LncRNAs functional archetypes. Schematic diagrams of four lncRNA archetypes. Decoys: acting as a 'molecular sink'; scaffolds: providing a platform to assemble different effector molecules to function together; guides: binding to protein and directing the localization of the ribonucleoprotein complex to specific target genes; signals: function as molecular signals to indicate gene regulation in space and time. Modified from Chen et al., 2014.

The first archetype includes "signal" lncRNAs, which serve as indicators of specific biological processes. Their transcription profiles are indeed peculiar to specific developmental stages, cell types or external stimuli. In this way, such lncRNAs coordinate gene expression in a specific time window, in a given cell population or during adaptive responses to environmental changes. Some of these transcripts exert regulatory functions by establishing physical interactions with genes, proteins or other transcripts; others are instead only by-products of transcription, and it is the transcriptional process itself which becomes regulatory. For example, lncRNAs such as Kcnq1ot1 and Air mediate the transcriptional silencing of multiple genes by

interacting with chromatin and recruiting the chromatin modifying machinery in an allele-specific manner (Mohammad et al., 2009). A different role is carried out by lncRNAs from the mammalian Hox loci: HOTAIR, Frigidair and HOTIP, which act as signals for anatomical position. Their expression was indeed found to be co-linear with the overall anatomic expression pattern of the HOX loci (Rinn et al., 2007).

The use of an RNA intermediate to convey information instead of a protein represents an advantage especially when cellular responses have to be fast, as translation times are bypassed. This is the case of two lncRNAs which are induced by p53 and involved in regulation of p53 transcriptional response to DNA damage: lincp-21 (Huarte et al., 2010) and PANDA (Hung et al., 2011). Finally, a new class of ncRNAs—enhancer or eRNAs—have been described that are produced by activity-dependent RNA polymerase II binding of specific enhancers (Kim et al., 2010). Enhancers have a more active "promoter-like" role in regulating gene expression. The level of eRNA expression at these enhancers positively correlates with the level of messenger RNA synthesis at nearby genes, suggesting an active engaging of these in promoting mRNA synthesis (Kim et al., 2010; Wang and Chang, 2011).

LncRNAs fulfilling the second archetype work as "decoys", acting as molecular sinks for diverse factors. Several examples involve lncRNAs as negatively regulator of an effector. They may target the ability for transcription factors to induce transcription, as in the case of the lncRNA Gas5 on the glucocorticoid receptor (Kino et al., 2010). Moreover, they may act as "sponge" of miRNAs activity (Poliseno et al., 2010), which are thus sequestrated and prevented from exerting their function. "Guide" lncRNAs represent the third archetype: these transcripts bind proteins and subsequently drive their localization to specific targets. With this mechanism, ribonucleoprotein complexes are recruited to DNA, resulting in modulation of gene expression and epigenetic changes. Examples of these RNAs acting either in *cis* or in *trans* have been provided. Irrespective of the adopted mechanism, target gene expression can be both repressed or enhanced. An example of *cis* regulation is the mechanism of mammalian X inactivation. The mammalan X inactivation center (Xic) (Plath et al., 2002) controls the silencing of one of the two X chromosomes in female mammals. This is achieved by the recruitment of Polycomb repressive

complex 2 (PRC2) on the extra X-chromosome, brought in *cis* by RepA ncRNA originating from the 5'end of Xist (Wutz et al., 2002). This results in the creation of a "heterochromatic state" (Sun et al., 2006) that is required for transcriptional induction of Xist, followed by the recruitment of Polycomb and their associated chromatin modifications to the inactive X chromosome (Xi). An example of *cis*-accumulating lncRNA acting in *trans* is represented by FIRRE lncRNA (Hacisuleyman et al., 2014), which is transcribed from the X chromosome and involved in the formation of five trans-chromosomal contacts at its site of transcription. This 3D organization of chromosomes enables long-range interactions between the regulatory genomic elements and the target gene locus.

The fourth and last archetype comprehends the most complex functional class, that is "scaffold" lncRNAs. In this model, lncRNAs provide the core element on which specific molecular components are brought together. Scaffold lncRNAs can simultaneously contact different effectors through different discrete domains, thus dictating and coordinating the dynamics of intermolecular interactions and signalling events (Spitale et al., 2011; Rinn and Guttman, 2012). Globally, this concept relies on the ability of RNA to interact with DNA, RNA and proteins as well. A characteristic example of lncRNAs working as scaffold is provided by the telomerase complex. This is a conserved reverse transcriptase involved in maintenance of genome stability by adding back telomeric DNA repeats lost from chromosome ends (Lustig, 2004). Its catalytic activity requires the association of two subunits: an integral RNA subunit, the telomerase RNA (TERC), that provides the template for repeat synthesis, and a catalytic protein subunit, the telomerase reverse transcriptase (TERT). The TERC, in particular, also possesses structures that contribute to TERT binding and catalytic activity, in addition to those that play major roles in stability of the complex (Collins, 2008). Of notice, Zappulla and Cech demonstrated that protein interaction domains in yeast telomerase RNA can be swapped and spacer regions deleted with almost no impact on RNA function, further supporting a role of flexible scaffold bringing together diverse proteins (Zappulla and Cech, 2004). Further examples of lncRNAs interacting with different effectors (proteins or nucleic acids) through specialized domains are provided by: 1. HOTAIR, containing two different domains for binding of polycomb repressive complex 2 (PRC2) and CoREST-LSD1

complex (Tsai et al., 2010); 2. Xist, harboring two independent functional domains mediating silencing (Rep A for PCR2) and localization (Rep C for hnRNP U and YY1) respectively (Wutz et al., 2002). In addition to domain architecture, lncRNAs can take advantage of the nucleic acid base pairing to specifically select and modulate target RNA and/or DNA molecules.

Interestingly, lncRNAs association with multiple regulatory complexes has been observed in about 30% of embryonic stem cells (ESs) lincRNAs, suggesting a potentially general pattern (Guttman et al., 2011). In this context, recently Guttman and Rinn proposed a model according to which lncRNAs may work as key elements for a potential "modular RNA code" ruling cell states and biological processes (Guttman and Rinn, 2012). A crucial point in deciphering this code is the identification and characterization of key elements within lncRNAs acting as functional domains providing indicators of biological activity (Kapusta and Feschotte, 2014).

A main issue in the investigation of lncRNAs functional domains is represented by their poor sequence conservation. For example, the aforementioned *Xist* exerts a conserved function in mammals; however its first exon, which contains the main characterized Xist functional elements, is one of the less conserved across mammals when sequence is considered. Rather, it has been proposed that tandem repeats in this region form secondary structures necessary for function, both in human and mouse (Duszczyk et al., 2011; Maenner et al., 2010).

Different studies showed that secondary and tertiary structures seem to be much more conserved throughout different lncRNAs than primary sequence, suggesting a strict link between structure and function (Li et al., 2016; Smith et al., 2013). Structures can be neutral to certain mutations and permit interactions with proteins or other nucleic acids. Interestingly, lncRNAs seem to be characterized by a higher degree of structuring compared to protein-coding transcripts (Incarnato et al., 2014), which predominantly owe their biological activities to conserved sequence domains.

2 TEs

Long chromosomal regions which are not occupied by genes, intergenic regions and introns are covered with thousands of repetitive DNA sequences, most of which are mobile genetic elements or TEs. TEs are DNA sequences capable of moving and integrating into new genomic locations. They are diffused in all eukaryotic genomes and constitute between one and two thirds of the entire human genome (de Koning et al., 2011; Lander et al., 2001).

TEs have been often presented as “genomic parasites”, due to the detrimental effects often associated to their mobilization. The organisms' interactions with TEs curiously resemble the host-pathogen co-evolution, with cells being continuously engaged in a genetic “arms race” with endogenous retrotransposons (Dawkins and Krebs, 1979; Goodier, 2016). To this, host genomes have constantly evolved new strategies to fight TEs mobilization and, TEs, on their behalf, have contrived ways to escape such control (Kato et al., 2007). Indeed, some copies of the TEs might retain limited activity bypassing their extinction, as a consequence of the accumulation of neutral mutations. While deleterious mutations are promptly eliminated by selection, neutral ones can occasionally turn out to be beneficial. Accordingly, despite their parasitic nature, TEs can rather be seen as genomes' symbionts (Elbarbary et al., 2016; Goodier, 2016).

Previously considered as genomic “junk”, TEs are now known to play pivotal roles in shaping genomic complexity and dynamicity thanks to their ability to occasionally introduce functional sequence elements at their insertion sites (Cordaux and Batzer, 2009; Elbarbary et al., 2016). Material deriving from TEs has been constantly “recycled” by host genomes for the most diverse purposes. As a result, TEs play a crucial role in regulation of gene expression at different levels. TEs can work as functional DNA elements, like promoters or enhancers (Bejerano et al., 2006; Faulkner et al., 2009; Su et al., 2014) and drive the expression of downstream genes. Often, they provide interfaces for binding with transcription factors or chromatin-remodelling complexes (Bourque et al., 2008; Lynch et al., 2011; Wang et al., 2007). Furthermore, TEs can be “exonized” in transcripts, giving rise to regulatory RNA domains as well. They can promote alternative splicing of pre-mRNAs (Lev-Maor et

al., 2008; Sorek et al., 2002) or, when embedded in 3′ UTRs of mRNAs, affect their stability (An et al., 2004). Phylogenetic analysis showed that different TEs have contributed to the rise of mRNAs polyadenylation sites through the course of evolution (Lee et al., 2008). Furthermore, embedded retrotransposons can provide miRNA-binding sites in target mRNAs (Piriyapongsa et al., 2007). Interestingly, a growing body of evidence suggests that embedded TEs can confer specific functional features to lncRNAs as well (Johnson and Guigó, 2014). Accordingly, several TE families are enriched in lncRNAs, with consistent interspecific variation in the coverage and types of TEs (Kapusta et al., 2013; Kelley and Rinn, 2012; Kim et al., 2016). Over two thirds of mouse and human lncRNA sequences harbor retrotransposon insertions (Kapusta et al., 2013; Volders et al., 2015), that is a considerably higher percentage compared to protein-coding sequences, untranslated regions or small RNAs (Kapusta et al., 2013). More interestingly, the association of retrotransposons with promoters, TSS or polyA sites is by far more frequent in lncRNAs than in protein-coding mRNAs, in both species. Enrichment patterns of different TEs classes seem to be conserved in mouse and human lncRNAs, with ERVs being the most enriched class. An exception is provided by SINEs, which are less diffused in human, but enriched in mouse lncRNAs (Kelley and Rinn, 2012). These observations suggest a strong contribution of TEs in lncRNAs evolution and function (Kapusta et al., 2013). For this, they have been recently proposed to represent the molecular basis of domain organization in lncRNAs (Carrieri et al., 2012; Zucchelli et al., 2015a; Johnson and Guigó, 2014; Kapusta and Feschotte, 2014).

2.1 Classification of TEs

According to the mechanism adopted to mobilize, TEs are divided into two main classes (Huang et al., 2012). In particular, DNA transposons (class II) encode a transposase enzyme which catalyzes their “cut and paste” into new genomic sites; retrotransposons (class I) are instead firstly transcribed into an RNA intermediate, which is subsequently retrotranscribed and integrated at a different site of the genome (“copy and paste”) (Wicker et al., 2007). Retrotransposons constitute the most abundant class of TEs and, conversely to DNA transposons, are limited to

eukaryotes. Three subclasses can be distinguished: long terminal repeat (LTR) elements, Long INterspersed Elements (LINEs) and Short INterspersed Elements (SINEs).

LTR elements are autonomous retrotransposons evolved from retroviruses and, as a consequence, resemble these in terms of genomic structure and mechanism of amplification. They are 5-12 kilobases (kb) long and contain LTRs flanking an ORF encoding for an RNA-dependent DNA polymerase (retrotranscriptase, RT), which catalyzes their amplification (Havecker et al., 2004). Although a variety of LTR retrotransposons exist, only the vertebrate-specific endogenous retroviruses (ERVs) appear to have been active in mammalian genomes (Lander et al., 2001).

Unlike LTRs, LINEs and SINEs lack terminal repeats and are globally referred to as non-LTR retrotransposons.

LINEs are autonomous retrotransposons of 6 kilobases (kb) widespread among eukaryotic genomes. In particular, LINE-1s (L1s), which represent the only active subfamily in mammals, occupy the 19% and 17% of mouse and human genomes, respectively (Ostertag and Kazazian, 2001). Retrotransposition-competent LINE-1s encode the machinery necessary to support their own replication. A typical L1 transcript contains two open reading frames, ORF1 and ORF2. ORF1 encodes a 40 kDa protein (ORF1p) with RNA-binding activity, while ORF2 encodes a 150 kDa protein (ORF2p) with endonuclease and reverse-transcriptase activity. The LINE-1 endonuclease domain is responsible for the dsDNA break at the insertion target site, whereas the reverse transcriptase (RT) activity generates the cDNA copy of LINE-1 to be inserted into the new genomic location (Ostertag and Kazazian, 2001).

SINEs are non-autonomous retrotransposons ranging from 85 to 500 bases in length. Eukaryotic genomes can contain tens or hundreds of thousands of SINE copies (Deragon and Zhang, 2006; Kramerov and Vassetzky, 2005). As SINEs do not encode for their own machinery to retrotranspose, they exploit the endonuclease and RT activity of a partner LINE-encoded protein (Dewannieux et al., 2003).

LINEs and SINEs constitute about the 30% of the human genome and their distribution across it seems to be not casual (Lander et al., 2001). Indeed, while LINEs are generally found in intergenic regions, SINEs are instead abundant in gene-rich regions (Medstrand et al., 2002). This non-random distribution may reflect a

selection against insertion of large LINEs sequences in correspondence of genes, probably to limit occasional detrimental effects. Conversely, SINEs are more likely to be tolerated, probably due to their smaller size, and have interestingly acquired a number of regulatory roles in gene expression (Lander et al., 2001).

2.1.1 SINEs

Like all TEs, SINE elements have been originally presented as genomic parasites (Doolittle and Sapienza, 1980), due to the disruptive effects on gene expression caused by their insertions into coding or regulatory sequences in genomes, often related to disease conditions. However, in some cases, their insertions within coding regions seem to have resulted in the rise of functional elements (Lander et al., 2001).
SINEs are widespread among eukaryotes, even if not as wide as other TEs. They can be found in all mammals, reptile, fishes and some invertebrates, as well as in plants. Of notice, D. melanogaster species lack SINEs. In addition, SINEs are missing in most unicellular eukaryotes (Elbarbary et al., 2016).

Structure and classification of SINEs

Three parts can be typically distinguished into a SINE: a *5′ head*, a *body* and a *3'tail*.
The 5' head, which harbors the promoter, is used to classify SINEs into three superfamilies, according to their derivation from, and thus similarity to, cellular Pol III-transcribed RNAs, that is, tRNAs, 7SL RNA and 5S RNA (Kramerov and Vassetzky, 2011; Wicker et al., 2007). Examples of structures characterizing SINEs belonging to different families are reported in **Figure 3**.
SINEs originating from tRNAs are particularly abundant across different organisms; in particular, the SINEB2s represent one of the main families in the mouse genome, with more than 350 000 copies.
7SL RNA-derived SINEs have been instead identified only in rodents (Krayev et al., 1980; Veniaminova et al., 2007), primates (Deininger et al., 1996; Zietkiewicz and Labuda, 1996) and tree shrews (Nishihara et al., 2002; Vassetzky et al., 2003). In mouse, this family encompasses the SINEB1; in human, it includes Alus.
The number of SINE families originating from 5S rRNA is relatively small; they have been found in some fishes (Kapitonov and Jurka, 2003; Nishihara et al., 2006) and in few mammals: fruit bats (Gogolevsky et al., 2009) and springhare

(Gogolevsky et al., 2008).

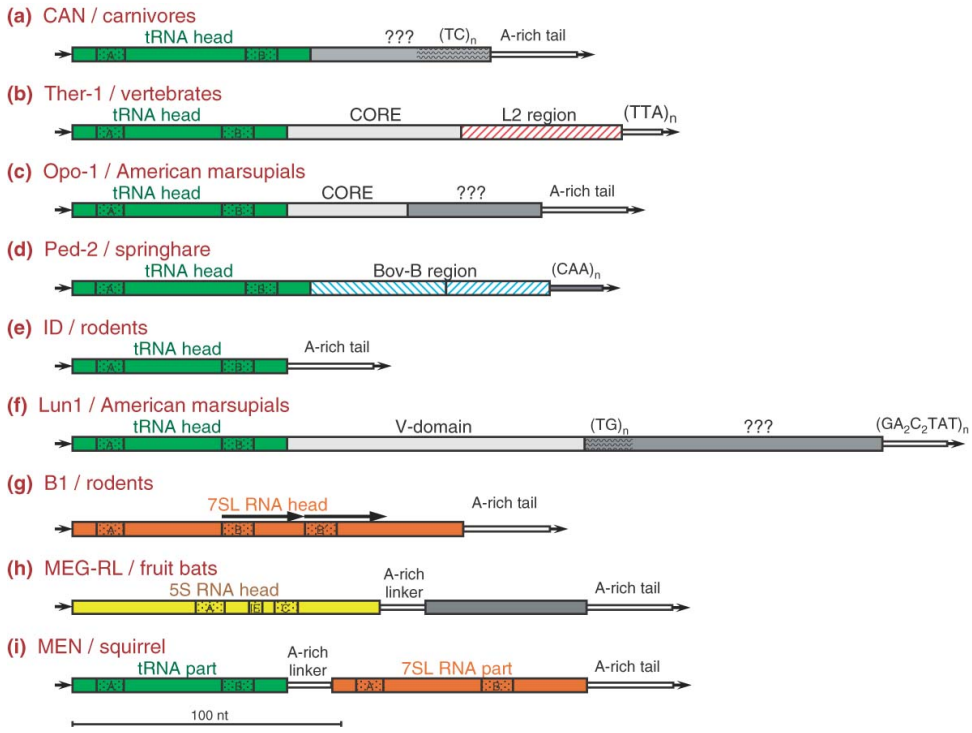


Figure 3. Short interspersed element (SINE) structure examples. (**a**) CAN, tRNA-derived SINE with a unique region of unknown origin and a (TC)n stretch; (**b**) Ther-1, tRNA-derived SINE with a CORE domain and a LINE-derived region; (**c**) Opo-1, tRNA-derived SINE with a CORE domain and a unique region; (**d**) Ped-2, tRNA-derived SINE with a bipartite LINE-derived region; (**e**) ID, simple tRNA-derived SINE; (**f**) Lun-1, tRNA-derived SINE with a V-domain, a (TG)n stretch, and a unique region; (**g**) B1, 7SL RNA-derived SINE with an internal duplication; (**h**) MEG-RL, 5S rRNA-derived SINE with a unique region; (**i**) MEN, dimeric tRNA/7SL RNA-derived SINE. Boxes with dotted background correspond to pol III promoter regions; '???' correspond to body parts of unknown origin; direct repeats including terminal target site duplications are indicated by arrows**.** Modified from Kramerov and Vassetzky, 2011b.

Similarly to the RNAs from which they derive, SINEs have an internal pol III promoter. The promoter in tRNA and 7SL RNA genes consists of two boxes (A and B) of about 11 nt spaced by 30–35 nt, while the 5S rRNA genes have three such boxes: A, IE and C (Schramm and Hernandez, 2002). The presence of the promoter within the transcribed sequence is critical for SINE amplification, as the promoter is preserved in new SINE copies. If SINE retrotransposition relies on Pol III, numerous exonic SINE copies are transcribed by Pol II within pre-mRNAs or other transcripts (i.e. non-coding), making nuclear RNA rich in SINEs (Kramerov and Vassetzky,

2005).

The body of SINE consists of a mostly family-specific central sequence of unknown origin; however, part of this can contain domains shared by different SINE families or, as it happens more frequently and across different species, sharing sequence similarity with the 3'-terminal of a LINE, whose RT is exploited in SINEs amplification (Matveev and Okada, 2009; Ohshima and Okada, 2005). The LINE-derived regions of SINEs can be required or not for the recognition by LINE RTs; accordingly, SINEs are divided into the "stringent" and "relaxed" recognition groups.

SINEs 3' tail is characterized by the presence of repeated mono-, di-, tri-, tetra- or pentanucleotides. The tail of many SINEs is a poly(A) or irregular A-rich sequence (A-tail); the latter can contain the signals of transcription termination and polyadenylation (Borodulina and Kramerov, 2001, 2008).

SINEs in mammalian genomes

Almost all human SINEs belong to a single family known as Alu. They took their origin form the 7SL RNA gene following head-to-tail fusion and have subsequently amplified. Alu elements are primate-specific repeats of about 300 bp which count over 1 million copies in the human genome (constituting the 11% of the entire genome) (Lander et al., 2001). It has been estimated a frequency of more than one Alu for every 3000 bp of genomic DNA; accordingly, their presence in a large number of genes and transcripts has been reported.

Alus inherited the highly conserved A and B boxes from the 7SL RNA promoter gene. However, these elements are not sufficient to drive transcription *in vivo*, so Alus depend on flanking sequences for their expression (**Figure 4**). RNA Pol III-dependent transcripts of this class are referred to as free Alus.

Alu elements are found in gene rich regions, generally within non-coding segments of transcripts, such as in introns and untranslated regions, where they are transcribed by RNA polymerase II (Versteeg et al., 2003). In particular, "exonized" or "embedded" Alus can provide regulatory *cis* elements (Nekrutenko and Li, 2001). They have been associated to regulation of alternative splicing (Lev-Maor et al., 2008; Sorek et al., 2002) and mRNAs turnover. Indeed, the poly(T) sequence that exists in AS Alu elements is the source of ~40% of identified 3′UTR AU-rich

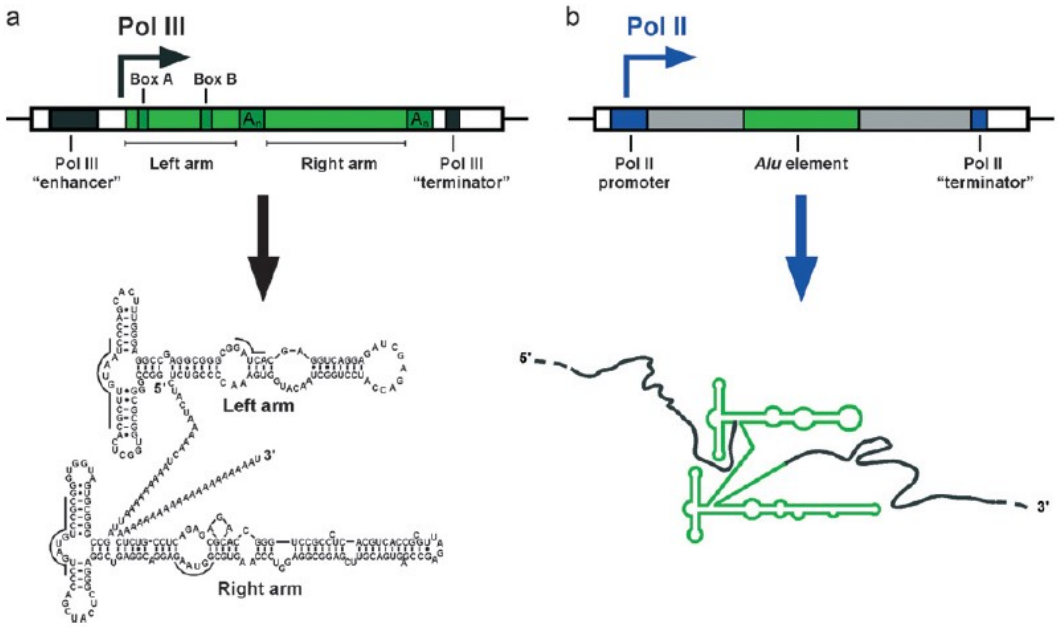


Figure 4. Transcription of Alu elements by Pol II and Pol III. (**a**): Free Alu RNA. This hypothetical Alu element, shown in green, is transcribed by its internal Pol III promoter elements (Box A and Box B), which are helped by an upstream Pol III enhancer (or upstream promoter elements) for efficient transcription. A Pol III terminator (TTTT-3') is depicted downstream of the Alu element. Both Pol III enhancer and terminator are provided by the locus of integration of the Alu element. The secondary structure of the resulting Alu RNA was drawn based on the secondary structure determined by Sinnett et al., 1991 and adapted to the sequence of the Alu element of intron 4 of the *a*- Fetoprotein gene, which was shown to bind the SRP9/14 proteins. Underlined letters indicate the binding sites of SRP9/14 by analogy to SRP RNA. (**b**) Embedded Alu RNA. This hypothetical Alu element, shown in green, is located inside of a protein-coding gene. It might be inserted into an intron, a UTR, and exceptionally into a coding region. This Alu element is then transcribed by Pol II, embedded inside a larger transcription unit. Depending on the sequence environment, as well as on its own sequence, the embedded Alu RNA might adopt a typical Alu fold. Modified from Hasler et al., 2007.

elements (AREs), which regulate mRNA half-life through the competitive binding of proteins that stabilize or destabilize the transcript (An et al., 2004).

As a consequence of the high homology among members of Alu subfamilies, Alu elements can trigger intermolecular pairing and double-stranded (ds) RNA formation (Neeman et al., 2006). In particular, human Alu elements embedded in a group of lncRNAs, called 1/2sbs-RNAs, provide an RNA recognition motif that base pairs with complementary Alu sequences in the 3' UTRs of protein-coding mRNAs to

drive recruitment of STAUFEN1 (STAU1) and STAU1-mediated decay (Gong and Maquat, 2011). Interestingly, Alus have been shown to undergo post-translational modifications, such as editing. DsRNA formation triggers indeed the activity of adenosine deaminase (ADAR) proteins, which converts adenosine residues to inosines (A-to-I) (Savva et al., 2012). Athanasiadis and colleagues determined an interesting correlation between RNA editing and oppositely oriented Alu elements in close proximity. They demonstrated that each edited Alu has a reverse-oriented partner nearby, which appears to be edited as well. The extent of editing appears to depend on the distance between two inverted Alu repeats. Indeed, inverted repeated Alu elements (IRAlus) tend to form 300 bp long dsRNAs as a result of base-paring, thus triggering ADAR activity (Athanasiadis et al., 2004; DeCerbo and Carmichael, 2005; Blow et al., 2004). It has been shown that this process extensively involves Pol II (but not Pol III)-transcribed Alus, which constitute the 90% of human A-to-I editing targets. Edited transcripts are enriched in subnuclear bodies known as paraspeckles. Paraspeckle protein $p54^{nrb}$ and its partners likely participate in the regulation of the nuclear/cytoplasmic distribution of IRAlus-RNAs (Chen and Carmichael, 2008). However, the precise role of A-to-I editing is however still mysterious.

In rodents, SINE sequences occupy the 7.6% of the genome, and are found, on average, every few thousands nucleotides, providing potential templates for transcripts' heterogeneity and ultimately molecular evolution. Two major families of SINEs can be identified in mouse: B1 and B2 (Krayev et al., 1982). SINEB1 originated from 7SL RNAs. Elements of the B1 subclass named ID are involved in dendritic RNA localization when embedded in introns (Buckley et al., 2011). SINE B2s presumably derive from tRNA and have more than 350000 copies in the mouse genome (Kramerov and Vassetzky, 2011a).

Similarly to human Alus, murine SINEs have been shown to provide functional elements involved in the regulation of diverse biological processes. Interestingly, Lunyak and colleagues showed that SINEB2 may serve as chromatin boundary elements. In particular, their bidirectional transcription causes a developmentally relevant change in chromatin structure, establishing a permissive environment that allows transcription of mouse growth hormone (GH) gene (Lunyak et al., 2007).

An example of conserved function of SINEs in both species is provided during stress response. Even if the majority of SINEs are silent in somatic tissues, stressful insults, like heat shock, can induce their transcription via their Pol II promoters, resulting in massive up-regulation of Alu and mouse B2 RNAs, which can in turn inhibit expression of most genes, excluding those up-regulated during heat shock, by binding to Pol II (Allen et al., 2004; Mariner et al., 2008).

Collectively, human and mouse SINEs have been described to modulate protein translation in at least two different manners. They can act as *trans* regulatory factors when transcribed by Pol III and assembled in Alu RNP, and act as *cis* regulatory elements when transcribed by Pol II in 5′- and 3′-UTRs (Häsler and Strub, 2006; Lunyak et al., 2007).

3 SINEUPs: a new functional class of AS lncRNAs

One of the main features of genomes is that different genes residing in opposite DNA strands can co-exist within the same genomic region. As a result of bidirectional transcription, overlapping natural sense/antisense (S/AS) pairs are generated. (Derrien et al., 2012; Katayama et al., 2005; Werner, 2005). In this context, AS transcription provides a florid source of lncRNAs in all kingdoms, from bacteria to humans (Lasa et al., 2012; van Dijk et al., 2011; Katayama et al., 2005).

S/AS pairs may present all the possible combinations between protein-coding and lncRNAs genes and are classified according to their reciprocal genomic organization. In particular, 5' head-to-head divergent, 3' tail-to-tail convergent and fully overlapping configurations can be observed (**Figure 5**).

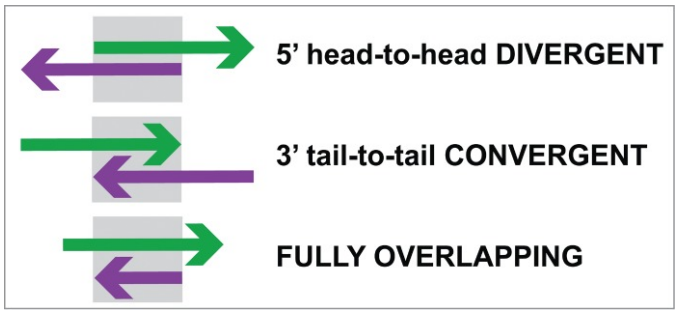


Figure 5. Classification of sense/antisense (S/AS) pairs. Sense genes are in green, AS genes in purple. Arrows indicate 5′ to 3′ direction. Gray box indicates regions of overlap. Adapted from Zucchelli et al., 2015a.

Of notice, approximately 61-72% of all transcribed regions in mouse and human present lncRNAs that are in antisense orientation to adjacent protein-coding genes (Katayama et al., 2005; Cheng et al., 2005). Such pairs in which a lncRNA is transcribed from the opposite strand of a protein-encoding gene are the most studied ones and, in a number of cases, these AS lncRNAs have been demonstrated to modulate cognate protein-coding gene expression in different modes (Pelechano and Steinmetz, 2013). AS lncRNAs may affect the epigenetic state of chromatin (Yu et al., 2008), exert transcriptional control, regulate splicing (Tripathi et al., 2010) and mRNAs stability (Spigoni et al., 2010). Recently, a new class of AS lncRNAs regulating translation of partially overlapping protein-coding transcripts has been

described (Carrieri et al., 2012).

AS Uchl1 and the discovery of natural SINEUPs

In search for AS transcripts likely to regulate the expression of Parkinson's disease (PD)-associated genes, Carrieri and colleagues identified a spliced lncRNA transcript in the murine Ubiquitin carboxyl-terminal hydrolase L1 (Uchl1)/PARK5 gene, mapping in antisense orientation to its protein-coding counterpart.

This 5' head-to-head divergent lncRNA, that initiates within the second intron of Uchl1 and overlaps the first 73 nt of the sense (S) mRNA including the AUG codon (-40/+33 from ATG) was named antisense Uchl1 (AS Uchl1). The non-overlapping part of the transcript contains two embedded repetitive sequences, SINEB1 of the F1 subclass (Alu) and SINEB2 of the B3 subclass (Carrieri et al., 2012) (**Figure 6**).

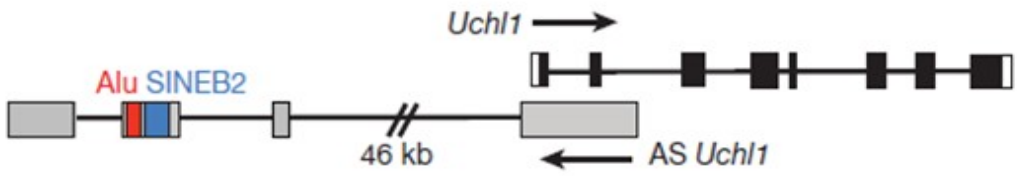


Figure 6. Uchl1/AS Uchl1 genomic organization. Uchl1 exons are in black: 3' and 5' UTRs are in white; AS Uchl1 exons are in gray; repetitive elements are in red (Alu) and in blue (SINEB2); introns are indicated as lines. Adapted from Carrieri et al., 2012.

In mouse, AS Uchl1 is expressed 40% of tissues that expressed Uchl1 mRNA, but no AS Uchl1 is found in the absence of sense transcript (Carrieri et al., 2015). S and AS Uchl1 are highly expressed in the ventral midbrain, with a specific enrichment observed in dopaminergic cells. The two transcripts are localized in different subcellular compartments, with mature Uchl1 mRNA being predominantly detected in the cytoplasm, while AS Uchl1 in the nucleus. Expression of both S and AS transcripts has also been reported in mouse neuronal dopaminergic cell line (MN9D) (Carrieri et al., 2012).

Overexpression of AS Uchl1 in MN9D cells produces no changes in Uchl1 mRNA levels, but it is curiously accompanied by an increase of UCHL1 endogenous protein, suggesting that AS Uchl1 regulation on Uchl1 expression occurs at post-transcriptional level (Carrieri et al., 2012). Selective deletion of AS Uchl1 sequence elements lead to the identification of two functional domains responsible for AS

Uchl1-mediated translation upregulation, that is, the 5' overlapping region and the inverted SINEB2 element.

In physiological conditions, AS Uchl1 accumulates in the nucleus of dopaminergic neurons, but it translocates to the cytoplasm upon cellular stress as induced by rapamycin, an inhibitor of CAP-dependent translation (**Figure 7**). Once in the cytoplasm, AS Uchl1 promotes translation of sense protein-coding mRNA by enhancing its association to heavy polysomes (Carrieri et al., 2012). Importantly, rapamycin-triggered increase of UCHL1 synthesis relies on the presence of a functional AS Uchl1.

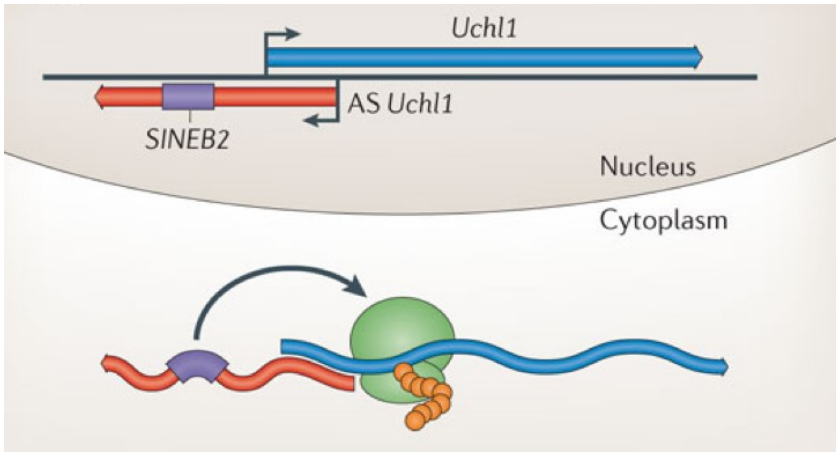


Figure 7. AS Uchl1 up-regulates translation of Uchl1. The 5′ region of an AS transcript to the ubiquitin carboxyl-terminal hydrolase L1 gene (AS Uchl1) recognizes its sense transcript and increases the translation efficiency of Uchl1 — an effect that depends on its SINEB2 domain. Modified from Pelechano and Steinmetz, 2013.

In summary, AS Uchl1 lncRNA enhances UCHL1 translation in stress conditions with a post-transcriptional mechanism that requires functional overlapping and inverted SINEB2 sequences.

Interrogation of FANTOM3 dataset allowed the identification of 31 natural S/AS pairs sharing similar features to S/AS Uchl1, that is, 5' head-to-head overlapping and with the AS transcript harboring an inverted SINEB2 element. Of these, AS Uxt proved to act similarly to AS Uchl1: its overexpression in MN9D cells triggered an increase of endogenous Uxt protein synthesis, with unaffected Uxt mRNA levels (Carrieri et al., 2012). More recently, functional validation has been successfully carried out for other AS RNAs in the list including AS to elastin, a secreted protein.

(Patrucco et al., 2015b). In this context, AS Uchl1 can therefore be considered as the representative member of a new functional class of natural AS lncRNAs capable of up-regulating translation of sense overlapping transcripts. Biological activity depends on the combination of two RNA elements: the overlapping region (Binding Domain, BD) confers target specificity while the embedded inverted SINEB2 element (Effector Domain, ED) is required for translation enhancement (**Figure 8**). These lncRNAs are referred to as SINEUPs, as they rely on a SINEB2 sequence to UP-regulate translation in a gene-specific manner (Zucchelli et al., 2015a; Zucchelli et al., 2015b).

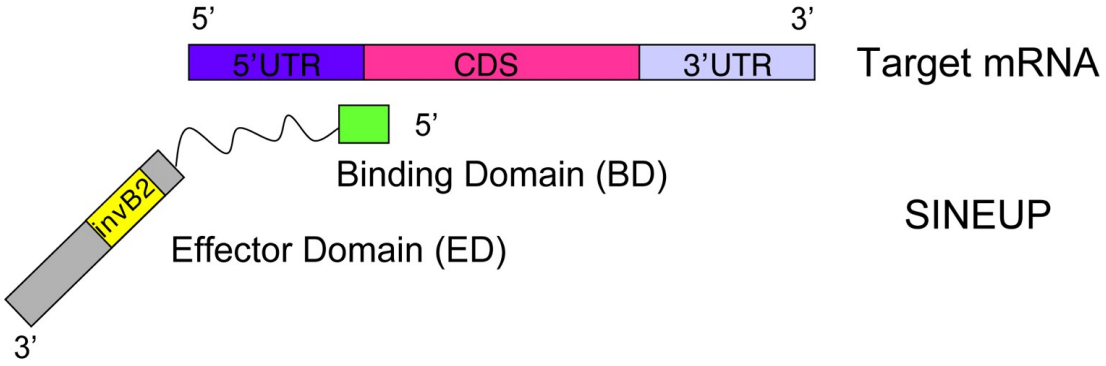


Figure 8. Schematic representation of SINEUPs. SINEUP modular structure. SINEUP binding domain (green): SINEUP sequence that overlaps, in antisense orientation, to the sense protein-coding mRNA. SINEUP effector domain (yellow): non-overlapping portion of SINEUPs (gray), containing the inverted SINEB2 element (invB2) that confers activation of protein synthesis. 5′ to 3′ orientation of sense and antisense RNA molecules is indicated. Structural elements of protein-coding mRNA are shown: 5′ untranslated region (5′UTR, purple), coding sequence (CDS, pink) and 3′ untranslated region (3′UTR, light blue). Adapted from Zucchelli et al., 2015b.

The modular organization of SINEUPs strongly reflects the role of embedded TEs in shaping lncRNAs functional features (Carrieri et al., 2012; Zucchelli et al., 2015a). In particular, TEs could provide binding sites for specific molecular complexes regulating SINEUP activity. At the same time, AS overlapping regions may confer target specificity through RNA/RNA and RNA/DNA pairing.

However, the exact mechanism underlying the activity of the inverted SINEB2 as ED of SINEUPs remains elusive.

Synthetic SINEUPs

The joined activity of the domains found in SINEUPs suggests the potential use of AS Uchl1-derived lncRNAs as enhancers of target mRNAs translation. Taken advantage of SINEUPs domain architecture it is indeed theoretically possible to design artificial SINEUPs to virtually any mRNA of choice by solely manipulating the BD. In particular, Carrieri and colleagues engineered a chimeric construct, named AS GFP, by swapping AS Uchl1 BD with a complementary sequence to EGFP mRNA in antisense orientation. AS GFP derives indeed from natural AS Uchl1 deprived from its 5' end containing the overlapping sequence to Uchl1 (Δ5' AS Uchl1). The whole 3' tail (about 1200 nucleotides), harboring the ED (inverted SINEB2) and the partial Alu repeat, is entirely kept (**Figure 9**).

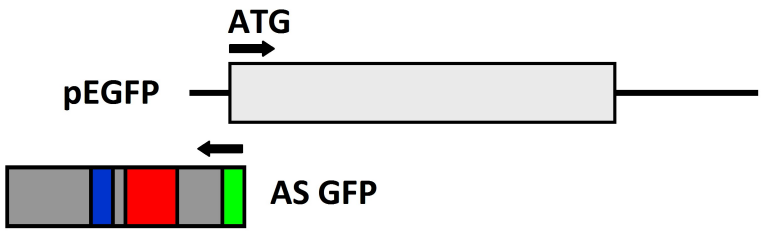


Figure 9. Synthetic AS lncRNA increases target protein levels. Scheme of AS GFP construct. Δ5' AS Uchl1 with repetitive elements (SINEB2, red; Alu, blue) and the overlap (green) regions are indicated.

Coherently with what expected, AS GFP succeeded in up-regulating GFP protein translation when co-transfected with corresponding sense GFP-encoding DNA in HEK 293T cells, with no effects on GFP mRNA levels. Further supporting SINEUPs action at translational level, pulse-labelling experiments showed that induction of GFP was due to an increase of newly synthesized protein. Interestingly, recent studies by Yao and co-workers showed that GFP mRNA is recruited to heavy polysomes in AS GFP (there referred to as RNAe)-transfected cells (Yao et al., 2015).

Collectively, the translational regulatory properties of natural SINEUPs are retained by synthetic ones, paving the way for their applications as tools to selectively modulate gene expression *in vitro* and *in vivo*. Furthermore, common molecular mechanisms are likely to be exploited.

The precise molecular basis of SINEUP activity remains unclear. It is conceivable that the formation of secondary structures could be involved, providing binding sites or signalling cues for recruitment of regulatory elements and molecular effectors. In this context, the study of the structure/function of the ED, as well as the identification of protein partners modulating the activity of SINEUPs would provide insights into the biology of these natural lncRNAs and would give essential information for the application of their synthetic derivatives.

4 RNA-protein interactions

A growing body of literature over recent decades has shown that, in biological environments, RNA is regularly bound and modulated by RNA-binding proteins, (RBPs) (Gerstberger et al., 2014). Association with protein partners is observed throughout the entire life cycle of mRNAs (processing, export, localization, stability, translation and degradation), with significant implications in transcriptional and post-transcriptional gene expression regulation (Glisovic et al., 2008).

RNAs often function together with proteins in ribonucleoproteins (RNPs) providing dynamic structures controlling multiple cellular processes (Mitchell and Parker, 2014; Moore, 2005). RNPs can be promptly remodelled according to the cell's needs, especially under conditions that require adaptive changes (Beckmann et al., 2016).

Even if the largest portion of published research focuses on mRNA-binding proteins (mRBPs) and messenger RNPs, regulation by RBPs is not limited to mRNAs, but it also includes processes acting on ncRNAs. Indeed, it is reasonable to assume that all ncRNAs rely on protein-binding to exert their functions in cells (Eddy, 2001), and this is true for both catalytic (rRNA, small nuclear RNAs -snRNAs- and small nucleolar RNAs -snoRNAs-) (Fica et al., 2013; Leung et al., 2011) and other regulatory RNA species (miRNAs and piwi-associated RNAs -piRNAs-) (Bartel, 2004). In this context, the higher level of complexity characterizing lncRNAs regulatory pathways is known to be strictly linked to physical interactions with diverse proteins modulating their activity (Guttman and Rinn, 2012; Rinn and Chang, 2012; Ulitsky and Bartel, 2013; Wang et al., 2011). Furthermore, functional features of lncRNAs are strictly related to the presence of complex, conserved secondary and tertiary structures (Li et al., 2016). Proteins, on their behalf, are prone to bind RNAs in correspondence of structural domains, like stem–loops and bulges (Nagai, 1996). Recruitment of molecular effectors thus results in the formation of ribonucleoprotein complexes associated with plenty of regulatory outputs (Guttman and Rinn, 2012). Moreover, a number of studies showed that proteins can modulate basic properties of lncRNAs, like stability or localization within cells, ultimately affecting their biological activity. Protein association could result in both a "positive" and a "negative" regulation of lncRNAs function. For example, hnRNP U directs FIRRE-

mediated *trans*-chromosomal association, by positioning the lncRNA at the site of action (Hacisuleyman et al., 2014). Conversely, binding with RBP HuR decreases lincRNA-p21 stability, thus preventing this from repressing JUNB and CTNNB1 mRNAs translation (Yoon et al., 2012). Furthermore, proteins can work as carriers regulating lncRNAs trafficking across cell compartments, as recently shown for HuR and GRSF1, which control the nuclear export and mitochondrial localization of the lncRNA RMRP (Noh et al., 2016). Although the investigation of ncRNA protein partners represents an area of active research (McHugh et al., 2014), the full spectrum of ncRNAs-binding proteins is still unknown.

Collectively, deciphering RNA/protein interactions represents a prerequisite for the dissection of RNA regulatory processes and, more widely, for a better understanding of the physiology of cells.

4.1 RBPs

RBPs are crucial players in gene expression regulation due to their critic role in controlling mRNA metabolism and ncRNAs function. Supporting the importance of RNA-protein interactions in cellular homeostasis is the link between RBPs dysfunction and disease (Castello et al., 2013; Darnell, 2010a).

RBPs functionality relies on their ability to selectively recognize and bind specific motives within target RNAs through an array of diverse RNA-binding domains (RBDs). RBDs are deeply conserved in bacteria, archaea and eukaryotes (Gerstberger et al., 2014), suggesting their contribution to essential biological processes across all kingdoms. Even if different RBDs adopt different strategies to bind RNA, some general features exist characterizing RBPs–RNA interactions.

RNA recognition is driven by the overall protein fold (involving hydrogen bonds with backbone atoms) as well as by specific amino acid side chain–nucleotide interactions. In particular, two modes of binding can be distinguished, that is, groove or β-sheet binding. In the former case, a structured domain of the protein is positioned into the groove of an RNA helix, while in the latter one, binding pockets are formed on β-sheets surfaces, interacting with unpaired RNA bases (Jones et al., 2001). Target specificity is often accomplished by way of hydrogen bonding and electrostatic interactions; binding affinity relies on electrostatic and stacking

interactions (Aviv et al., 2006; Jones et al., 2001).

Individual RBDs usually contact a few nucleotides, and combinations of RBDs within the same protein are frequently observed, likely to increase affinity and specificity (Cook et al., 2015). Collectively, recognition of sequence as well as structural elements within the target RNAs may occur, according to the specific characteristics of different RBDs. Among the most well-characterized RBDs we list RNA recognition motives (RRM), K-homology (KH) domains (type I and type II), RGG (Arg-Gly-Gly) boxes, zinc fingers (ZnF, mostly C-x8-X-x5-X-x3-H), double stranded RNA-binding domains (dsRBD) and Pumilio/FBF (PUF or Pum-HD) domains (Cook et al., 2015). Features of the main canonical RBDs are summarized on **Table 1**. Of notice, not all RBPs contain canonical RBDs (Aviv et al., 2006; Battle and Doudna, 2001).

Domain	Topology	RNA-recognition surface	Protein–RNA interactions	Representative structures (PDB ID)
RRM	αβ	Surface of β-sheet	Interacts with about four nucleotides of ssRNA through stacking, electrostatics and hydrogen bonding	U1A N-terminal RRM[18] (1URN)
KH (type I and type II)	αβ	Hydrophobic cleft formed by variable loop between β2, β3 and GXXG loop. Type II: same as type I, except variable loop is between α2 and β2	Recognizes about four nucleotides of ssRNA through hydrophobic interactions between non-aromatic residues and the bases; sugar-phosphate backbone contacts from the GXXG loop, and hydrogen bonding to bases	Nova-1 KH3 (type I)[41] (1EC6), NusA (type II)[37] (2ASB)
dsRBD	αβ	Helix α1, N-terminal portion of helix α2, and loop between β1 and β2	Shape-specific recognition of the minor–major–minor groove pattern of dsRNA through contacts to the sugar-phosphate backbone; specific contacts from the N-terminal α-helix to RNA in some proteins	dsRBD3 from Staufen[51] (1EKZ)
ZnF-CCHH	αβ	Primarily residues in α-helices	Protein side chain contacts to bulged bases in loops and through electrostatic interactions between side chains and the RNA backbone	Fingers 4–6 of TFIIIA[56] (1UN6)
ZnF-CCCH	Little regular secondary structure	Aromatic side chains form hydrophobic binding pockets for bases that make direct hydrogen bonds to protein backbone	Stacking interactions between aromatic residues and bases create a kink in RNA that allows for the direct recognition of Watson–Crick edges of the bases by the protein backbone	Fingers 1 and 2 of TIS11d[57] (1RGO)
S1	β	Core formed by two β-strands with contributions from surrounding loops	Stacking interactions between bases and aromatic residues and hydrogen bonding to the bases	Ribonuclease II[121] (2IX1), exosome[99] (2NN6)
PAZ	αβ	Hydrophobic pocket formed by OB-like β-barrel and small αβ motif	Recognizes single-stranded 3′ overhangs of siRNA through stacking interactions and hydrogen bonds	PAZ[73] (1SI3), Argonaute[76] (1U04), Dicer[77] (2FFL)
PIWI	αβ	Highly conserved pocket, including a metal ion that is bound to the exposed C-terminal carboxylate	Recognizes the defining 5′ phosphate group in the siRNA guide strand with a highly conserved binding pocket that includes a metal ion	PIWI[75] (1YTU), Argonaute (1U04)[76]
TRAP	β	Edges of β-sheets between each of the 11 subunits that form the entire protein structure	Recognizes the GAG triplet through stacking interactions and hydrogen bonding to bases; limited contacts to the backbone	TRAP[122] (1C9S)
Pumilio	α	Two repeats combine to form binding pocket for individual bases; helix α2 provides specificity-determining residues	Binding pockets for bases provided by stacking interactions; specificity dictated by hydrogen bonds to the Watson–Crick face of a base by two amino acids in helix α2	Pumilio[84] (1M8Y)
SAM	α	Hydrophobic cavity between three helices surrounded by an electropositive region	Shape-dependent recognition of RNA stem–loop, mainly through interactions with the sugar-phosphate backbone and a single base in the loop	Vts1[123] (2ESE)

dsRBD, double-stranded RNA-binding domain; KH, K-homology; OB-like, oligonucleotide/oligosaccharide binding-like; PDB ID, Protein Data Bank identification; RRM, RNA-recognition motif; siRNA, small interfering RNA; ssRNA, single-stranded RNA; ZnF, zinc finger.

Table 1. Characteristics of RNA-binding domains. Adapted from Lunde et al., 2007.

4.2 Methodologies to study RNA/protein interactions

Recent technological developments have allowed characterization of RNA–protein interactions at an unprecedented scale. The methods currently employed can be broken down into two general categories: ‘protein-centric’ and ‘RNA-centric’. These alternative, but often complementary approaches allow to selectively focus on different aspects of RNA-protein interactions.

“Protein-centric” approaches are usually chosen to identify consensus RNA-binding motives for RBPs or, more widely, to profile target RNAs of a desired known RBP. The most popular *in vitro* “protein-centric” approaches include SELEX (systematic evolution of ligands by exponential enrichment) (Ellington and Szostak, 1990; Tuerk and Gold, 1990) and its more recent combination with NGS SEQRS (Campbell et al., 2012), RNAcompete (Ray et al., 2009) and RNA Bind-n-Seq (Lambert et al., 2014). Collectively, these methods allow the selection of preferred high-affinity RNA-binding sites for specific RBPs, starting from randomized pools of RNAs. *In vivo,* “protein-centric” approaches generally rely on the ability to purify a protein (Hogan et al., 2008), or class of proteins (Ingolia et al., 2009), followed by sequencing of the associated RNAs. The first genome-wide analysis for characterization of RBP-RNA interactions involved immunoprecipitation of RBP–RNA complexes using antibodies against endogenous proteins or epitope tags followed by microarray analysis (RIP-chip) (Tenenbaum et al., 2000) or high-throughput sequencing (RIP-seq). Improvements of such methods based on cross-linking the RBP to the RNA using UV radiation before immunoprecipitation (CLIP) ensures maintaining of *in vivo* contacts (Ule et al., 2003). In particular, coupling of CLIP to high-throughput sequencing (HITS-CLIP or CLIP-seq) enables genome-wide identification of RBPs binding sites (Darnell, 2010b; Hogan et al., 2008; Zhang and Darnell, 2011). Global data lead to the observation that RBP-to-mRNA interactions are, in general, many-to-many, with each RBP interacting with diverse mRNAs, and each mRNA being regulated by several RBPs (Hogan et al., 2008). Moreover, such approaches provided initial insights into proteins interacting and regulating lncRNAs' action (Khalil et al., 2009; Rinn et al., 2007; Zhao et al., 2008).

As these approaches require knowledge of the protein, they are of more limited utility for defining the proteins that associate with a given RNA transcript. The advent of mass spectrometry (MS) resulted in the development of complementary genome-wide approaches to determine selected RNAs interactomes, as well as to discover new RBPs. In this context, "RNA-centric" methods rely on the capture of a given RNA, or class of RNAs, and identify the associated proteins by using MS (Baltz et al., 2012; Castello et al., 2012). Thus, hundreds of RBPs have been identified and validated, with a considerable portion of these lacking previous functional annotation as RBPs or containing non-canonical RBDs. Despite having great potential, MS-based approaches do not lack some major caveats. First, these procedures rely on complex multi-step protocols which need to be specifically optimized for different RNP complexes. Moreover, sensitivity limitations of MS may lead to underrepresentation of those complexes made-up of low abundant transcripts and/or RBPs (Castello et al., 2012) and, at the same time, to overrepresentation of proteins with low complexity sequences (Baltz et al., 2012). Some of these limitations could be overcome by approaches based on the screening of protein libraries, both *in vivo* (Harada et al., 1996; Koh and Wickens, 2014) and *in vitro* (Danner and Belasco, 2001; Laird-Offringa and Belasco, 1996; Siprashvili et al., 2012). Any format of protein library can faithfully represent the complexity of a desired full-length proteome, either as an array of candidate proteins or as products of selected Open Reading Frame (ORF) libraries.

In particular, most *in vitro* approaches are based on selection of RBPs or RBDs exposed to immobilized or labelled target RNAs. Usually, different rounds of selection and amplification of the progressively selected libraries are carried out.

The main drawbacks related to *in vitro* selection of protein libraries may derive from limited accessibility of binding sites, lack of *in vivo*-supported post-translational modifications (and subsequent impossibility to isolate proteins that bind RNA by virtue of these), overrepresentation of certain clones within the starting library (and generation of false positives) and occasional failure in proper protein folding (DiDonato et al., 2004; Roberts and Szostak, 1997). However, some of these issues could be partially or completely overcome: for example, during their constructions, libraries can be "filtered", thus exclusively displaying correctly folded, stable, active

proteins. If this is still a difficult task for full-length proteins, it is instead more easily achievable for protein domains. Indeed, small structurally-conserved protein domains can be independently expressed while still preserving their individual functions (Fields and Song, 1989; Heger and Holm, 2003). Thus, screening a library encompassing the entire collection, or mostly, of functional domains encoded by a genome (Domainome) could provide a simple method to annotate gene products, including those encoding RBPs. When combined with next generation sequencing (NGS), these "domainome" libraries can be used as universal screening tools for the most diverse purposes, including rapid interactome profiling (D'Angelo et al., 2013; Di Niro et al., 2010). Among the methodologies for *in vitro* selection of protein libraries, phage display has been successfully employed to identify protein partners of desired target RNAs (Laird-Offringa and Belasco, 1996; Patrucco et al., 2015).

In vitro phage display selection to study RNA-protein interactions

Phage display was created by G. Smith in 1985 (Smith, 1985) as a method for presenting polypeptides on the surface of lysogenic filamentous bacteriophages. Since then, this method has become one of the most effective ways for producing large amounts of peptides, proteins and antibodies.

Even if traditionally employed to investigate protein-protein interactions, phage display has been successfully used to investigate RNA-protein interactions as well (Danner and Belasco, 2001; Laird-Offringa and Belasco, 1996). Recently, Patrucco and colleagues proposed a novel platform exploiting phage display technology to profile the interacting proteome of selected RNAs (Patrucco et al., 2015). The approach combines the selection of a phage library displaying "filtered" open ORFs with next-generation DNA sequencing. In particular, the pipeline was validated by using a well-characterized RNA/RBP pair and proved to be efficient in isolating both known and new RBPs specific for the target RNA.

Phage display is based on a direct linkage between phage phenotype and its encapsulated genotype, which leads to presentation of molecule libraries on the phage surface. The gene encoding the displayed molecule is packed within the same virion as a single-strained DNA (ssDNA) and the displayed peptides or proteins are expressed in fusion with phage coat protein (Kaplan and Gershoni, 2012). This coupling between genotype and phenotype ensures that identical phage particles will

be obtained from the same infected *Escherichia coli* clone.

The phage display technology has provided the ability to create protein libraries containing a great number of phage particles, encoding and displaying different molecules. Phage libraries can be used to investigate interactions with desired targets (proteins, RNA or DNA). Biopanning—the procedure of specific binders selection—is essential for enriching the desired molecule level. The selection-driving target is immobilized on a solid support, incubated with the phages and repeated cycles of incubation, washing, amplification and re-selection of bound phages are carried out (**Figure 10**).

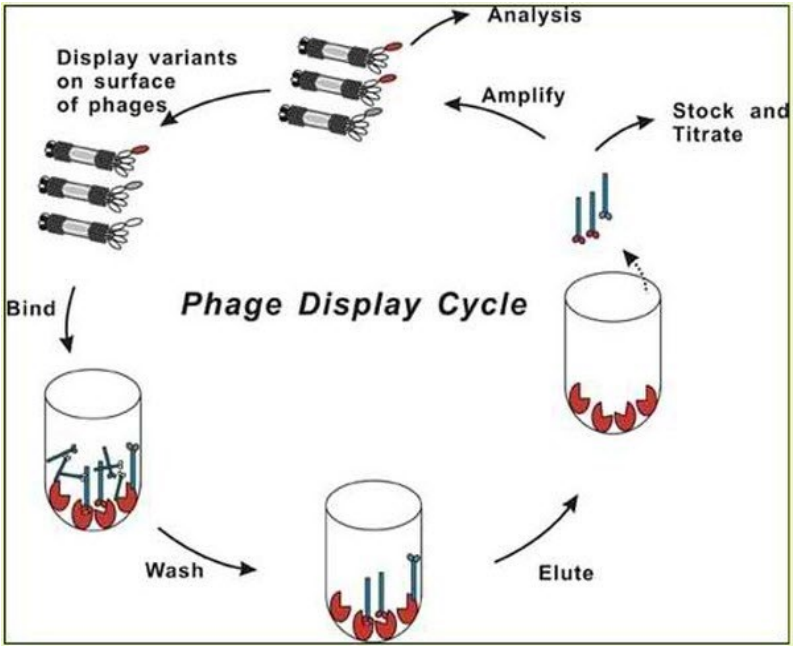


Figure 10. The phage display cycle. Phages are amplified and exposed to a desired target. Bound phages are eluted and subsequently re-amplified for further rounds of selection. Unbound phages are eliminated by stringent washes. Modified from http://phages.org/phage-display/.

Most phage libraries are built by taking advantage of phagemids. These are plasmids (4.6 kilobases) which encode a signal sequence, the phage coat protein and an antibiotic resistance marker. The fragment/polypeptide of interest is cloned upstream of the coat protein sequence and expression is controlled by the use of a promoter such as lacZ. A phagemid cannot produce infective phage particles alone, in fact, a helper phage is required, which provides the genes essential for phage replication and assembly, including a wild-type copy of the coat protein used for display (Carmen

and Jermutus, 2002).

Application of ß-lactamase-based filtering to randomly fragmented DNA from diverse sources allows generation of phage libraries enriched for genic ORFs (D'Angelo et al., 2011; Di Niro et al., 2010; Zacchi et al., 2003). This relies on the assumption that only fragments of functional ORFs are likely able to form the foldable domains which do not affect the correct folding and activity of the fused ß-lactamase reporter protein. Conversely, random ORFs do not fold into coherent domains and lead to aggregation, misfolding and inactivation of the reporter.

AIMS OF THE STUDY

Although lncRNAs represent the major transcriptional output of mammalian cells, only a limited number of these have been associated with specific functions. AS Uchl1 and, more in general, SINEUPs constitute a family of natural AS lncRNAs enhancing translation of complementary target mRNAs. Two discrete domains are essential in SINEUPs: an overlapping region to a target mRNA or BD dictates target specificity, while an embedded inverted SINEB2 element or ED confers biological activity. Taken advantage of SINEUPs modular organization, artificial SINEUPs can be re-directed against desired mRNA by solely manipulating the BD.

The domain architecture of SINEUPs support the model of lncRNAs working as flexible modular scaffolds, with each functional unit being independently capable of recruiting different molecular effectors. However, if the role attributed to the BD is to "guide" and mediate SINEUPs interactions with target mRNAs, the mechanisms determining the biological activity of the inverted SINEB2 as SINEUPs ED remain unclear. SINEB2 elements present a highly conserved secondary structure, deriving from the RNA molecules from which they emerged (tRNAs). Interestingly, in most lncRNAs a strict link between structure and function has been demonstrated, and often the presence of structural motifs is required for engaging molecular partners modulating lncRNAs biological activity.

In this context, in first place we investigated the molecular mechanisms regulating SINEUPs activity by focusing on:

1) the structural basis for activation of protein synthesis mediated by SINEUPs ED in natural and synthetic SINEUPs

2) the identification of SINEUPs protein partners.

As further long-term goal, to asses SINEUPs scalability for potential applications in biotechnology as well as in therapy, we aimed at:

3) validating SINEUPs as RNA tools to increase protein synthesis.

MATERIALS AND METHODS

Constructs

Plasmids expressing AS GFP, AS Uchl1, AS Uchl1 ΔB2, AS Uchl1 ΔAlu and AS Uchl1 ΔTE (previously referred to as ΔTOT) were previously described (Carrieri et al., 2012). PcDNA3.1- (Thermo Fisher Scientific) and pEGFP-C2 (Clontech) were commercially available vectors.

AS Uchl1 mutant lacking the SL1 domain within the invSINEB2 element (Δ68-77 nt) was obtained by gene synthesis and cloned between XbaI/HindIII sites in pcDNA 3.1(-).

AS GFP was used to create AS GFP ΔSL1 mutant. Mutagenesis PCR primers were designed by QuikChange Primer Design Program (Agilent).

For the bacterial expression of GST-fusion products, ORF fragments were excised with BssHII-NheI from the phagemid DNA and subcloned into a custom-designed pGEX-FLAG expression vector as previously described (Heger and Holm, 2003). The vector contains a FLAG-tag (DYKDDDDK) for C-terminal tagging of expressed proteins.

Biopanning procedures

For biopanning experiments, phage particles were resuspended in PBS buffer at a concentration of 10^{11} cfu/µl and for each selection 10^{12} phages were used. Production and rescue of phagemids were carried out according to published protocols (D'Angelo et al., 2013). The ORF phage library used in this study has been described previously (Di Donato et al., 2004).

Selections were driven by two SINEUP-related baits: i) AS Uchl1 Δ5' and ii) inverted SINEB2 from AS Uchl1. These were transcribed *in vitro* (IVT) with commercially available kits (MEGAscript® T7 Kit Ambion for AS Uchl1 Δ5'and MEGAshortscript™ T7 Transcription Kit for the inverted SINEB2), using original plasmids as template (Carrieri et al., 2012). IVT RNAs were 3' biotinylated using Pierce RNA 3' End Biotinylation kit (Thermo Fisher Scientific).

Pre-clearing of phages was performed prior to start with selection: 20 ul of streptavidin-coated magnetic beads (New England Biolabs) were washed in TENT

buffer (10 mM Tris HCl pH 8.0, 1 mM EDTA, 250 mM NaCl, 0,5% triton X-100) and incubated with 10^{12} phages in TENT buffer for 30' at room temperature. This step significantly reduced the presence of "sticky" phages binding to beads or plastics. Unbound phages were recovered and employed for selection.

Selections were carried out as follows: 3 pmoles of IVT RNA bait were diluted in TENT buffer implemented with RNAse inhibitors (SUPERase In™ RNase Inhibitors, Thermo Fisher Scientific) and incubated with pre-washed streptavidin magnetic beads for 20' at room temperature and subsequently washed. Pre-cleared phages were then added and left for 45' at room temperature, in the presence of competitors. In particular, for each bait we carried out two parallel selection protocols, different in the competitor used (1 ug/ul tRNAs from *E. coli*, SIGMA ALDRICH or ssDNA oligos from herring sperm). Complexes were extensively washed, to get rid of unbound phages. RNA-binding phages were finally eluted in the presence of RNAse A (10 ug/ml in 10 mM Tris pH 8.0, 1 mM EDTA, 15 mM NaCl, 2' treatment at room temperature). These were used for infection of *E. Coli* DH5α at OD_{600} 0.5 at 37°C for 45', thus giving rise to one selected library specific for each RNA bait. The selected phage pool was amplified in *E. coli* DH5α and the biopanning procedure was repeated for a second round. Higher stringency was achieved by increasing the number of washing steps. Two consecutive selection cycles previously proved to be efficient in screening out without introducing significant restrictions in output diversity (Fan and Steitz, 1998). After the second round of selection, colonies growing on agar plates were harvested and plasmid DNA was isolated by standard miniprep procedure.

NGS and bioinformatic analysis

CDNA inserts were PCR-amplified with barcoded Molecular Identifier (MID) tagged–primers (MID-primers) and sequenced by Illumina SMARTSeq platform. Bioinformatic analysis were performed with NGS-Transcriptome profile explorer (Trex) system (Boria et al., 2013). Sequences were mapped onto the human genome (NCBI build 36) and matching sequences were compared with annotated genes.

Rescue of phagemid clones by inverse PCR

Rescue of phagemid clones was performed as described previously (Patrucco et al.,

2015a). Briefly, a pair of specific back-to-back outward primers was design for each of the tested genes, centering on the nucleotide region identified by the overlapping reads. PCR was performed with a Phusion High-Fidelity DNA Polymerase (Thermo Scientific). PCR products were gel purified, phosphorylated with T4 polynucleotide kinase, ligated by T4 DNA ligase and transformed. Colonies growing on ampicillin plates were randomly picked and grown into 1ml 2X TY medium.

GST-fusion proteins expression and purification

ORF fragments subcloned in pGEX-FLAG were transformed into *E.Coli* BL21 (DE3) cells. Bacterial cultures (100 ml) were grown at 28°C until OD_{600} = 0.5, induced with 1 mM IPTG for 3 hours and centrifuged. Bacterial pellets were resuspended in lysis buffer (PBS containing 1% Triton X-100, 200 µg/ml lysozyme, 20 µg/ml DNAse, protease inhibitors, Roche), incubated for 30' at 4°C and sonicated for 2-3 minutes. Cell debris were removed by centrifugation and supernatants combined with glutathione-agarose beads (Sigma-Aldrich) for 1h at 4°C under gentle rotation. After three washes in PBS-Tween 0.1% followed by three more in PBS, GST fusion proteins were eluted in 750 ul elution buffer (50 mM reduced glutathione, 100 mM NaCl, pH 8.0). Proteins were dialyzed against PBS and checked for purity and concentration by SDS-PAGE. Quantitative densitometry of Coomassie Blue-stained proteins was calculated with ImageJ software (Shu et al., 2006) using BSA as reference for protein quantification. GST-fusion proteins integrity was determined by western blotting, using two different monoclonal antibodies: anti-GST (clone GST-2, Sigma-Aldrich) and anti-FLAG (Sigma-Aldrich).

ELISA

Screening of selected clones in ELISA-based assays, either in the phage format or as soluble GST-fusion polypeptides, was performed according to standard protocols described previously (Danner and Belasco, 2001), with some modifications. Briefly, phage ELISA was performed with Microlon plates (Greiner) coated overnight at 4°C with 10 µg/ml streptavidin. After blocking and rinsing wells in TENT buffer, biotinylated RNA oligonucleotides (5 pmoles/well, diluted in 100 µl TENT buffer implemented with RNAse inhibitors) were captured on the plates. Phage-containing

supernatants of individual clones, diluted 1:1 in TENT buffer with RNAse inhibitors, were added to the wells and incubated for 45'. Following three washing steps, incubation with HRP-conjugated anti-M13 monoclonal antibody (GE Healthcare) for 1 hour at room temperature was carried out. Signal was revealed with TMB (3, 3', 5, 5' tetramethylbenzidine) and read at A_{450} using a Victor^{TM} X4 multilabel plate reader (Perkin Elmer). ELISA on soluble GST-fusion polypeptides was performed as described above. In particular, baits-coated wells were subsequently incubated 1 hour at room temperature with the purified proteins, extensively washed in TENT buffer and again incubated 1 hour with a mouse monoclonal anti-GST antibody (clone GST-2, Sigma-Aldrich) 1:5000 in TENT buffer. Following 1 hour incubation with a HRP-conjugated secondary antibody (Sigma-Aldrich), signal coming from RNA-protein binding was revealed as described above.

Cell culture and transfection

Neuro2a cells were obtained from ATCC (Cat. No. ATCC-CCL-131) and maintained in culture with Minimum Essential Medium Earle's Salt + GlutaMAX™-I (MEM, Gibco by Life Technologies, Cat. No. 41090-028) supplemented with 10% FBS (Sigma) and 1% antibiotics (penicillin/streptomycin), as suggested by the vendor. HEK 293T/17 cells were obtained from ATCC (Cat. No. ATCC-CRL-11268) and maintained in culture with Dulbecco's Modified Minimum Essential Medium (DMEM, Gibco) supplemented with 10% FBS (Sigma) and 1% antibiotics (penicillin/streptomycin), as suggested by the vendor. MN9D cells were obtained from prof. Michael Zigmond at University of Pittsburgh and maintained in culture with Dulbecco's Modified Minimum Essential Medium (DMEM, Gibco) supplemented with 10% FBS (Sigma) and 1% antibiotics (penicillin/streptomycin).
For structure/function validation experiments, Neuro2a and HEK 293T/17 cells were plated in 6 well-plates the day before transfection at 50% confluency and transfected with AS Uchl1 or AS GFP plasmids, respectively, using Lipofectamine® 2000 (Life Technologies™) and following manufacturer's instructions.
For RNA immunoprecipitation (RNA-IP) experiments, 2.5X 10^6 HEK 293T/17 were plated in 10 cm plates and transfected after 16-20 hours with AS Uchl1 plasmid using FuGENE® HD Transfection Reagent (Promega), following manufacturer's

instructions. Data of RNA and protein were obtained from the same transfection in each replica.

For nucleocytoplasmic fractionation experiments in HEK 293T/17, 4X10^5 cells were plated in 6-well multiwell and transfected after 16-20 hours with AS Uchl1, AS Uchl1 ΔB2, AS Uchl1 ΔAlu or AS Uchl1 ΔTE, following Fugene HD manufacturer's instructions. For nucleocytoplasmic fractionation experiments in N2A and MN9D, 3X10^6 cells were plated in 100-mm dishes and transfected after 16-20 hours with AS Uchl1, using Lipofectamine® 2000 (Life Technologies™) and following manufacturer's instructions.

RNA-IP

All the stock solutions were prepared with clean and pure reagents diluted in DEPC/Ambion water (AM9932, Ambion™), using RNAse-free equipment. Lysis and washing buffers were freshly-prepared and kept on ice; all steps, including centrifugation, were performed at 4°C.

In vivo formaldehyde fixation

At 48 hours post-transfection cells were washed once in ice-cold PBS, scraped and subsequently washed twice in 10 ml of cold PBS (3000 rpm, time: 4'). Cells were fixed in 1% formaldehyde (AR grade, Mallinkrodt) in PBS for 10' at room temperature with slow mixing and quenched in 0.25 M glycine (pH 7) at room temperature for 5'. Cells were subsequently harvested by centrifugation at 3000 rpm for 4' and washed twice with ice-cold PBS.

Beads blocking and coating

100 ul of M-280 Sheep anti-Mouse dynabeads (Life Technologies) were washed 3 times in washing buffer (0.1% BSA in PBS), blocked with 3 washes in 0.5% BSA in PBS and finally washed twice in RIP buffer (25 mM Tris-HCl pH 7.4, 150 mM KCl, 0.5% IGEPAL CA-630, 5 mM MgCl2, 0.5 mM DTT, protease inhibitors and 20 U/ml Superase RNA inhibitors). Coating with antibody/control IgG was carried out by overnight incubation of blocked beads with 20 ug of anti-DRBP76 (BD Science) or 20 ug of normal mouse IgG (Sigma-Aldrich) in a final volume of 180 ul.

Lysis and IP

Lysis was performed using 1 ml RIP buffer. Lysates were solubilized by sonication

with two short pulses (15 sec). Between the two cycles samples were kept on ice for at least 2'. Insoluble material was removed by centrifugation at 13,000 rpm for 10'. Total lysate was pre-cleared via incubation with 100 ul of uncoated blocked beads for 30' at 4°C with gentle rotation. After recovering from beads, the lysate was split and incubated with specific antibody or control IgG-coated beads overnight on a rotary platform at 4°C. 1/20 of total pre-cleared lysate was kept before splitting as INPUT.

Washing steps

Beads/Ab/lysate complexes were washed six times (5' the first and last wash/ 1' the remaining ones) in RIP washing buffer (same as RIP buffer, but with 300 mM KCl) in a cold room.

Reversal of cross-linking and elution

For reversal of cross-linking and elution, beads containing the IP samples were resuspended in 100 ul of elution buffer (50 mM Tris-Cl pH 7.0, 5 mM EDTA, 10 mM DTT and 1% SDS) and incubated at 70°C for 45'. Supernatants were recovered and resuspended in 1 ml of Trizol (Ambion™/Life Technologies) and both RNA and proteins were extracted according to the manufacturer's instructions.

RNA isolation, reverse transcription and quantitative Real Time-PCR (qRT-PCR)

RNA was extracted using TRIZOL reagent (Ambion™/Life Technologies), following manufacturer's instructions. RNA was eluted and treated with TURBO DNA-free Kit (Ambion) for 15' at 37°C, to avoid plasmid DNA contamination. RNA quality was finally checked on a formaldehyde agarose gel.

cDNA was prepared from 250 ng of purified RNA using iSCRIPT^{TM} cDNA Synthesis Kit (Bio-Rad), according to manufacturer's instructions. In particular, for RNA-IP experiments, equal volumes of DNAse-treated RNA samples were used for reverse transcription.

In order to monitor the efficiency of DNAse treatment, an equal amount of each RNA sample was retrotranscribed in the absence of RT.

qRT-PCR reaction was performed on diluted cDNA (1:2.5) using SYBR-Green PCR Master Mix (Biorad) and an iCycler IQ Real time PCR System (Bio-Rad). In

particular, in RNA-IP experiments, undiluted cDNA was used as qRT-PCR input.
Oligonucleotide sequences of primers used in this study for GAPDH, beta-actin, Uchl1, GFP, AS Uchl1 (primers 5') and AS Uchl1 (primers 3', used for detection of AS GFP) were previously described (Carrieri et al., 2012). Primers used for detection of UBC were previously described (Kuwano et al., 2008). Primers used for detection of precursor-rRNA (pre-RNA) were previously described (Murayama et al., 2008; Oie et al., 2014).
The amplified transcripts were quantified using the comparative Ct method and relative gene expression was calculated with the ΔΔCt method (Schmittgen and Livak, 2008).

Western Blot

For Western Blot (WB) analysis, cell pellets were directly dissolved in Laemmli sample buffer. For RNA-IP experiments in particular, ILF3 IP efficiency was monitored by loading the whole fraction of proteins recovered from the organic phase after Trizol extraction, resuspended in Laemmli sample buffer. All lysates were briefly sonicated, boiled and loaded on 10% (for ILF3) or 12% (for UCHL1 and GFP) poly-acrylamide gels. Immunoblotting was performed with the following primary antibodies: anti-UCHL1 (Millipore, Cat. No. AB1761-I) 1:5000; anti-GFP (Living Colours), 1:8000; anti-DRBP76 (BD Science), 1:500 overnight; anti-β-actin (SIGMA ALDRICH) 1:2000; anti-NONO (SIGMA ALDRICH), 1:500. Signals were revealed after incubation with horseradish peroxidase-conjugated (HRP) secondary antibodies (DakoCytomation, Glostrup, Denmark) 1:1000 or anti-protein A HRP (Sigma-Aldrich) for 1 hour at RT, in combination with ECL (GE Healthcare, Cat. No. RPN2105). Image detection was performed with Alliance LD2-77WL system (Uvitec, Cambridge).

Cell fractionation

Nucleocytoplasmic fractionation was performed as previously described (Wang et al., 2006). Cell handling and centrifugation steps were done at 4°C; all buffers were kept on ice.
Briefly, at 48 hours post-transfection cells were washed in ice-cold PBS, scraped on ice and collected in 1.5 ml eppendorf tubes. A fraction (usually 1/10) of cells was

kept for total lysate analysis. Cells were harvested by centrifugation (3', 4000) at 4°C, washed twice in PBS and resuspended in 1 ml of RSB (10 mM Tris, pH 7.4, 10 mM NaCl, 3 mM MgCl2). Lysates were incubated 3' on ice and subsequently centrifuged at 4000 rpm for 3' at 4 °C.

Supernatants were discarded and cell pellets resuspended in a volume of RSBG40 (10 mM Tris, pH 7.4, 10 mM NaCl, 3 mM MgCl2, 10% glycerol, 0.5% Nonidet P-40, 0.5 mM dithiothreitol -DTT, and 10 U/ml SUPERase) equal to 4X pellet volume, by slow pipetting. Nuclei were pelletted at 7000 rpm for 3' and supernatants recovered and saved as the first (soluble) cytoplasmic fraction. Nuclear pellets were again resuspended in RSBG40 implemented with 1/10 volume of detergent (3.3%-wt/wt sodium deoxycholate and 6.6%-vol/vol Tween 40) and incubated on ice for 5'. Following centrifugation (3', 7000 rpm) supernatant (insoluble cytoplasmic fraction) was pooled with the previous, whereas pellets were washed three times more in 1 ml of RSBG40 and finally spinned at 10,000 rpm for 5'. The resulting pellet corresponded to nuclear RNA fraction. RNA samples were resuspended in 1 ml of Trizol (Invitrogen) and RNA isolated using following manufacturer's instruction. RNA was eluted and treated with Turbo DNAse (Ambion). The purity of fractions was confirmed by qRT-PCR on pre-ribosomal RNA and GAPDH.

ILF3 silencing

4X10^5 HEK 293T/17 were harvested on a 6-well multiwell and co-transfected with 4 ug of AS Uchl1 plasmid and 4 ug of ILF3 siRNA (Mission esiRNA, mouse ILF3, SIGMA ALDRICH)/control siRNA (All Stars Neg. Control siRNA, Qiagen) with 10 ul of Lipofectamine® 2000 (Life Technologies™) in serum-free DMEM with no antibiotics. After 24h, a second round of transfection was performed, using 2 ug of both plasmid and siRNA. On the following day medium was changed with 10% FBS-DMEM. At 48h from the second transfection cells were collected for fractionation. 1/20 of total cells were resuspended in Laemmli sample buffer for WB analysis of ILF3 protein levels in silenced/control cells. Nucleocytoplasmic fractionation was performed as described before and cell fractions resuspended in 1 ml of Trizol.

ILF3/p54^{nrb} co-immunoprecipiation (co-IP)

ILF3 IP was performed as described above (section: RNA-IP). The presence of p54^{nrb} in IP complexes was revealed with anti-NONO antibody (SIGMA) prior to detect ILF3, on the same nitrocellulose membrane.

Immunofluorescence microscopy

Cells were fixed in 4% paraformaldehyde in PBS for 10' at room temperature, washed twice in PBS and treated with glycine 0.1 M in PBS for 5' at room temperature. Following two more washes in PBS, fixed cells were permeabilized with 0.1% Triton X-100 for 4' at room temperature and blocked with 0.2% BSA, 1% FBS, and 0.1% Triton in PBS for 5' at room temperature. Cells were incubated 90' with anti-DRBP76 (BD Bioscience) 1:50 in blocking solution at room temperature, washed in PBS three times and finally stained with Alexa Fluor-488 or Alexa Fluor-594 (Invitrogen)-labelled anti-mouse or anti-rabbit secondary antibodies, 1:250 in blocking buffer. Nuclei were visualized with DAPI (1 ug/ml). Homemade anti-DJ-1 1:250 (Foti et al., 2010) or anti-thyrosin hydroxylase (TH) (Chemicon) 1:1000 were used to counterstain cell cytoplasm. Images were captured with a confocal microscope (LEICA TCS SP2).

Statistical analysis

All data are expressed as mean ± standard deviation on n≥3 replicas. Statistical analysis was performed using Excel software. Statistically significant differences were assessed by Student's t-test. Differences with $p<0.05$ were considered significant.

UNPUBLISHED DATA OF PROFESSOR S. GUSTINCICH LABORATORY IN COLLABORATION WITH PROFESSOR J. PLAVEC LABORATORY

Secondary structure of the inverted SINEB2 embedded in AS Uchl1 RNA

The secondary structure of the inverted SINEB2 element (invSINEB2/183) contained in AS Uchl1 was determined using chemical footprinting. The RNA was *in vitro* transcribed from a plasmid containing the invSINEB2 sequence of AS Uchl1. DMS (dimethyl sulfate) and CMCT (1-cyclohexyl-(2-morpholinoethyl) carbodiimide metho-p-toluene sulfonate) were used as methylating agents. DMS preferentially methylates positions N1 and N3 of adenines and cytosines, respectively, while CMCT methylates position N3 of uridines and to a lesser extent N1 of guanines. The level of methylation is directly related to the accessibility to solvent of potential modification sites. Therefore, hydrogen bonded nucleotides are not methylated, while non-hydrogen bonded are. The methylation sites were analyzed by reverse transcribing RNA into cDNA starting from a fluorescently labelled DNA primer. The DNA oligos were analyzed on large sequencing gels and visualized on a densitometer. Data from footprinting studies has been used as an input for restrained mFOLD secondary structure prediction. It is noteworthy that data from either of the chemicals was sufficient for an unambiguous secondary structure determination. The invSINEB2/183 RNA folds into a structure with mostly helical secondary structure elements (**Figure 11**).

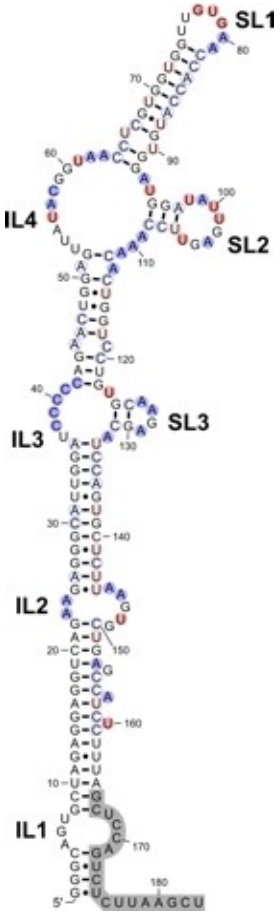


Figure 11. Secondary structure of the inverted SINEB2/183 ED of AS Uchl1. tDMS and CMCT reactive nucleotides are shaded in blue and red, respectively. Internal loops and stem-loops are labelled as ILx and SLx, respectively. Non-reactive nucleotides are only circled. The segment shaded in gray corresponds to the DNA primer hybridization site.

The structure exhibits several bulges, asymmetric internal loops and hairpins. Consisting of nucleotides 5-8 and 167-172, the internal loop (IL1) could not be directly probed with DMS or CMCT due to hybridization of the fluorescent DNA primer to the 3' of the RNA. This internal loop is followed by a helical region, which contains three single nucleotide bulges, with nucleotides G154, A157 and U160 showing weak reactivity with methylating agents. The asymmetric internal loop (IL2) is comprised of nucleotides 23-25 and 144-149. Data suggests that G26:U144 base pair is not formed according to chemical footprinting as U144 is reactive with CMCT, while U143 remains protected. This is to be expected, due to the relatively low stability of GU base pairs, especially at the termini of helical regions.

The invSINEB2/183 construct features two more internal loops, which branch out

into short hairpins. Comprised of nucleotides 37-41 and 123-132, the internal loop (IL3) is branched into a short stem-tetraloop element (SL3). Similarly, nucleotides 53-63 and 93-111 form a larger internal loop (IL4) with a stem-heptaloop motif (SL2). The terminal hairpin (SL1) includes nucleotides 64-92 and exhibits a G/C rich stem with an A/U rich loop region. All stem nucleotides up to C64 and G92 are protected from methylation, including U66:U90 mismatch nucleotides. On the other hand, loop nucleotides G77, U78 and G79 are all susceptible to methylation by CMCT. Importantly, U80 can be methylated by DMS while U81 exhibits very weak reactivity. Partial solvent access suggests that the two A:U base pairs are involved in an equilibrium between opened and closed states.

RESULTS

1 Structural basis for activation of protein synthesis mediated by the ED in natural and synthetic SINEUPs

1.1 The terminal SL1 hairpin contributes to AS Uchl1 ability to increase UCHL1 protein levels

As previously introduced, the embedded invSINEB2 element acts as ED in natural SINEUP AS Uchl1. Deletion of invSINEB2 sequence, but not of the embedded Alu repeat, abolishes UCHL1 protein up-regulation mediated by AS Uchl1 in mouse neuroblastoma cell lines (Carrieri et al., 2012). We investigated whether secondary structure components of invSINEB2 affect the function of AS Uchl1 RNA. We focused our attention on the terminal SL1 as, according to chemical footprinting data, it is the most stable secondary structure element within invSINEB2. The terminal stem-loop hairpin structure was thus disrupted by deleting nucleotides 68-77 of invSINEB2 (ΔSL1) from full length AS Uchl1 (ΔSL1 mutant). This deletion would not affect the other helical regions. To investigate invSINEB2-ΔSL1 activity when embedded in full length AS Uchl1, we took advantage of murine neuroblastoma Neuro2a cells, as they express Uchl1 mRNA but do not contain detectable levels of endogenous wild type (WT) AS Uchl1. AS Uchl1 activity was defined as UCHL1 protein increase in the presence of unchanged mRNA levels, as quantified by western blotting and qRT-PCR, respectively (**Figure 12**). AS Uchl1 WT was able to increase UCHL1 protein levels while maintaining stable Uchl1 mRNA levels, as expected for a post-transcriptional regulatory mechanism. Conversely, ΔSL1 deletion mutant abolished the ability of AS Uchl1 RNA to up-regulate UCHL1 protein levels. Indeed, UCHL1 amounts were comparable in cells transfected with ΔSL1 mutant and in control samples.

Taken together, these data indicate that the terminal SL1 hairpin of embedded invSINEB2 is a structural determinant required for AS Uchl1 ability to increase protein levels as synthesized from its target mRNA.

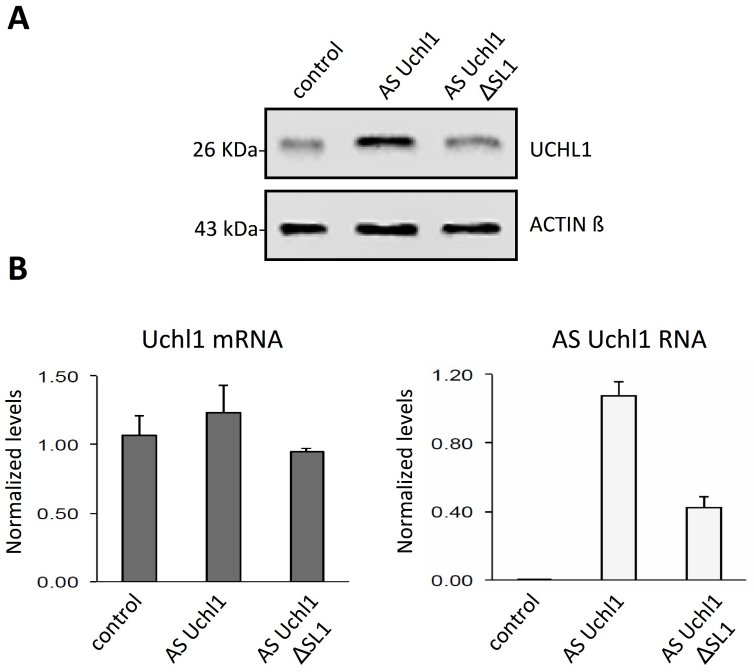


Figure 12. The SL1 hairpin contributes to AS Uchl1 activity. Murine neuroblastoma N2a cells were transfected with AS Uchl1 and ΔSL1 mutant constructs. Control cells were transfected with an empty control plasmid. 48 hours after transfection, cells were lysed and processed for protein and RNA analysis. (**A**) Western blot was performed with anti-UCHL1 antibody. β-actin was used as loading control. (**B**) Expression of Uchl1 mRNA (gray bars) and AS Uchl1 (white bars) were monitored by qRT-PCR using specific primers. Data indicate mean ± st. dev. Data are representative of N=5 independent replicas.

1.2 Conservation of the structural basis for activation of protein translation in synthetic SINEUPs

A synthetic SINEUP against EGFP, named AS GFP, was generated previously (Carrieri et al., 2012). AS GFP is able to increase GFP protein translation when double transfected with corresponding S GFP DNA in HEK 293T cells (Carrieri et al., 2012). As in natural SINEUPs, the ability to enhance protein synthesis relies on the embedded inverted SINEB2 element. Following the identification of the SL1 as structural determinant for the ED-mediated translation activation in natural SINEUP AS Uchl1, we assessed whether similar regions of structural stability are involved in the activity of synthetic SINEUPs as well. To this purpose, we generated an AS GFP ΔSL1 mutant and tested its ability to modulate GFP translation in HEK 293T/17 cells. Cells were co-transfected with S GFP and canonical or mutated AS GFP. As negative control, GFP was co-transfected with an empty vector. At 48 hours post-

transfection cells were collected for protein and RNA analysis. Conversely to canonical AS GFP, which successfully enhanced GFP protein production, AS GFP ΔSL1 mutant failed in up-regulating GFP translation (**Figure 13**).

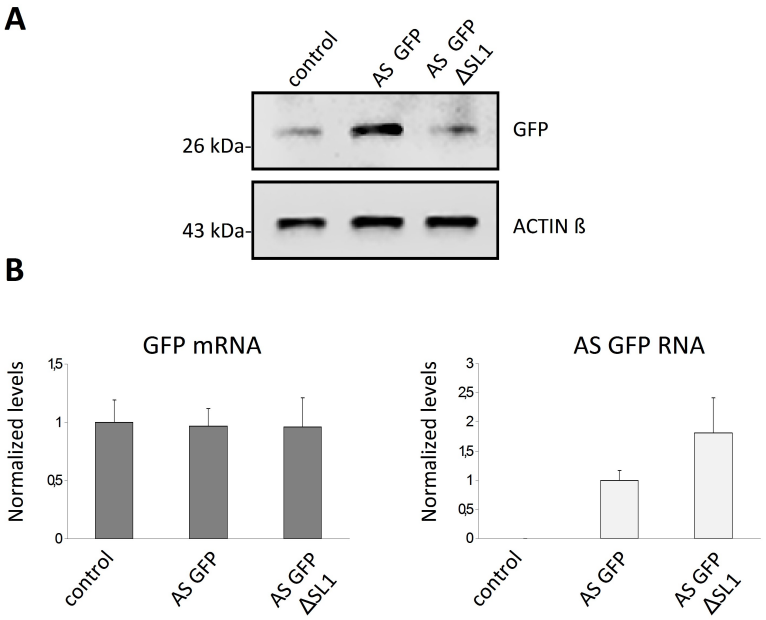


Figure 13. The SL1 is a structural determinant for synthetic SINEUPs activity. HEK 293T/17 cells were co-transfected with pEGFP and AS GFP or AS GFP ΔSL1 plasmids. Control cells received an empty control plasmid. 48 hours after transfection, cells were lysed and processed for protein and RNA analysis. (**A**) Western blot was performed with anti-GFP antibody. β-actin was used as loading control. (**B**) Expression of GFP mRNA (gray bars) and AS GFP (white bars) were monitored by qRT-PCR using specific primers. Data indicate mean ± st. dev. Data are representative of N=3 independent replicas.

Stable GFP mRNA levels in control, canonical and mutated AS GFP-transfected samples excluded that differences in protein quantity could be a consequence of different transfection efficiency. Taken together, these results suggest that the structural-based SL1 domain within inverted SINEB2 element is essential to maintain the ED activity in synthetic SINEUPs as well.

2 Identification of SINEUPs protein partners

2.1 Identification of proteins that bind SINEUPs ED through ORF phage display selection

Once defined the structure of the inverted SINEB2, as well as the presence of highly-stable structural determinants involved in the activation of protein synthesis, we found intriguing to search for proteins binding and regulating the activity of SINEUPs ED. Indeed, many RBPs recognize specific structural motifs, like stem-loops or bulges, within RNA molecules.

To this purpose, we employed a high-throughput protein expression and interaction analysis platform aimed at identifying the RNA-Interacting Domainome (RIDome). The approach combines the selection of a phage library displaying "filtered" ORFs and NGS (Di Niro et al., 2010; Patrucco et al., 2015).

The typical outcome is a list of ranked genes that directs analysis and validations to the best candidates. High-scoring ORFs are recloned from the library by inverse PCR, and their interactions are confirmed by ELISA-based assays. The interacting partners are then validated by RNA-IP and other functional assays.

The library selection was performed as reported previously (Patrucco et al., 2015), with minor modification regarding the synthesis of the RNA baits which were generated by *in vitro* transcription (IVT) from a DNA template and then enzymatically biotinylated at the 3' end (see Materials and Methods). All constructs employed in screenings and validation experiments are schematized in **Figure 14**. A summary of the RIDome pipeline is reported in **Figure 15**.

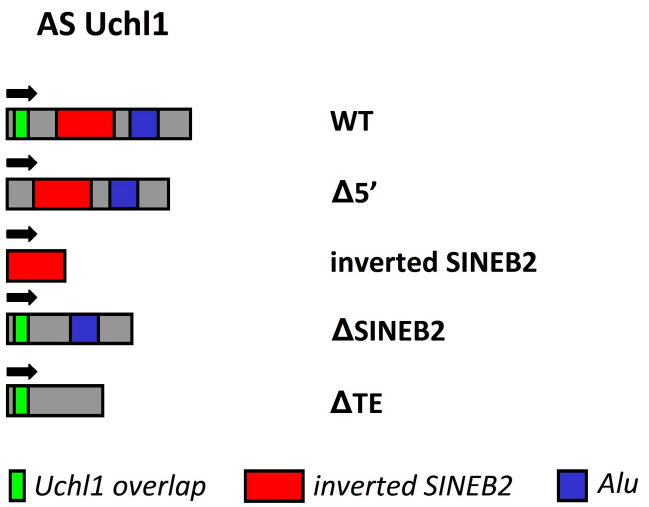


Figure 14. Constructs used for selection and validation experiments. All constructs derive from AS Uchl1 wild type (WT). AS Uchl1 WT contains the BD (green), the ED or inverted SINEB2 (red) and the Alu (blue). AS Uchl1 Δ5' lacks the 5' end, including the BD. The inverted SINEB2 is from AS Uchl1. AS Uchl1 ΔSINEB2 lacks the ED. AS Uchl1 ΔTE lacks both SINEB2 and Alu. Abbreviations: BD=binding domain; ED=effector domain; TE=Transposable Elements.

To explore the contribution of the ED in protein binding irrespective of its collocation within an AS lncRNA, we employed an IVT inverted SINEB2 as phage display selection-driving bait (**Figure 15**, first square). In parallel, we carried out a wider investigation of SINEUP-binding proteins by selecting the library on an IVT AS Uchl1 Δ5'. Lacking the BD, in which target specificity of different SINEUPs is encoded, AS Uchl1 Δ5' provides the common backbone on which synthetic SINEUPs are built and includes the ED in an embedded format (Carrieri et al., 2012). In particular, we carried out two parallel selections for each bait, differing for the competitor used (tRNA or ssDNA).

After two cycles of selection, phagemid DNAs were recovered, ORF inserts amplified and sequenced according to an Illumina SMARTSeq protocol. To limit the time-consuming analysis of large amounts of data, we decided to sequence a minimal fraction of the selection output, sufficient to contain an overview of the most likely repertoire of SINEUP-interacting ORFs.

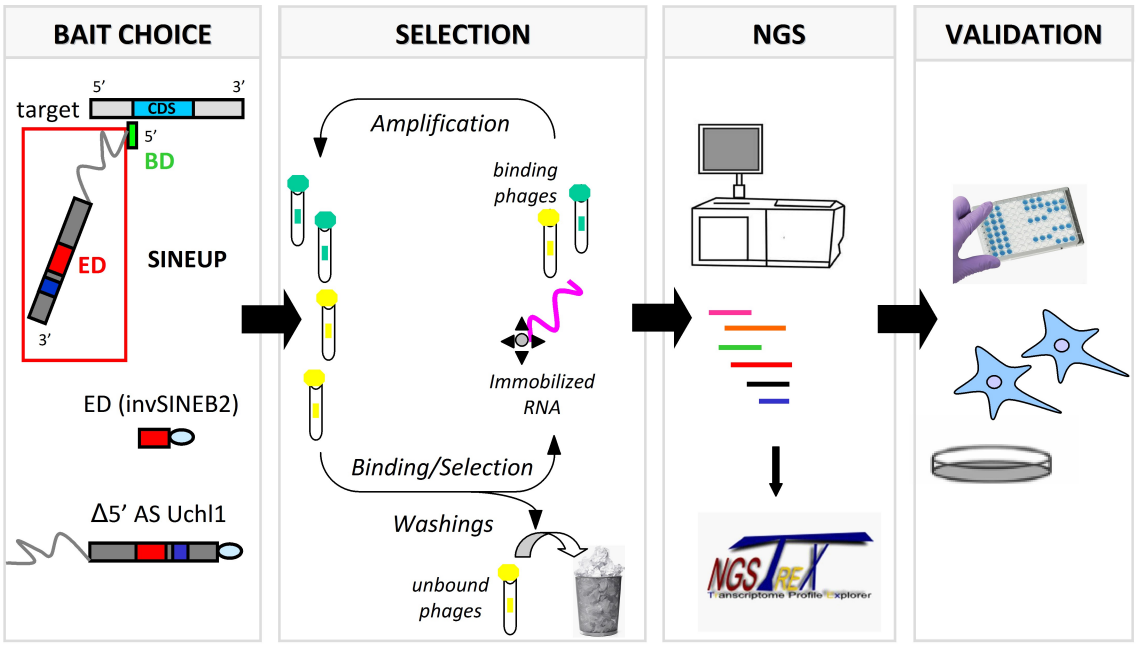


Figure 15. SINEUP-driven phage display selection. First square: schematic representation of a SINEUP and its target mRNA. The red square highlights the region where the baits are contained in SINEUPs. The inverted SINEB2 and AS Uchl1 Δ5' were used as selection-driving baits. Baits were biotinylated at their 3' end. Second square: schematic representation of the selection protocol. Screenings were carried out by taking advantage of a "filtered" Open Reading Frames (ORF) phage library displaying human protein domains. Biotinylated baits were captured by streptavidin-binding magnetic beads and incubated with amplified phage library in the presence of tRNA/ssDNA competitor. Stringency washes allowed selective isolation of phages involved in bait-binding. Third square: after two cycles of selection, selection outputs were sequenced according to Illumina SMARTSeq protocol and sequencing results analyzed with T-Rex NGS software. Fourth square: high-scoring ORF domains were cloned by inverse PCR and binding with target baits validated *in vitro* (ELISA) and *in vivo* (RNA-IP) (forth square). Abbreviations: BD=binding domain; ED=effector domain; CDS=coding sequence; RNA-IP=RNA immunoprecipitation.

2.2 NGS data analysis revealed ILF3 as the dominant candidate interactor

We collected around 100.000 reads from each selected library that were analysed with the NGS-Trex system (Boria et al., 2013) and mapped onto the human genome (NCBI build 36). Sequences matching annotated genes were then ranked as described (Di Niro et al., 2010).

As expected, the outcome of phage selections consisted of lists containing several hundreds of genes, most of them represented by very few reads and that form the "noise" of the phage selection. To remove most of such background, we arbitrarily set a threshold, and considered in successive analysis only those genes that were represented by at least 20 reads in the selected libraries and by at least four reads in the non-selected (NS) library (**Table 2**).

Bait	Competitor	Number of Reads	Mapping Reads	Average length	Genes	Genes with ≥ 20 reads
invSINE B2	ssDNA	89017	71329	235	5129	198
invSINE B2	tRNA	115501	96403	201	5255	218
AS Uchl1 Δ5'	ssDNA	94695	74686	258	3938	95
AS Uchl1 Δ5'	tRNA	137116	110929	241	5655	295
						Genes with ≥ 4 reads
NS	N/A	155880	85448	113	8128	3803

Table 2. Summary of NGS results. For each selection the total number of reads, the mapping reads and their average length are reported. The number of selected genes is shown as well. Arbitrary parameters were applied to narrow the number of selected genes. Threshold was fixed to ≥ 4 and ≥ 20 for unselected and selected libraries, respectively.

This restriction allowed us to limit the analysis to few hundreds genes in total. Enrichment analysis was performed to assess those genes that were positively selected after biopanning experiments and calculated by dividing the normalized number of reads (reads per million, RPKM) in the selected libraries *vs* the NS library. Each gene was then plotted on a dispersion graph showing the fold enrichment *vs* the total number of reads. Results from each selection are shown in **Figure 16.** Comparing the four selections, it became evident that a single gene was strongly enriched during phage selection. In fact, unlike most genes included in the analysis which showed a fold enrichment in the range 2-20 fold, the interleukin enhancer-binding factor 3 (ILF3) was enriched by >1000 times in 3/4 selections.

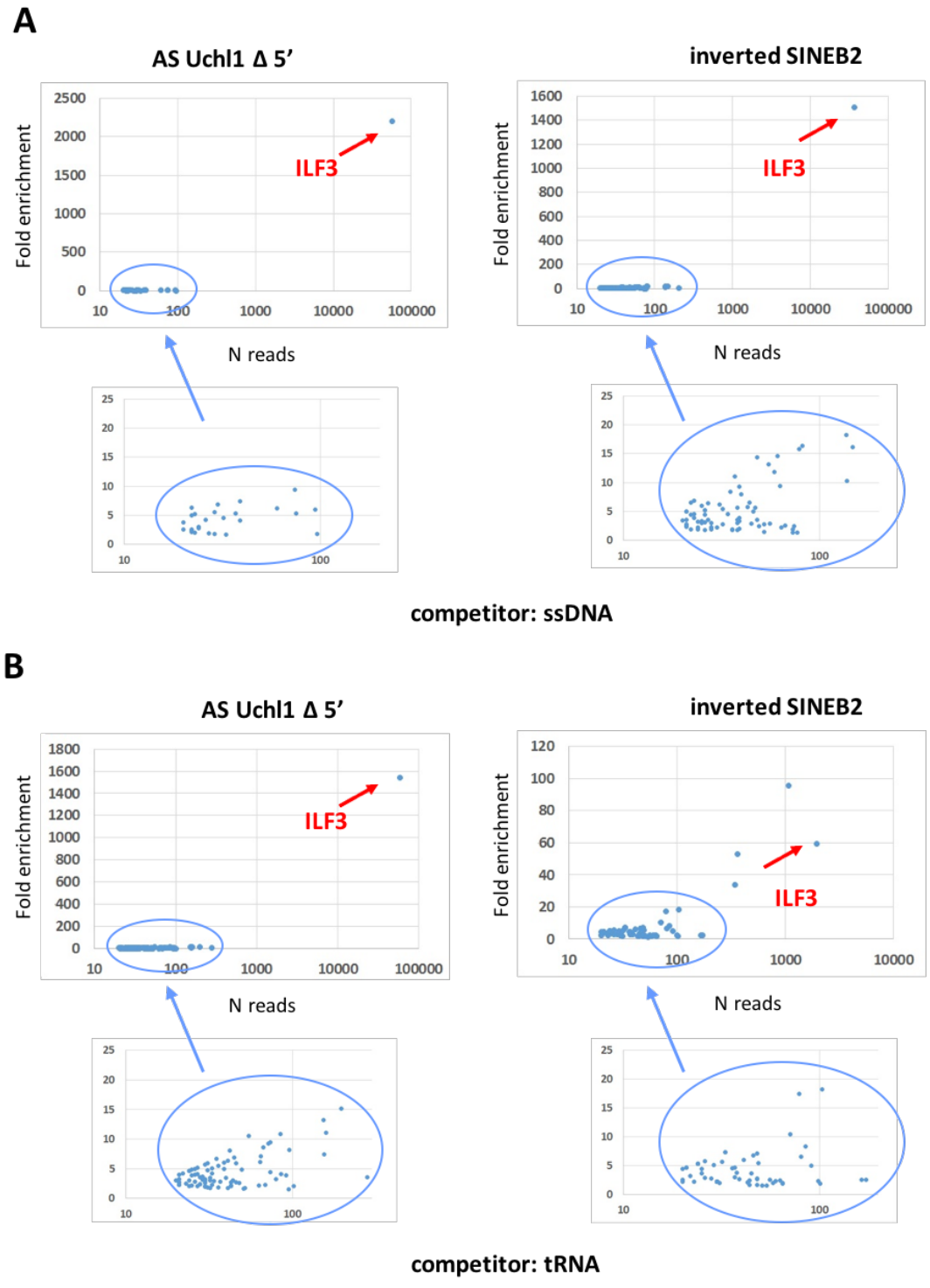


Figure 16. Summary of NGS analysis. Results from inverted SINEB2 and AS Uchl1 Δ5'selections are shown. Selections in **A** were carried out with single-stranded DNA competitor; in selections represented in **B** tRNA was employed. Enrichment analysis were performed by dividing the normalized number of reads (reads per million, RPKM) in the selected libraries *vs* the non-selected library. Genes were plotted on a dispersion graph showing the fold enrichment *vs* the total number of reads. The blue circles correspond to enlarged areas in each chart.

We then compared the gene lists from the four selections. The Venn diagram depicted in **Figure 17** shows that 18 genes are common to all four selected libraries, and 21 more genes are present in at least three samples.

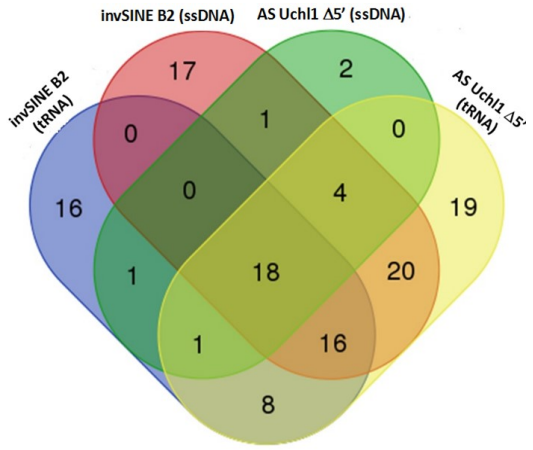


Figure 17. Comparison of enriched genes in different selections. The Venn diagram illustrates genes specifically enriched in each selection (outer area) or commonly identified in two or more selections (inner area). Blue= inverted SINEB2_tRNA; red=inverted SINEB2 ssDNA; orange=AS Uchl1 Δ5'_tRNA; green=AS Uchl1 Δ5'_ssDNA.

2.3 *In vitro* validation of selected ORFs

As a first general round of validation, we were interested in assessing whether the differential enrichment between ILF3 and the large majority of selected ORFs corresponded to differential binding specificity. To this purpose, together with ILF3, we selected a couple of clones showing lower fold enrichment (in the range of 20). In particular, we chose two well characterized RBPs, HNRNPA3 and SRSF5. HNRNPA3 was present in all selections, while SRSF5 was found in 3/4 (**Appendix Table 1 and Table 2**). Thus, corresponding clones were recovered from selected libraries, sequenced and tested in phage ELISA on the inverted SINEB2. Interestingly, ILF3 proved to be the unique clone giving a specific signal (**Figure 18**), which was coherent with the higher fold enrichment emerged from sequencing data. Therefore, we decided to focus our attention on this extensively characterized double-stranded RNA-binding protein (dsRBP) regulating several steps of mRNAs life cycle (Castella et al., 2015).

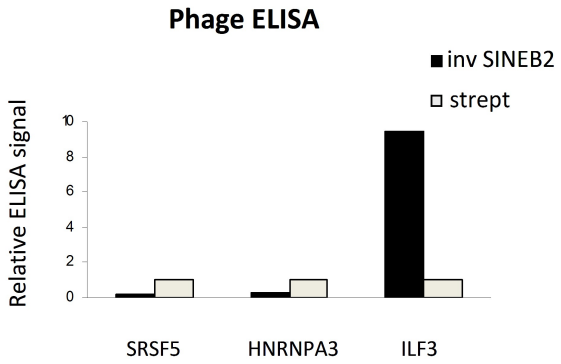


Figure 18. Preliminary screening of higher-ranking ORFs. Specificity of clones SRSF5, HNRNPA3 and ILF3 was preliminary tested by phage ELISA on the inverted SINEB2. Streptavidin served as control. Values are indicated as the fold signal *vs* background (uncoated wells).

ILF3 RNA-binding capability is mainly dependent on the presence of two double-stranded RNA-binding motives (dsRBMs), referred to as dsRBM1 and dsRBM2, shared by ILF3 main isoforms (**Figure 19B**). In particular, ILF3 clone employed in preliminary phage ELISA corresponded to dsRBM2, which represented the most enriched ILF3 domain in all selections. Indeed, individual alignment of ILF3-related reads from selected libraries followed by protein blast revealed predominant mapping on ILF3 dsRBM2 (**Figure 19A**). A minor portion of sequences mapped instead on dsRBM1 and other domains. These observations were further supported by preliminary screening and sequencing of randomly picked clones from selected libraries (data not shown). Moreover, comparison of reads landscape in selected *vs* non-selected (NS) libraries allowed to confirm that ILF3 dsRBM2 enrichment was restricted to SINEUP baits-driven selections (both with AS Uchl1 Δ5' and inverted SINEB2), thus excluding any artifacts coming from overrepresentation of the clone in the starting library (**Figure 19A**).

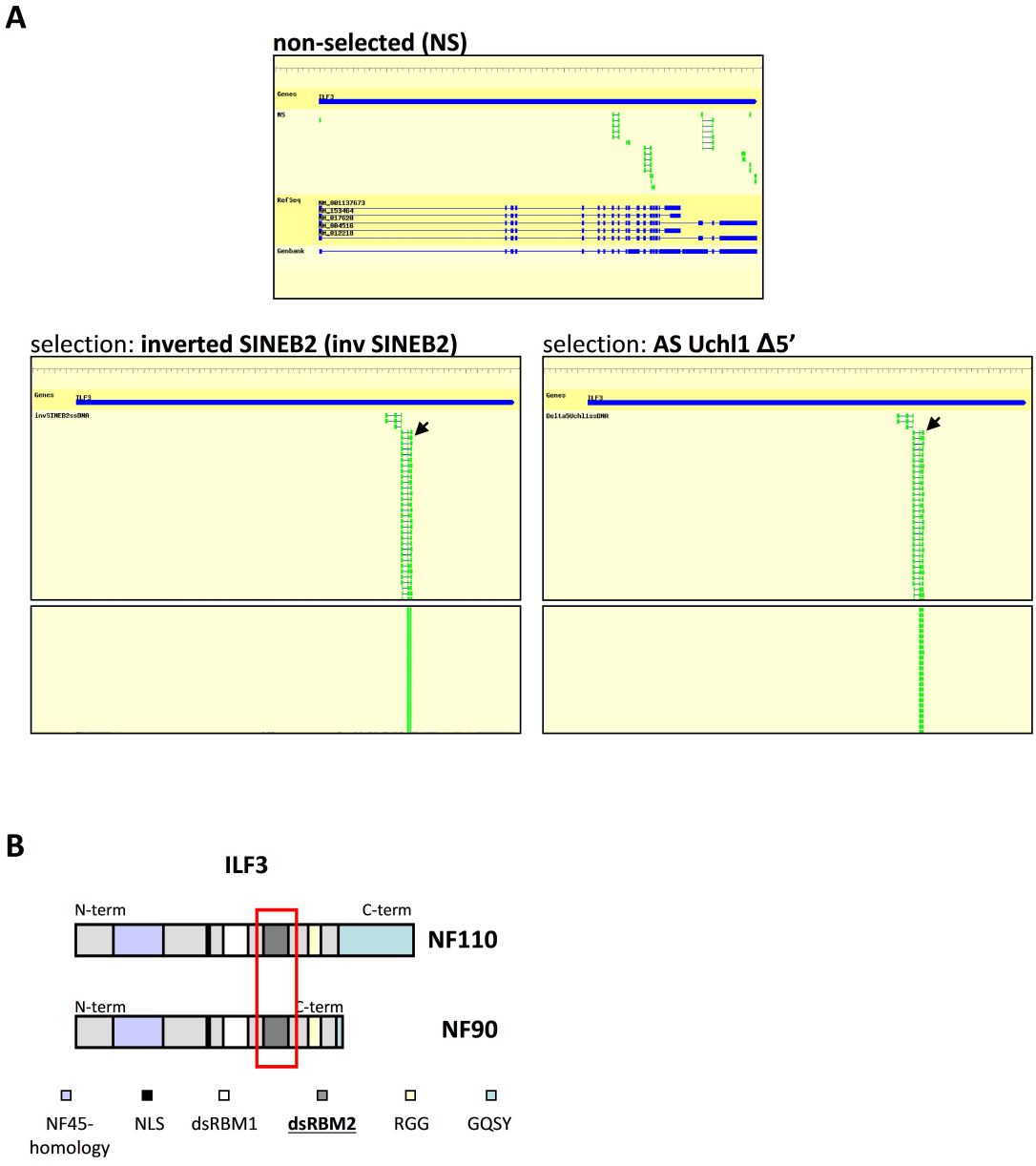


Figure 19. DsRBM2 is specifically enriched in SINEUP-driven selections. (**A**) Reads alignment to ILF3 gene. Data relative to AS Uchl1 Δ5'_ssDNA (middle) and inverted SINEB2_ssDNA (bottom) selections are shown. DsRBM2 (black arrows) is enriched in selected libraries, but not in the unselected (NS, top) library. Blue bars (top) indicate the gene; green bars (bottom) correspond to exons. (**B**) Schematic representation of ILF3 domains in the two main protein isoforms (NF110 and NF90): NF45-homology domain; nuclear localization signal (NLS); double-stranded RNA-binding motives (dsRBM) 1 and 2; RGG motif; GQSY domain. Abbreviations: NF=nuclear factor (*alias* for ILF3).

Hence, we kept on validating binding of ILF3 dsRBM2 with each of the baits employed for selection by phage ELISA (**Figure 20**).

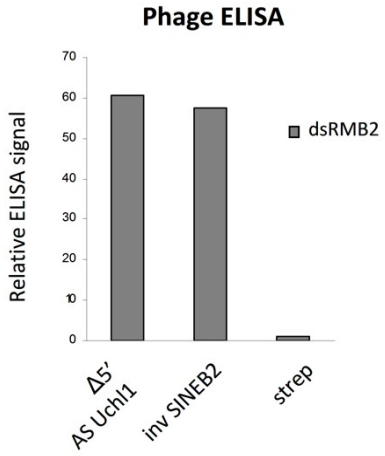


Figure 20. *In vitro* validation of SINEUP/ILF3 interaction by phage ELISA. DsRBM2 was screened by phage ELISA on SINEUP baits. Streptavidin served as control. Values are indicated as the fold signal *vs* background (uncoated wells).

Data showed that binding occurred with both AS Uchl1 Δ5' and inverted SINEB2 in a similar fashion, thus corroborating the specific nature of SINEUP/ILF3 dsRBM2 interaction *in vitro*. As further biochemical characterization, we were interested in comparing binding profiles of the two dsRBMs present in ILF3, differentially enriched during selections. Two representative ORFs encoding for ILF3 dsRBM1 and dsRBM2, respectively, were thus cloned as GST-fusion polypeptides and screened in ELISA (**Figure 21**).

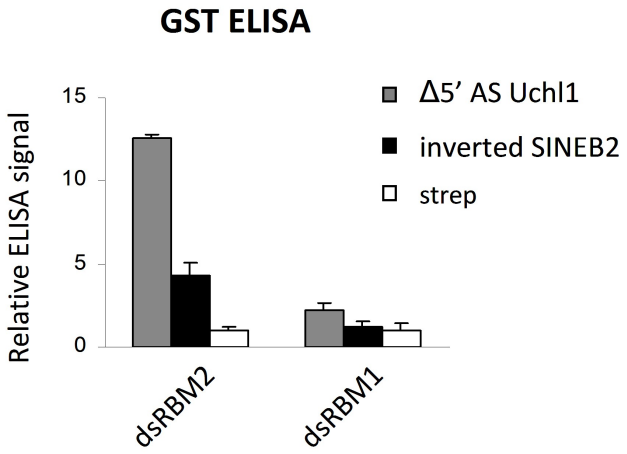


Figure 21. *In vitro* validation of SINEUP/ILF3 interaction by GST ELISA. Comparison between dsRBM1 and dsRBM2. Domains were subcloned into a compatible pGEX vector, purified as GST-fusion polypeptides and screened on SINEUP baits. Data indicate mean ± st dev. Data are representative of n=3 independent replicas.

Considering dsRBM2, results pointed out that binding occurred in a specific fashion with both baits. Notably, binding to AS Uchl1 Δ5'-embedded inverted SINEB2 was characterized by a higher signal/noise ratio. Conversely, dsRBM1-related ELISA signal was similar to the negative control with both baits.

Taken together, these results support a direct binding between the inverted SINEB2 and ILF3 dsRBM2, which provides the specific domain mediating the interaction with SINEUP baits *in vitro.* Of notice, we registered a remarkable increase in GST ELISA signal when dsRBM2 was exposed to AS Uchl1 Δ5'-embedded inverted SINEB2.

2.4 AS Uchl1 and ILF3 interact *in vivo* and binding requires the inverted SINEB2 repeat

RNA-IP preliminary setup

Once binding *in vitro* was validated, we focused on characterizing AS Uchl1/ILF3 interaction *in vivo* by RNA-IP. To overcome technical difficulties arising from low expression of endogenous AS Uchl1 in cells, we decided to overexpress this and co-immunoprecipitate it with endogenous ILF3.

Since endogenous AS Uchl1 is mainly localized in the nucleus (Carrieri et al., 2012), we investigated the subcellular distribution of transfected AS Uchl1 in three different

cell lines: mouse neuronal dopaminergic (MN9D), mouse neuroblastoma (N2A) and human embryonic kidney (HEK) 293T/17. After carrying out subcellular fractionation, AS RNA levels in nuclear and cytoplasmic compartments were quantified by qRT-PCR and expressed as relative percentages of total AS RNA. Nuclear and cytoplasmic fractions purity was controlled by monitoring levels of GAPDH and pre-rRNA, respectively. Comparison of distribution patterns in different cell lines confirmed a predominant nuclear localization of exogenous AS Uchl1 in all cell lines, with >60% of AS RNA detected in nuclei (**Figure 22**).

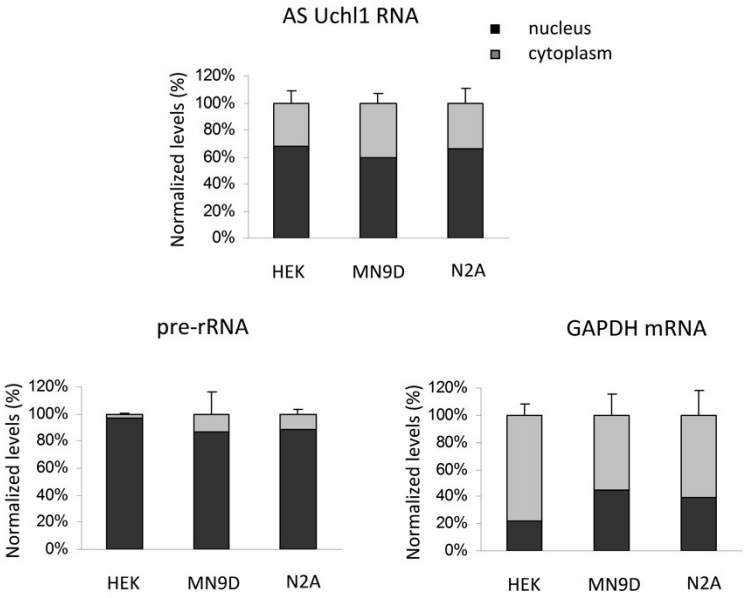


Figure 22. AS Uchl1 subcellular distribution in different cell lines. Subcellular distributions of overexpressed AS Uchl1 in HEK 293T/17, N2A and MN9D were compared. RNA quantification in cellular fractions was made by qRT-PCR. Nuclear (dark gray) and cytoplasmic (light gray) RNA levels were expressed as percentages of total RNA. Purity of fractions was assessed by evaluating levels of pre-rRNA and GAPDH. Data are representative of three independent experiments and indicate mean ± st dev.

In the same cells, we also investigated ILF3 localization by immunofluorescence (IF). Consistently with what found in literature, ILF3 was almost totally nuclear in all cells (**Figure 23A**). Of notice, in neuronal cells MN9D and N2A ILF3 staining was characterized by bright spots co-localizing with nucleoli. This has been shown previously, even if in a different line (Viranaicken et al., 2011), and is likely due to

the expression of a splicing variant predominantly found in nucleoli. In summary, these experiments confirmed the co-existence of overexpressed AS Uchl1 and endogenous ILF3 in the same cell compartment in different cell lines.

We then completed ILF3 characterization by monitoring protein expression levels. WB analysis showed that our antibody was specifically recognizing the two main isoforms of ILF3 (NF90 and NF110, at 90 kDa and 110 kDa, respectively, **Figure 23B**) in all cell lines.

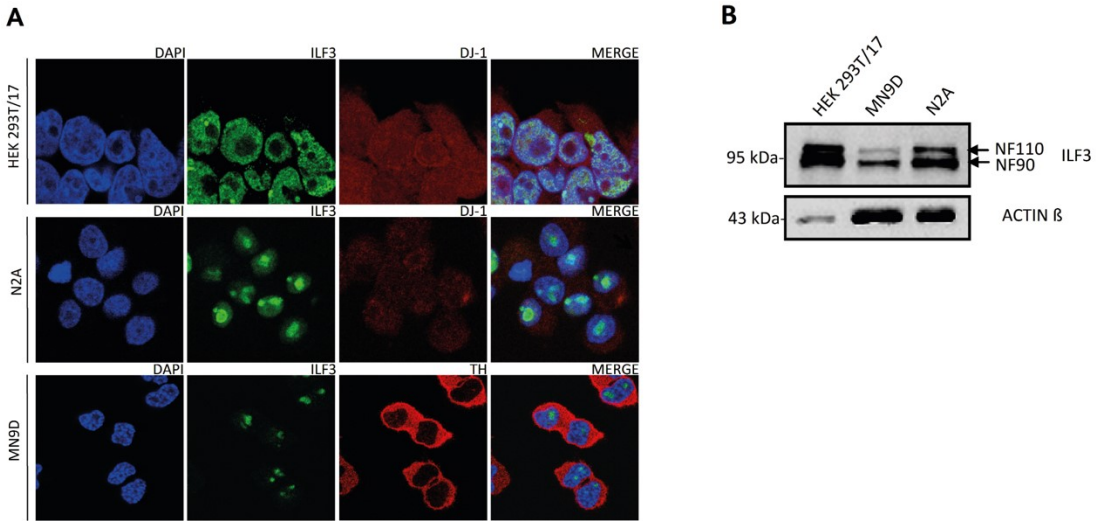


Figure 23. ILF3 in different cell lines. (**A**) IF on ILF3. Endogenous ILF3 was detected with anti-ILF3 antibody. Cytoplasmic counter-staining was carried out with anti DJ-1 antibody in HEK 293T/17 and N2A; in MN9D anti-TH antibody was used. Nuclei were stained with DAPI. (**B**) ILF3 protein levels in different cell lines. Endogenous ILF3 protein levels were monitored by WB in HEK 293T/17, MN9D and N2A. Immunoblotting was carried out with anti-ILF3 antibody. Actin-β was used as loading control.

In particular, expression levels in HEK 293T/17 were considerably higher than in neuronal cells.

Therefore, given their higher transfectability compared to neuronal lines and the higher expression of ILF3, we chose to carry out RNA-IP experiments in HEK 293T/17.

RNA-IP

Following AS Uchl1 overexpression and cross-linking of RNA-protein complexes, endogenous ILF3 was immunoprecipitated with specific antibodies or control IgG.

As we were interested in addressing the contribution of SINEUP ED in protein binding *in vivo*, an AS Uchl1 mutant lacking the inverted SINEB2, referred to as AS Uchl1 ΔSINEB2 was also employed (described in Carrieri et al., 2012 and schematically represented in **Figure 24A**). Parallel RNA-IPs were thus performed with WT and ΔSINEB2 AS Uchl1, and enrichment patterns compared. The presence of target RNA in ILF3 IP fractions or control IgG was quantified by qRT-PCR and normalized to the mRNA level of the housekeeping gene Ubiquitin C (UBC), previously described as non-interacting with ILF3 (Kuwano et al., 2010). IP efficiency was monitored by western blot. As shown in figure **Figure 24B**, AS Uchl1 was specifically enriched in ILF3 immunoprecipitates, confirming that AS Uchl1 and ILF3 interact *in vivo*. Of notice, WB analysis showed that both HEK endogenously expressed ILF3 isoforms, NF110 and NF90, were successfully immunoprecipitated (**Figure 24C**).

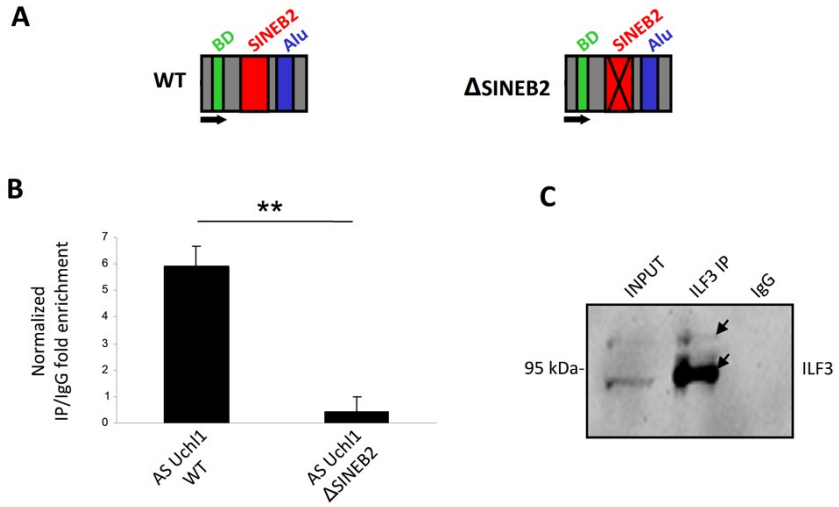


Figure 24. Validation of AS Uchl1/ILF3 interaction *in vivo*. (**A**) Schematic representation of AS Uchl1 WT (left) and AS Uchl1 ΔSINEB2 (right). (**B**) Endogenous ILF3 co-IP with overexpressed AS Uchl1 WT or AS Uchl1 ΔSINEB2 was carried out in HEK 293T/17. As control of IP specificity, IgG were immunoprecipitated in parallel. WT or mutated AS Uchl1 enrichment in ILF3 IP or IgG was quantified by qRT-PCR (left) and expressed as (2^ ΔCT)*100 ILF3 IP/ (2^ ΔCT)*100 IgG. ΔCT was calculated on INPUT. RNA content in IgG/IP was normalized on UBC mRNA. (**C**) ILF3 IP efficiency was monitored by WB. Immunoblotting was performed with anti-ILF3 antibody, recognizing both 90 KDa and 110 KDa ILF3 isoforms. Data are representative of three independent experiments and indicate mean ± st dev. Differences with $p<0.05$ were considered significant.

Interestingly, removal of the inverted SINEB2 resulted in a remarkable decrease of AS RNA detected in ILF3 IP fraction (**Figure 24B**), suggesting a role of the embedded repeat in establishing contacts with this RBP. Taken together, these results support an interaction between AS Uchl1 and ILF3 in vivo and a crucial role of the embedded inverted SINEB2 repeat as interaction interface with the nuclear protein partner ILF3.

2.5 The inverted SINEB2 and Alu direct localization of AS Uchl1 to ILF3-containing nuclear complexes

So far, we confirmed AS Uchl1/ILF3 binding *in vitro* and *in vivo*, pointing out the pivotal role of the ED in mediating contacts between transcript and protein *in vivo*. As a natural prosecution of such findings, we investigated the biological function of this interaction in cells. Transcripts with embedded TEs have been reported to be often retained in nuclei. Being privileged sites for editing (Athanasiadis et al., 2004; Kim et al., 2004; Levanon et al., 2004), repetitive elements and, in particular, SINEs often provide interfaces for association to nuclear protein complexes, which subsequently control their export and cytoplasmic availability (Chillón and Pyle, 2016; Zhang and Carmichael, 2001). On the other hand, nuclear ILF3 has been shown to actively contribute to regulation of basic metabolic properties of RNAs, including subcellular localization and stability (Castella et al., 2015). In this context, we investigated the role of ILF3 in trapping SINEUPs in the nuclear compartment, via binding the inverted SINEB2. To address the involvement of ILF3 in modulating SINEUPs localization, we overexpressed AS Uchl1 in ILF3-silenced HEK 293T/17 (**Figure 25A**) and checked its subcellular distribution by nucleocytoplasmic fractionation. Interestingly, data showed a 10-20% increase of AS Uchl1 WT in the cytoplasmic fraction of knocked-down cells (**Figure 25B**).

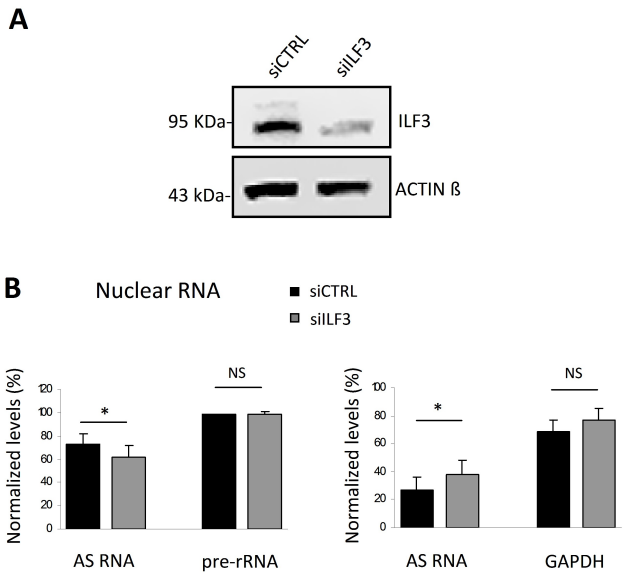


Figure 25. AS Uchl1 subcellular localization upon ILF3 silencing. (**A**) ILF3 was silenced in HEK 293T/17 and (**B**) overexpressed WT AS Uchl1 subcellular localization was evaluated in knocked-down cells (siILF3, gray bars) by nucleocytoplasmic fractionation. Control cells received WT AS Uchl1 and control siRNA (siCTRL, black bars). RNA quantification in cellular fractions was made by qRT-PCR. Nuclear (left) and cytoplasmic (right) RNA levels were expressed as percentages of total RNA. Purity of fractions was assessed by evaluating levels of pre-rRNA and GAPDH. Data are representative of four independent experiments and indicate mean ± st dev. Differences with $p<0.05$ were considered significant.

As a complementary approach, we verified whether the inverted SINEB2 excision had any impact on AS Uchl1 localization. Therefore, we carried out cell fractionation from HEK 293T/17 transfected with AS Uchl1 WT or with its mutant deprived of the SINEB2. We observed that AS Uchl1 distribution was partially perturbed in response to removal of the inverted SINEB2, again with a 20% increase of AS RNA detected in the cytoplasmic compartment, compared to the WT variant. This was strikingly reminiscent of what observed following ILF3 silencing (**Figure 26**).

Taken together, these data strengthen a role of ILF3 in AS Uchl1 nuclear entrapment through its interaction with the inverted SINEB2 element.

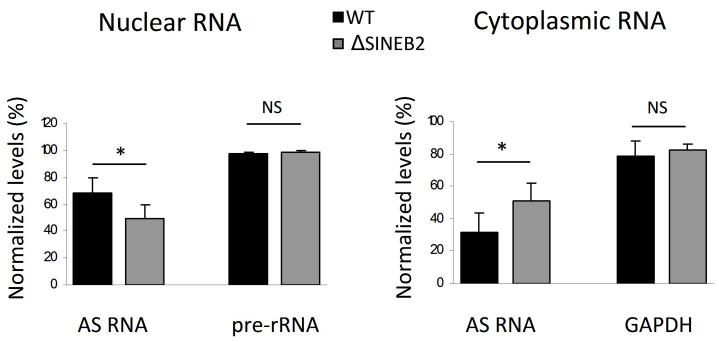


Figure 26. Subcellular localization of AS Uchl1 WT and its ΔSINEB2 mutant. Subcellular distributions of AS Uchl1 WT (black bars) and ΔSINEB2 (gray bars) were compared. RNA quantification in cellular fractions was made by qRT-PCR. Nuclear (left) and cytoplasmic (right) RNA levels were expressed as percentages of total RNA. Purity of fractions was assessed by evaluating levels of pre-rRNA and GAPDH. Data are representative of four independent experiments and indicate mean ± st dev. Differences with $p<0.05$ were considered significant.

However, the deletion of the inverted SINEB2 proved not to be sufficient for a complete relocalization of AS Uchl1 in the cytoplasm. In addition to the SINEB2, another repetitive element, that is a partial Alu sequence, was previously unmasked in the AS Uchl1 third exon (Carrieri et al., 2012). Though not necessary for translational activity, we decided it was interesting to investigate its potential role in AS Uchl1 subcellular localization. We thus assessed the effects of combined removal of the SINEB2 and the Alu repeat by taking advantage of AS Uchl1 Δ Transposable Elements (AS Uchl1 ΔTE) mutant (**Figure 27A**, right). Results showed a dramatic inversion of AS RNA distribution within the cell with 60-70% localizing in the cytoplasmic fraction (**Figure 27B**).

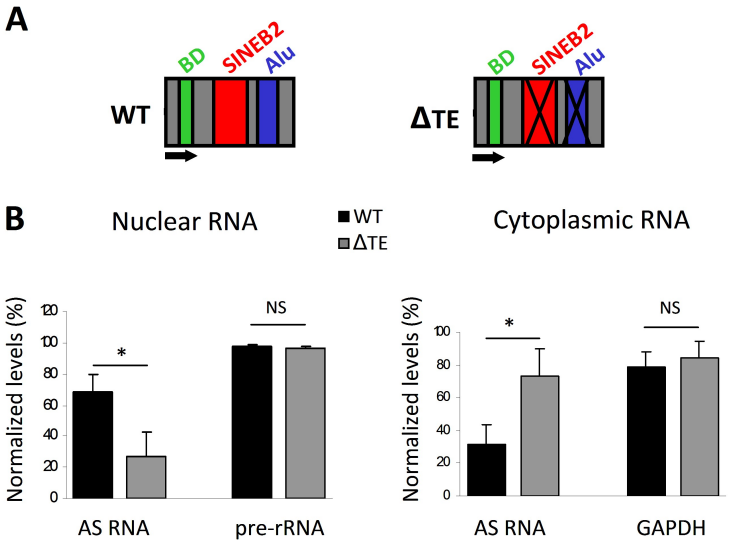


Figure 27. Subcellular localization of AS Uchl1 ΔTE mutant. (**A**) Schematic representation of AS Uchl1 WT and AS Uchl1 ΔTE mutant. ΔTE results from combined removal of embedded inverted SINEB2 (red) and partial Alu (blue). (**B**) Subcellular distributions of WT AS Uchl1 (black bars) and ΔTE (gray bars) were compared. RNA quantification in cellular fractions was made by qRT-PCR. Nuclear (left) and cytoplasmic (right) RNA levels were expressed as percentages of total RNA. Purity of fractions was assessed by evaluating levels of pre-rRNA and GAPDH. Data are representative of four independent experiments and indicate mean ± st dev. Differences with $p<0.05$ were considered significant.

To summarize, these results suggest that embedded repeats, SINEB2 and Alu, may act in a synergic fashion to produce the motif responsible for AS Uchl1 nuclear retention.

2.6 ILF3 interacts with paraspeckle protein p54^{nrb} *in vivo*

We then asked whether ILF3 could be part of a major nuclear complex in charge of regulating AS Uchl1 availability in the cytoplasm. Therefore we focused our attention on paraspeckles. These are subnuclear bodies localizing in the interchromatin space in the proximity of nuclear speckles (Fox and Lamond, 2010). Paraspeckles are critical to the control of gene expression through the nuclear retention of transcripts containing dsRNA regions that have been subject to A-to-I editing.

We assessed the interaction between ILF3 with p54^{nrb}, that is one of the main protein components of paraspeckles. To this purpose, we carried out an assay of co-IP of endogenous ILF3 and p54^{nrb} in HEK 293T/17. Following protein extraction, lysates were analyzed by WB. As shown in **Figure 28**, p54^{nrb} was enriched in ILF3 immunoprecipitates, suggesting that the two proteins are likely to be found in the same complex *in vivo.*

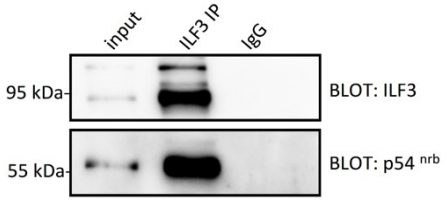


Figure 28. ILF3 interacts with paraspeckle protein p54^{nrb} *in vivo.* Co-IP of ILF3 and p54^{nrb}. Endogenous ILF3 was immunoprecipitated in HEK 293T/17 cells. ILF3 IP efficiency was revealed by immunoblotting with anti-ILF3 antibody. The presence of p54^{nrb} in immunoprecipitates was assessed by probing the membrane with anti-p54^{nrb} antibody.

3 Validation of SINEUPs as RNA tools to increase protein synthesis

Attachment: "SINEUPs are modular AS long non-coding RNAs that increase synthesis of target proteins in cells"

Similarly to proteins, AS Uchl1 and natural SINEUPs display discrete domains that in combination determine their function. The BD and the ED provide "modules" conferring specificity and biological activity, respectively. As previously shown with synthetic SINEUP AS GFP, the domain organization of AS Uchl1 can be exploited to engineer artificial RNAs enhancing translation of selected mRNAs (Carrieri et al., 2012). These observations pave the way for the application of SINEUPs as tools to increase synthesis of target proteins in a number of different contexts.

In biopharmaceutical industry, SINEUPs could be employed in systems for large-scale production of recombinant proteins, with the clear advantage of increased purity of the target protein over contaminants.

The use of SINEUPs as molecular biology reagents would sound promising as well. Indeed, from a certain perspective, SINEUPs can be considered as the opposite counterpart of siRNAs. As siRNAs have become a useful tool in the hands of experimental biologists when inhibition of gene expression is required to address specific biological questions, there are as many instances where scientists are interested in perturbing biological systems by increasing the amount of a specific endogenous protein, or a whole set of proteins, within physiological ranges. Eventually, SINEUP molecules may represent ideal RNA therapeutic tools for increasing gene expression *in vivo.* In particular, diseases arising from haploinsufficiency would strongly benefit from the discovery of RNAs that can increase protein levels of genes for which low expression is pathogenic.

So far, the potential scalability of AS Uchl1-derived synthetic lncRNAs to a platform of mRNA-specific translation enhancers remains to be addressed.

In the following attached paper, we validated and optimized SINEUPs as molecular tools to increase translation of selected targets.

Validation came along different steps. First, we demonstrated the efficacy and reproducibility of synthetic SINEUPs in different cell lines. As the modular structure of SINEUPs theoretically allows the design and generation of synthetic SINEUPs against desired targets by swapping the BD with an appropriate sequence, we then

tested the flexibility of BD design. To this purpose, we created a SINEUP targeting the commonly used FLAG tag sequence and showed that it successfully managed to up-regulate synthesis of different N-terminally FLAG-tagged proteins in cells, supporting interesting applications of SINEUPs in protein manufacturing. Next, we provided evidence about the activity of SINEUPs domains as independent units presenting their own structure and function. Indeed, SINEUP activity is retained in miniSINEUPs, which exclusively contain the BD and ED, isolated from the rest of AS Uchl1 sequence. MiniSINEUPs represent a miniaturized version of SINEUPs, within the range of small RNAs length, thus facilitating delivery *in vivo*. Finally, we explored the potential application of synthetic SINEUPs to modulate endogenous genes expression. In this context, we successfully design SINEUPs targeting endogenous Parkinson's disease-associated DJ-1 and validated activity in different neuronal cell lines.

In summary, here we propose SINEUPs as scalable tools to increase synthesis of chosen proteins, with important applications in molecular biology experiments, protein manufacturing as well as in RNA-based molecular therapy.

frontiers in Cellular Neuroscience

ORIGINAL RESEARCH
published: 13 May 2015
doi: 10.3389/fncel.2015.00174

SINEUPs are modular antisense long non-coding RNAs that increase synthesis of target proteins in cells

Silvia Zucchelli[1,2†], *Francesca Fasolo*[1†], *Roberta Russo*[1], *Laura Cimatti*[1], *Laura Patrucco*[2], *Hazuki Takahashi*[3], *Michael H. Jones*[4], *Claudio Santoro*[2], *Daniele Sblattero*[2], *Diego Cotella*[2], *Francesca Persichetti*[2], *Piero Carninci*[3] *and Stefano Gustincich*[1*]

[1] Scuola Internazionale Superiore di Studi Avanzati, Area of Neuroscience, Trieste, Italy, [2] Dipartimento di Scienze della Salute, Universita' del Piemonte Orientale, Novara, Italy, [3] Division of Genomic Technologies, RIKEN Center for Life Science Technologies, Yokohama, Japan, [4] Cell Guidance Systems, Cambridge, UK

OPEN ACCESS

Edited by:
Rosanna Parlato, Ulm University, Germany

Reviewed by:
Mohammadreza Hajjari, Shahid Chamran University of Ahvaz, Iran
Rohit Mathur, M.D. Anderson Cancer Center, USA

***Correspondence:**
Stefano Gustincich, Scuola Internazionale Superiore di Studi Avanzati, Sector of Neurobiology, Via Bonomea 265, 34136 Trieste, Italy
gustinci@sissa.it

†*These authors have contributed equally to this work.*

Received: *10 December 2014*
Accepted: *20 April 2015*
Published: *13 May 2015*

Citation:
Zucchelli S, Fasolo F, Russo R, Cimatti L, Patrucco L, Takahashi H, Jones MH, Santoro C, Sblattero D, Cotella D, Persichetti F, Carninci P and Gustincich S (2015) SINEUPs are modular antisense long non-coding RNAs that increase synthesis of target proteins in cells. Front. Cell. Neurosci. 9:174. doi: 10.3389/fncel.2015.00174

Despite recent efforts in discovering novel long non-coding RNAs (lncRNAs) and unveiling their functions in a wide range of biological processes their applications as biotechnological or therapeutic tools are still at their infancy. We have recently shown that AS Uchl1, a natural lncRNA antisense to the Parkinson's disease-associated gene Ubiquitin carboxyl-terminal esterase L1 (Uchl1), is able to increase UchL1 protein synthesis at post-transcriptional level. Its activity requires two RNA elements: an embedded inverted SINEB2 sequence to increase translation and the overlapping region to target its sense mRNA. This functional organization is shared with several mouse lncRNAs antisense to protein coding genes. The potential use of AS Uchl1-derived lncRNAs as enhancers of target mRNA translation remains unexplored. Here we define AS Uchl1 as the representative member of a new functional class of natural and synthetic antisense lncRNAs that activate translation. We named this class of RNAs SINEUPs for their requirement of the inverted SINEB2 sequence to UP-regulate translation in a gene-specific manner. The overlapping region is indicated as the Binding Doman (BD) while the embedded inverted SINEB2 element is the Effector Domain (ED). By swapping BD, synthetic SINEUPs are designed targeting mRNAs of interest. SINEUPs function in an array of cell lines and can be efficiently directed toward N-terminally tagged proteins. Their biological activity is retained in a miniaturized version within the range of small RNAs length. Its modular structure was exploited to successfully design synthetic SINEUPs targeting endogenous Parkinson's disease-associated DJ-1 and proved to be active in different neuronal cell lines. In summary, SINEUPs represent the first scalable tool to increase synthesis of proteins of interest. We propose SINEUPs as reagents for molecular biology experiments, in protein manufacturing as well as in therapy of haploinsufficiencies.

Keywords: SINEUP, long non-coding RNA, antisense, protein expression, cell lines

Introduction

Large genomic projects such as ENCODE (Djebali et al., 2012) and FANTOM (Forrest et al., 2014) have shown that the majority of the mammalian genome is transcribed, thus generating a previously underestimated complexity in gene regulatory networks. Protein encoding genes present a large repertory of alternative Transcription Start Sites (TSSs) that may drive transcription in a cell type-specific manner (Valen et al., 2009). Different 5′UTRs may contain information for mRNA sorting to neuronal compartments as well as for stimulus-dependent translation. Furthermore, in addition to 25000 genes encoding for proteins, at least an equal number of long non-coding RNA (lncRNA) genes have been identified so far. These generate >200 base pairs long transcripts that do not encode for proteins. About one third of annotated lncRNAs overlaps with protein-coding genes (Derrien et al., 2012). Many of these are transcribed from the opposite strand forming sense/antisense (S/AS) pairs (Katayama et al., 2005; Derrien et al., 2012).

The nervous system appears as a privileged site for lncRNA expression, as the vast majority of these transcripts is brain-enriched and regulates neuronal development and functions (Qureshi and Mehler, 2012). Furthermore, a complex network of natural S/AS pairs may participate in brain development and homeostasis in physiological conditions. Interestingly, an increasing number of lncRNAs are associated with brain dysfunction and extensive AS transcription has been measured in *loci* associated to hereditary neurodegenerative diseases (Zucchelli et al., submitted).

Manipulating RNA expression *in vivo* has been proposed as new strategy for molecular therapy. Special attention has been devoted to small antisense oligonucleotides (ASOs) and siRNAs as tools to decrease gene expression of pathological target genes such as, for example, mutant huntingtin in Huntington's disease (Kordasiewicz et al., 2012; Yu et al., 2012).

An equally challenging large group of diseases would strongly benefit from the discovery of RNAs that can increase protein levels of genes for which low expression is pathogenic. Several hereditary intellectual and cognitive disabilities are haploinsufficiencies where only a single functional copy of a gene is unable to produce sufficient protein to maintain a physiological condition (Van Bokhoven, 2011). Therefore an RNA-based drug that can restore physiological amounts of the target protein can in principle be curative. Unfortunately, no molecules that can increase protein levels of a specific mRNA type *in vivo* have been found to date.

In search for AS transcripts that may regulate the expression of Parkinson's disease (PD)-associated genes, we recently discovered AS Uchl1, a natural lncRNA antisense to Ubiquitin carboxyl-terminal esterase L1 (Uchl1/PARK5) (Carrieri et al., 2012). AS Uchl1 is a nuclear-enriched transcript, that is expressed in dopaminergic neurons in the *Substantia Nigra*, the target of PD neurodegeneration, and is down-regulated upon PD-mimicking intoxication *in vitro* and *in vivo* (Carrieri et al., 2015). AS Uchl1 activity increases UchL1 protein synthesis at the post-transcriptional level. Upon stressful insults, AS Uchl1 shuttles from the nucleus to the cytoplasm, where it induces Uchl1 mRNA association to heavy polysomes to increase its translation (Carrieri et al., 2012). AS Uchl1 activity depends on two functional domains: the overlapping region that defines target specificity and the inverted SINE element of B2 subclass (invSINEB2) that confers protein synthesis activation. This functional organization is shared with other lncRNAs part of S/AS pairs in the mouse genome. Finally, by substituting the overlapping region with a sequence antisense to the Green Fluorescent Protein (GFP) mRNA, this synthetic RNA was able to increase GFP protein synthesis with a post-transcriptional mechanism (Carrieri et al., 2012).

The potential scalability of AS Uchl1-derived synthetic lncRNAs to a platform of mRNA-specific translation enhancers remained to be addressed.

Here we propose that AS Uchl1 is the representative member of a new functional class of natural and synthetic RNAs that increase protein synthesis. We name these RNAs as SINEUPs for their activity requires an invSINEB2 element (SINE) to UP-regulate translation of partially overlapping sense mRNAs. The overlapping region is indicated as the Binding Doman (BD) while the embedded inverted SINEB2 element is the Effector Domain (ED). By swapping BD, synthetic SINEUPs are designed targeting mRNAs of interest.

This work shows that synthetic SINEUPs are a versatile tool to increase synthesis of target proteins of interest paving the way for future applications of SINEUPs as molecular biology reagents and for manufacturing of recombinant proteins. Most importantly, SINEUPs may be directed to selected mRNA species *in vivo* representing a new type of RNA-based drug for molecular therapy.

Materials and Methods

Constructs

Plasmids expressing target proteins were previously described. In particular, for this study we used pEGFP-C2 (Carrieri et al., 2012), pcDNA3-2XFLAG-DJ-1 (Herrera et al., 2007; Zucchelli et al., 2009), pcDNA3-2XFLAG-TTRAP (Zucchelli et al., 2009; Vilotti et al., 2012), pcDNA3-2XFLAG-Hba (Biagioli et al., 2009) (Codrich et al., manuscript in preparation) and pcDNA3-2XFLAG-TRAF6 (Zucchelli et al., 2010, 2011).

Target specific SINEUPs were constructed using pcDNA3-Δ5′-ASUchl1 as backbone (Carrieri et al., 2012). SINEUP-backbone lacks the region of overlap (BD) to Uchl1 and retains AS Uchl1 ED with inverted SINEB2, Alu sequence and 3′ tail. SINEUP target-specific BDs were designed, in antisense orientation, around the ATG of protein-coding sequence with a −40/+32 anatomy.

SINEUP targeting EGFP (AS-GFP, here named SINEUP-GFP) has been described in Carrieri et al. (2012). SINEUP targeting FLAG-tagged proteins (SINEUP-FLAG) was cloned with the following primers (5′ to 3′ orientation):

Abbreviations: lncRNA, long non-coding RNA; AS, antisense; SINE, short interspersed nuclear element; invSINEB2, inverted SINE of B2 subfamily; PD, Parkinson's disease; TRAF6, tumor necrosis factor receptor associated factor 6; TTRAP/TDP2, TRAF and TNF receptor associated protein/tyrosyl-DNA phosphodiesterase 2; Hba, hemoglobin alpha chain.

FWD AS 2xFLAG:
ATATCTCGAGAATTCCTTGTCATCGTCGTCCTTGTAGTCCATCAATTCCAGCACACTGGCGGCCGT
REV AS 2xFLAG:
GAGAGATATCCTCGGATCCACTAGTAACGGCCGCCAGTGTGCTGGAATTGATGGACTACAAGGACG
Primers were annealed, elongated by PCR, digested and ligated into XhoI-EcoRV sites of SINEUP-backbone.

Short SINEUP targeting GFP (miniSINEUP-GFP) was generated combining BD of SINEUP-GFP and ED of AS Uchl1. Briefly, inverted SINEB2 was PCR amplified and cloned into EcoRI and HindIII sites of pcDNA3.1(-). SINEUP-GFP BD was subsequently added to the inverted SINEB2-containing plasmid at XhoI and EcoRI sites to obtain miniSINEUP-GFP. The following primers were used:
FWD EcoRI InvSINEB2: TATAGAATTCCAGTGCTAGAGGAGG
REV HindIII InvSINEB2: GAGAAAGCTTAAGAGACTGGAGC
FWD ApaI all O/L: TATAGGGCCCTCTAGACTCGAG
REV EcoRI O/L GFP20: GAGAGAATTCCAGCACAGTGGCGGCCGC
SINEUPs targeting DJ-1 were generated by annealing and PCR-based method (−40/+32) or with annealing and ligation of phosphorylated oligonucleotides (−40/+4), using the following primers:
SINEUP-DJ-1 (−40/+32) FWD:
ATATCTCGAGGCCAGGATGACCAGAGCTCTTTTGGAAGCCATTTTTATGTTATATGTTT
SINEUP-DJ-1 (−40/+32) REV:
GAGAGATATCTTTCAGCCTGGTGTGGGGCTTGTAAACATATAACATAAAAATGGCTT
SINEUP-DJ-1 (−40/+4) FWD:
TCGAGCCATTTTTATGTTATATGTTTACAAGCCCCACACCAGGCTGAAA
SINEUP-DJ-1 (−40/+4) REV:
TTTCAGCCTGGTGTGGGGCTTGTAAACATATAACATAAAAATGGC
All constructs were verified by sequencing.

Cell Lines and Transfection

HEK 293T/17 cells were obtained from ATCC (Cat. No. ATCC-CRL-11268 293T/17) and maintained in culture with Dulbecco's Modified Eagle Medium (GIBCO) supplemented with 10% FBS (SIGMA) and 1% antibiotics (penicillin/streptomycin), 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 and SK-N-SH were cultured in Eagle's minimal essential medium (SIGMA) supplemented with 10% FBS, 1% antibiotics, 1% GlutaMAX and 1% non-essential aminoacids. HeLa cells were grown with DMEM supplemented with 10% FBS and 1% antibiotics as previously described (Angelini et al., 2007). SH-SY5Y cells were maintained in culture as previously described (Zucchelli et al., 2009). BE(2)M17 were grown in 1:1 MEM-Glutamax (GIBCO)/F12 (GIBCO) supplemented with 10% FBS, 1% antibiotics, and 1% non-essential aminoacids.

When required, HEK 293T/17 cells were treated with rapamycin (SIGMA) at 1 μM for 1 h or at 100 nM for 16 h. DNA damage was induced by doxorubicin (SIGMA), at 1 μM for 1 h or 500 nM for 16 h.

HEK 293T/17 cells were transfected with Fugene HD (Roche) and Lipofectamine 2000 (Life Technologies), following manufacture's instruction. HepG2, HeLa, SH-SY5Y, BE(2)M17 and SK-N-SH cells were transfected with Lipofectamine 2000. All cells were transfected with a 1:6 ratio between sense and SINEUP encoding plasmids, maintaining the conditions described for S/AS Uchl1 (Carrieri et al., 2012). In detail, cells were plated in 6-well plates the day before transfection at 60% (for Fugene HD protocol) or 80–90% (for Lipofectamine protocol) confluency. For transfection with Fugene 0.3 μg pEGFP and 1.8 μg SINEUP plasmid were used; for Lipofectamine 2000 0.6 μg pEGFP and 3.4 μg SINEUP. Cells were collected at 24 h (HeLa cells) or 48 h (HEK 293T/17 and HepG2 cells) after transfection and split in two samples for RNA extraction and Western Blot analysis.

For SINEUPs targeting endogenous mRNAs, SINEUP-encoding plasmid was transfected at the highest dose (4 μg) following manufacture's instructions.

Data of RNA and protein levels were obtained from the same transfection in each replica.

Western Blot

For Western blot analysis, cell pellets were directly resuspended in Laemli sample buffer, briefly sonicated, boiled and loaded on poly-acrilamide gels.

Primary antibodies used in this study include anti-GFP rabbit polyclonal antibody (Life Techologies, Cat. No. A6445), used 1:1000, anti-FLAG M2 (SIGMA, Cat. No. 3165), 1:1000, anti-β-actin (SIGMA), 1:5000, and anti-TRAF6 (Abnova), used 1:500. To detect endogenous DJ-1 protein an antibody produced in our laboratory was used (Zucchelli et al., 2009; Foti et al., 2010). For the detection, anti-mouse-HRP or anti-rabbit-HRP (Dako) in combination with ECL (GE Healthcare) was used. Image detection was performed with Alliance LD2-77WL system (Uvitec, Cambridge). Image quantification was done using Adobe Photoshop CS5.

RNA Isolation, Reverse Transcription and Quantitative RT-PCR (qRT-PCR)

Total RNA was extracted from cells using RNeasy Mini Kit (QIAGEN) following manufacturer's instructions. RNA was treated with on-column DNase I (QIAGEN) followed by a second DNase I digestion in solution (Ambion). Two rounds of DNase digestion were required to avoid plasmid DNA contamination in this experimental setting. Single strand cDNA was prepared from 1 μg of purified RNA using the iSCRIPT™ cDNA Synthesis Kit (Bio-Rad) according to manufacturer's instructions. qRT-PCR reaction was performed on diluted cDNA (1:20) using SYBR-Green PCR Master Mix (Applied Biosystem) and an iCycler IQ Real time PCR System (Bio-Rad). Relative expression was calculated with the ΔΔCt method (Schmittgen and Livak, 2008). Oligonucleotide sequences of primers used in this study for GFP and GAPDH (Carrieri et al., 2012), DJ-1 (Foti et al., 2010),

TTRAP (Zucchelli et al., 2009), TRAF6 (Zucchelli et al., 2010) and Hba (Biagioli et al., 2009) were previously described.

SINEUP-GFP, SINEUP-FLAG and SINEUP-DJ-1 were detected with primers designed on the 3′ end of AS-Uchl1 (mAS Uchl1 FWD and REV, Primers 3′) (Carrieri et al., 2012). MiniSINEUP-GFP was quantified using the following primers: pTSinvB2 FWD-RT (CAGTGCTAGAGGAGGTCAGAAGA) and pTSinvB2 REV-RT (GGAGCTAAAGAGATGGCTCAG CACTT).

Cellular Fractionation

For fractionation experiments, GFP/SINEUP-GFP were transfected in 10 cm plates at 1:6 ratio using Lipofectamine 2000. Nucleo cytoplasmic fractionation was performed as previously described (Wang et al., 2006). Nucleus and cytoplasmic RNAs were extracted using Trizol reagent (Invitrogen) following manufacturer's instruction. RNA was eluted and treated with DNase I. The purity of the cytoplasmic fractions was confirmed by qRT-PCR on pre-ribosomal RNA using the following primers (Murayama et al., 2008):

FWD 5′-GAACGGTGGTGTGTCGTTC-3′
REV 5′-GCGTCTCGTCTCGTCTCACT-3′

Statistical Analysis

All data are expressed as mean ± standard deviation on $n \geq 3$ replicas. Statistical analysis was performed using Excel software. Statistically significant differences were assessed by Student's *t*-test. Differences with $p < 0.05$ were considered significant.

Results

SINEUPs: Definition and Design

As shown in Carrieri et al. (2012), AS Uchl1 stimulates translation of partially overlapping sense protein-coding mRNAs with no effects on RNA levels.

Here we propose AS Uchl1 as the representative member of a new functional class of natural and synthetic antisense lncRNAs that activate translation. We named these lncRNAs as SINEUPs since they take advantage of an embedded invSINEB2 element to UP-regulate translation. Therefore SINEUPs can be considered the first example of gene-specific inducers of protein synthesis.

SINEUPs display a modular architecture (**Figure 1A**). In the 5′ region, SINEUPs contain the sequence that overlaps, in antisense orientation, to the sense protein-coding mRNA. We named this sequence as SINEUP's Binding Domain (BD) since it provides target selection and SINEUP specificity by RNA-RNA base pairing. In AS Uchl1, BD is 73 base pair long, centered across the ATG with a −40/+32 configuration, spanning part of Uchl1 5′UTR and a portion of its CDS (Carrieri et al., 2012). The remaining SINEUP sequence presents the embedded invSINEB2 element in the non-overlapping part of the transcript. Since this region has been proven essential for protein synthesis up-regulation, we defined it as Effector Domain (ED) (**Figure 1A**).

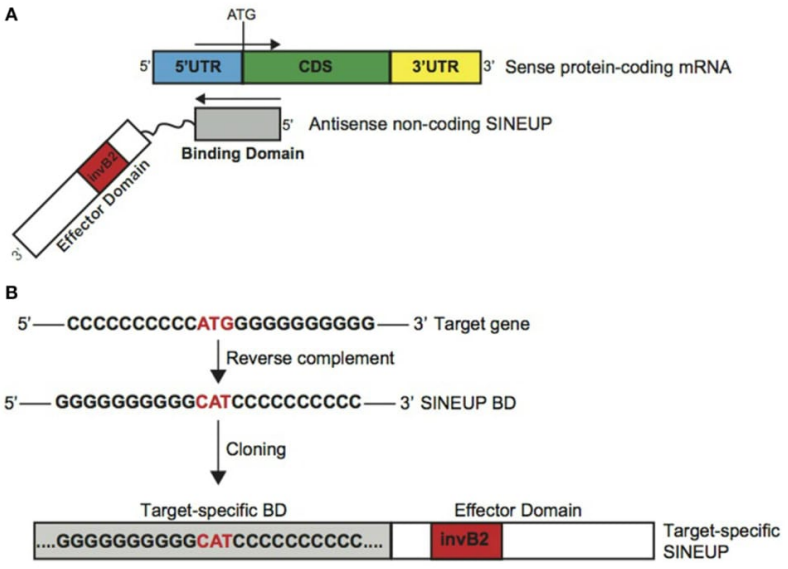


FIGURE 1 | Schematic representation of SINEUPs. (A) SINEUP modular structure. SINEUP binding domain (gray): SINEUP sequence that overlaps, in antisense orientation, to the sense protein-coding mRNA. SINEUP effector domain (red): non-overlapping portion of SINEUPs (white), containing the inverted SINEB2 element (invB2) that confers activation of protein synthesis. 5′ to 3′ orientation of sense and antisense RNA molecules is indicated. Structural elements of protein-coding mRNA are shown: 5′ untranslated region (5′UTR, blue), coding sequence (CDS, green) and 3′ untranslated region (3′UTR, yellow). **(B)** SINEUP design strategy. Schematic representation of the cloning strategy to generate target-specific SINEUPs. An artificial target gene sequence is indicated as example. SINEUP domains are represented as in **A**.

In Carrieri et al., we showed that this modular structure is shared with several natural lncRNAs component of S/AS pair in the mouse genome. For AS Uxt we proved it is able to increase UXT protein synthesis with no effects on Uxt mRNA levels. Therefore these define the first entries in the list of natural SINEUPs.

Given their modular structure, target-specific synthetic SINEUPs can be designed at will by manipulating AS Uchl1 sequence (**Figure 1B**). In Carrieri et al., we have designed the first synthetic SINEUP directing its activity to GFP mRNA by swapping BDs. However, any invSINEB2 sequence from other natural SINEUPs can in principle sustain activity and be considered potential ED.

Provided with the exact TSS for the target gene, sequence of interest is extrapolated centered across the initiating ATG. After reverse-complement manipulation, gene-specific BD is generated by annealing and PCR amplification of specific oligonucleotides. Target-specific SINEUP is then obtained by cloning specific BDs upstream to the SINEUP effector domain. For expression in mammalian cells, SINEUPs are cloned into pcDNA3.1 plasmid. Vectors for retroviral and lentiviral packaging can also be efficiently used (data not shown).

SINEUPs Work *in vitro* in Different Cell Lines

Mammalian cell cultures *in vitro* are routinely used as model systems to study the molecular mechanisms of gene functions as well as cell factories to produce therapeutic proteins. Considering the flexibility of SINEUP technology and its potential applications in molecular biology experiments, protein manufacturing and therapeutics, we investigated the efficacy and reproducibility of SINEUPs in different cell lines *in vitro*. To this purpose we selected hepatocellular carcinoma HepG2 cells, for their use as a model system by large multicenter consortia, epithelial carcinoma HeLa cells for their wide use in cell biology and in therapeutic protein production as well as HEK 293T/17 cells as positive control of SINEUP activity. As a representative synthetic SINEUP, we took advantage of SINEUP-GFP to increase GFP protein levels in transient overexpression experiments.

We estimated SINEUP activity as fold changes in protein levels encoded by targeted mRNAs in the presence/absence of SINEUP with mRNA amounts kept constant ($p > 0.05$).

Similar fold-changes were observed for SINEUP activity with GFP target at 24 and 48 h after transfection (data not shown). However timing for optimal activity was cell-line dependent, as best conditions were found at 24 h in HeLa and 48 h in HEK 293T/17 and HepG2 cells (**Figures 2A–C**).

In HEK 293T/17 cells we obtained an average 2.4 fold change (**Figure 2A**), confirming previously published data (Carrieri et al., 2012) on an independent batch of cells and on a larger cohort of transfections. SINEUP activity ranged from a minimum of 60% induction to a maximum of 400% (**Figure 2D**). SINEUP effect in transfected HEK 293T/17 cells was not enhanced upon stressful stimuli such as rapamycin and DNA-damage inducing drug doxorubicin (**Supplementary Figure 1**). HepG2 cells and HeLa cells proved to support SINEUP activity (**Figures 2B,C**), with an average induction of 1.65 and 1.82-fold, respectively. Minimal values were 20% in HepG2 and 40% in HeLa cells, and top effect was 250 and 220% (**Figure 2D**). No statistical differences could be measured in SINEUP activity between the three cell lines ($p > 0.05$), albeit HEK 293T/17 cells tended to be more effective (**Figure 2D**).

We observed that SINEUP activity is maintained independently of the reagent used for transfection. In HEK 293T/17 cells an average of 2.4 fold change could be measured with Lipofectamine ($n = 5$ experiments) and 2.3 with Fugene HD ($n = 6$ experiments) (data not shown). Under these experimental conditions, RNA from transfected SINEUP-GFP was detected in the cytoplasmic fraction although a prominent accumulation in the nucleus was evident (**Supplementary Figure 2**). Interestingly, sense GFP mRNA was equally distributed between cytoplasmic and nuclear compartments. Altogether these data indicate that SINEUPs can be used *in vitro* in different cell systems to up-regulate proteins of interest.

SINEUPs can be Designed to Increase Production of Target Proteins of Interest

The modular structure of SINEUPs predicts that by swapping the BD with an appropriate sequence it should be possible to redirect SINEUP activity to target mRNA of interest.

To test the flexibility of BD design, we generated a SINEUP molecule targeting the commonly used FLAG tag sequence. FLAG-specific SINEUP would be able to act at a post-transcriptional level increasing the quantities of proteins expressed in frame with an N-terminal FLAG tag. SINEUP targeting FLAG-tagged proteins (SINEUP-FLAG) was designed to mimic the molecular anatomy of SINEUP-GFP. In particular, SINEUP-FLAG BD encompasses −40 nucleotides (in pcDNA3 plasmid backbone) before FLAG-initiating Met and +32 bases covering the first FLAG tag sequence (**Figure 3A**).

We took advantage of a series of protein-coding genes available in the laboratory to test SINEUP-FLAG activity. We used human TRAF6 (RefSeq NM_004620) (Zucchelli et al., 2010, 2011), human TTRAP (RefSeq NM_016614) (Zucchelli et al., 2009; Vilotti et al., 2012), human DJ-1 (RefSeq NM_001123377) (Herrera et al., 2007; Zucchelli et al., 2009) and mouse Hba-a1 (RefSeq NM_008218) (Biagioli et al., 2009) (Codrich et al., manuscript in preparation) cloned in pcDNA3-2XFLAG. HEK 293T/17 cells were transfected with plasmids for FLAG-tagged targets in combination with SINEUP-FLAG (+SINEUP). Cells transfected with an empty vector were used as control (-SINEUP). SINEUP activity was measured quantifying protein levels by Western blot and RNA amounts by qRT-PCR. We found that the quantity of three of four FLAG-tagged proteins that we tested was modulated by co-expression of SINEUP-FLAG (**Figure 3B**). SINEUP effect was different toward the three targets, ranging from 1.5 to 3.0 fold changes. FLAG-TRAF6 showed the strongest SINEUP-mediated activation with a protein induction consistently in the range of 2.6 to 3.0 fold (**Figure 3B** and data not shown). No effect on TRAF6 mRNA was present, as expected ($p = 0.46$). Interestingly, increased TRAF6 levels could

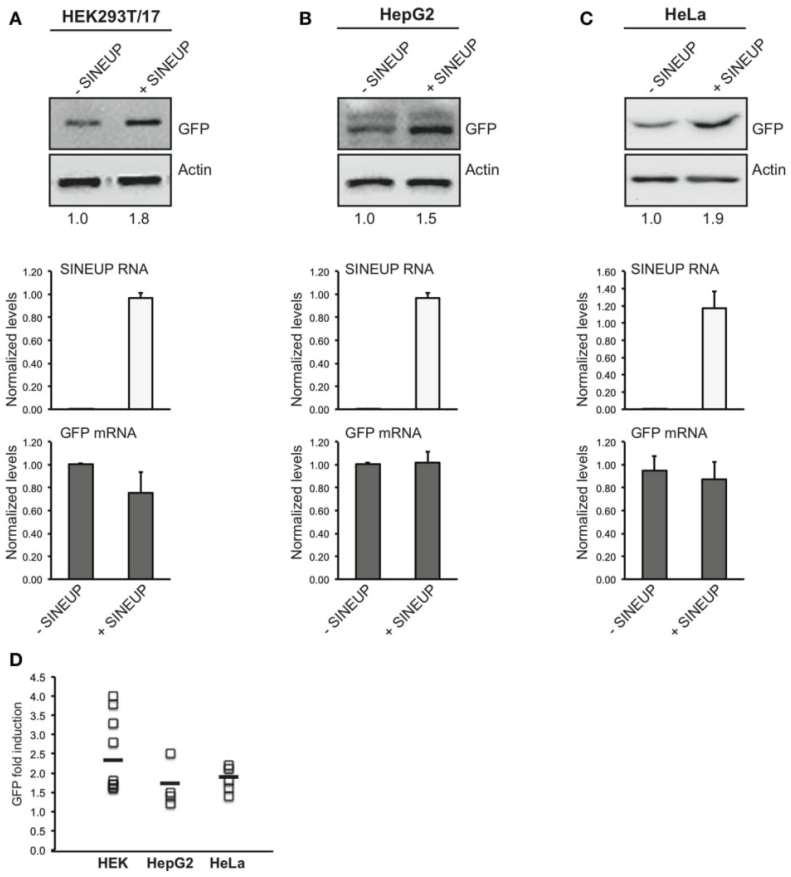


FIGURE 2 | SINEUP activity in cell lines. HEK 293T/17 **(A)**, HepG2 **(B)** and HeLa **(C)** cells were transfected with pEGFP-C2 and SINEUP-GFP at 1:6 ratio (+SINEUP). Control cells were transfected with pEGFP-C2 and an empty control plasmid (-SINEUP). 24 h (HeLa) or 48 h (HEK 293T/17 and HepG2) after transfection, cells were lysed and processed for protein (top) and RNA (bottom) levels. Western blot was performed with anti-GFP antibody. β-actin was used as loading control. Fold-induction was calculated on Western blot images normalized to β-actin and relative to empty control samples. Expression of SINEUP-GFP (white bars) and quantity of GFP mRNA (gray bars) were monitored by qRT-PCR using specific primers. Data indicate mean ± standard deviation. Data are representative of >3 independent replicas. **(D)** Graphical representation of SINEUP-mediated GFP up-regulation in HEK ($n = 10$), HepG2 ($n = 4$) and HeLa ($n = 5$) cells. No statistical difference was present between the three cell lines ($p > 0.05$).

be measured when probing lysates with anti-TRAF6 specific antibody (**Supplementary Figure 3**).

A 50% up-regulation was measured with FLAG-DJ-1 and 90% with FLAG-Hba. In both cases, sense mRNA in -SINEUP and +SINEUP transfections was not statistically different ($p = 0.36$ in FLAG-DJ-1 samples and $p = 0.38$ in FLAG-Hba).

A modest activation was observed with FLAG-TTRAP where a 10–20% increase was typically observed by Western blot. However, this was consistently accompanied by a similar modulation of TTRAP mRNA levels (**Supplementary Figure 4**), thus excluding this effect from SINEUP definition.

In summary synthetic SINEUPs can be designed to commonly used tag sequences and the same SINEUP can dictate translation of different tagged proteins of interest.

MiniSINEUPs Containing Exclusively BD and ED are Active

A major limitation in the use of naked RNA for *in vitro* and *in vivo* applications is the instability and low cellular permeability of long molecules. Chemical modifications can bypass such limits, but with specific constraints in RNA length. Synthetic SINEUPs derived from natural AS Uchl1 are about 1200 nucleotides (nt) long with BD of 72 base pairs, ED of 170 base pairs in addition to intervening sequences, a partial Alu element (73 base pairs) and a 3′ tail. This length is suitable for delivery systems such as viral vectors, but incompatible with the use of SINEUPs as naked RNA therapeutic molecules. Therefore, we aimed at synthesizing the shortest functional SINEUP that retains its translation enhancement activity. MiniSINEUP-GFP

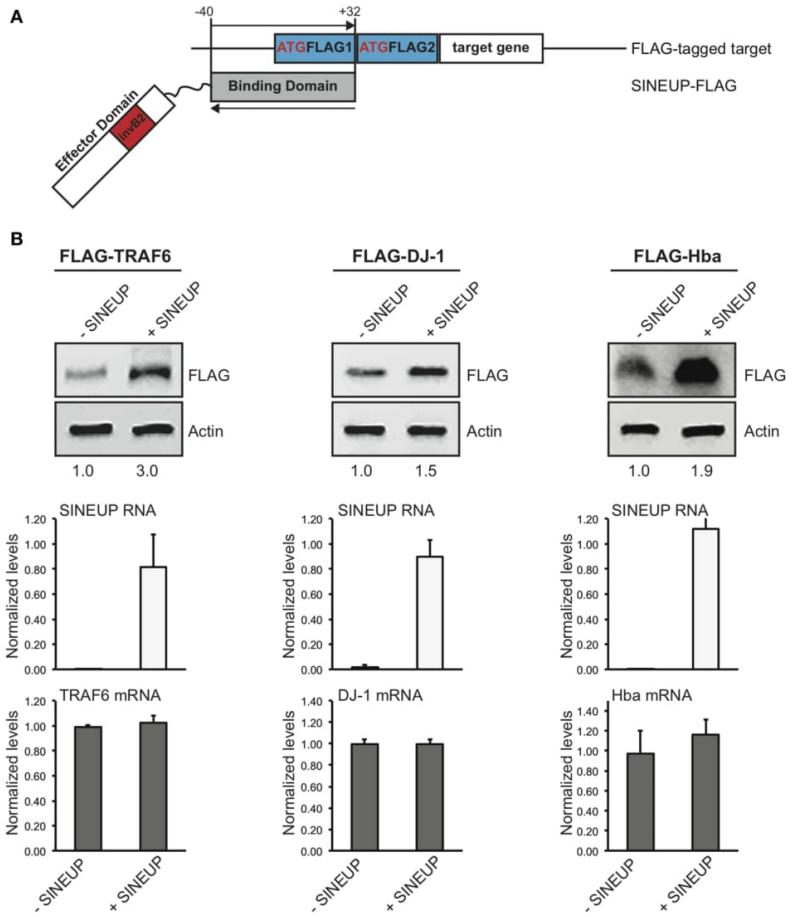


FIGURE 3 | Examples of target-specific SINEUPs. (A) Scheme of FLAG tag-specific SINEUP. SINEUP-FLAG was designed to target the −40/+32 region around the ATG of 2XFLAG expression plasmid (pcDNA3-2XFLAG). **(B)** Activity of SINEUP-FLAG was tested in HEK 293T/17 cells transfected with pcDNA3-2XFLAG-TRAF6, DJ-1 and Hba, as indicated. Protein (top) and RNA (bottom) levels were analyzed by Western blot with anti-FLAG antibody and by qRT-PCR with SINEUP (white) and target-specific (gray) primers, respectively. SINEUP activity was calculated as increase in protein levels relative to empty control samples (fold changes are shown). In all conditions, sense RNA quantities were stable ($p > 0.05$). Data indicate mean ± standard deviation and are representative of three independent experiments.

was obtained combining SINEUP-GFP BD and ED from AS Uchl1 (**Figure 4A**) giving rise to a ≈250 nt long transcript. When transfected in HEK 293T/17 cells miniSINEUP-GFP promoted a 2.5 fold increase in GFP protein levels with unaffected mRNA quantities ($p = 0.11$) (**Figure 4B**). Interestingly, the activity of miniSINEUP was comparable to that obtained with SINEUP-GFP (1.6 fold in this experiment, 2.4 average increase in HEK 293T/17 cells, **Figure 2D**). Under these conditions, we observed a 10-fold excess of miniSINEUP-GFP RNA relative to canonical SINEUP-GFP, as expected from its reduced size. Despite elevated RNA quantities, no impact was observed on GFP mRNA, proving that miniSINEUP retains the very same post-transcriptional mechanism of its full-length counterpart.

SINEUPs can be Targeted to Endogenous mRNAs of Interest

The use of SINEUPs as a versatile tool to increase protein synthesis is strictly dependent on their ability to act on endogenous, cellular mRNAs transcribed from a genomic locus that does not present a natural SINEUP antisense gene.

To prove this crucial point we designed synthetic SINEUPs targeted to endogenous DJ-1 mRNA, a gene involved in recessive familial Parkinson's Disease (PD). We generated two SINEUP-DJ-1 constructs with two different BD elements: −40/+32 (long, L), from −40 nucleotides (in annotated 5′ untranslated region) before DJ-1 translation start site to +32 bases in the coding sequence, as well as −40/+4 (short, S), ending at the

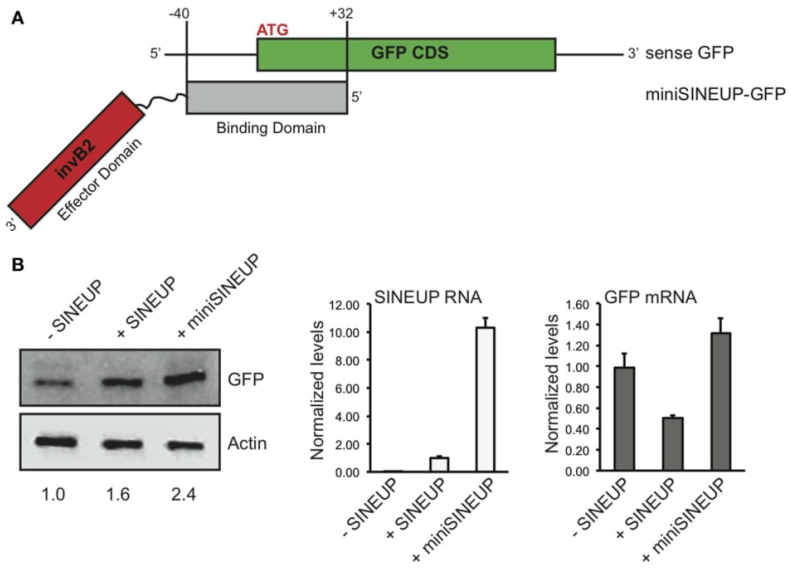


FIGURE 4 | Short SINEUPs are active. (A) Domain organization of miniSINEUP-GFP. Binding (gray) and effector (invB2, red) domains are indicated. **(B)** HEK 293T/17 cells were transfected with pEGFP-C2 and miniSINEUP-GFP (+miniSINEUP). SINEUP-GFP was included as positive control (+SINEUP). Control cells received pEGFP-C2 and pcDNA3.1(-) empty plasmid (-SINEUP). Total proteins (left) and RNA (right) were extracted and tested for GFP expression and GFP and SINEUP RNA quantities, respectively. Data are representative of three independent experiments and indicate mean ± standard deviation.

first nucleotide at 3′ of ATG (**Figure 5A**). Together with HEK 293T/17, we also carried out experiments in three human neuronal cell lines [SH-SY5Y, BE(2)-M17 and SK-N-SH cells] as representative of those frequently used as *in vitro* system to study the function of PD-associated genes. Cells transfected with an empty vector were used as control (-SINEUP).

As shown in **Figure 5B**, SINEUP activity on endogenous DJ-1 ranged from 1.7 to 2 fold with no changes in endogenous DJ-1 mRNA levels. SINEUP-DJ-1 activity was confirmed in all three human neuronal cell lines, proving the versatility of the technology (**Figures 5C–E**). Interestingly, the highest induction of DJ-1 protein levels was measured in SK-N-SH cells (3 fold).

In summary synthetic SINEUPs can be designed to mRNAs of interest that do not present a natural SINEUP in their genomic locus opening up the scalable use of SINEUPs to target endogenous protein coding transcripts of mammalian cells.

Discussion

To our knowledge SINEUPs are the only tool available so far that uses lncRNAs to enhance translation of target proteins. Their modular structure allows synthetic design of an overlapping region (BD) to target proteins of interest, without changing the overall structure of the original lncRNA. Indeed, by extracting solely the target-selecting BD and the ED from the original lncRNA, SINEUPs can be resized to miniSINEUPs retaining their activity.

Over competing technologies, SINEUPs have two major advantages: (1) they modulate translation of target mRNAs without introducing stable genomic changes into target cells; (2) their induction of selected protein is typically in a more physiological range (2-fold) than most conventional gene replacement strategies. These features render SINEUPs (and miniSINEUPs) a potentially interesting tool for a number of applications. First, SINEUPs may be used as reagents for molecular biology. As siRNAs have become an invaluable instrument to inhibit gene expression, many cases exist in which increasing the amount of a specific protein is required. SINEUPs may be designed to a single gene of interest or to a tag that is common to several targets and achieve translation enhancement, thus formally becoming the opposite counterpart of siRNAs. This is especially relevant in the nervous system where regulation of translation has an enormous impact on synaptic plasticity and memory formation. The synthesis of specific proteins requires a fine tuned control at single synapse and spines involving the translation of selective subtypes of mRNAs. The special challenges posed by the anatomo-functional organization of the brain require cellular machinery for controlling mRNAs localization according to the morphology and connectivity of single neuronal cell types. This is achieved at least in part by the use of different TSSs leading to the synthesis of mRNA isoforms with specific 5′UTR containing information for subcellular localization and translation (Baj et al., 2011). SINEUPs can thus induce translation from the mRNA isoform expressed in a defined

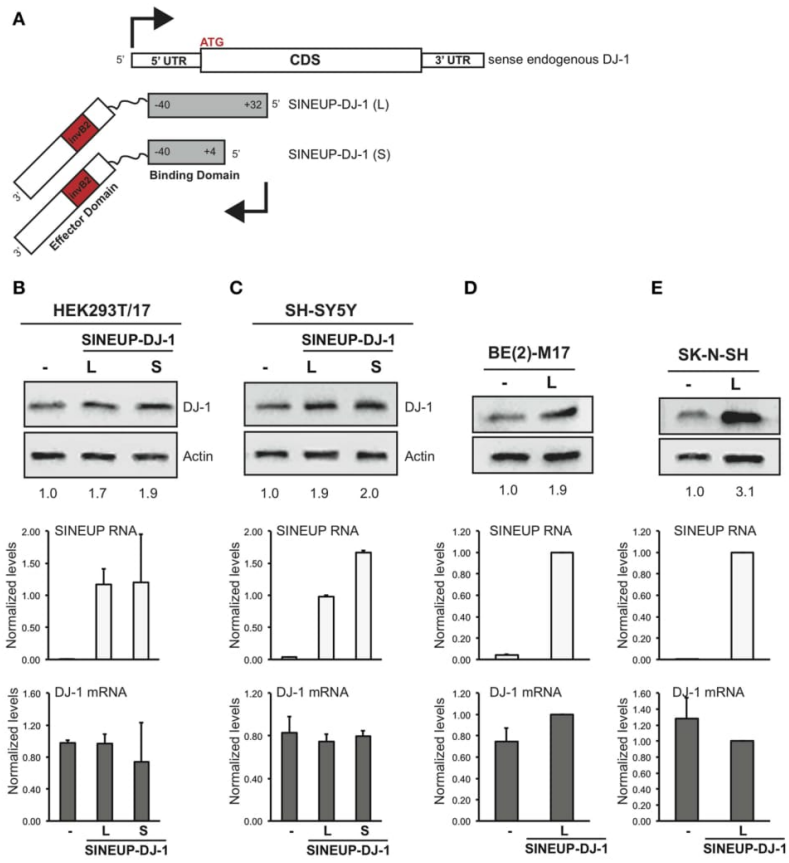


FIGURE 5 | SINEUP is active on endogenous mRNA. (A) Scheme of SINEUP-DJ-1. Two different SINEUPs were designed to target the −40/+32 (L) and −40/+4 (S) regions around the ATG of human DJ-1 gene. Activity of SINEUP-DJ-1 was tested in non-neuronal HEK 293T/17 **(B)**, and in neuronal SH-SY5Y **(C)**, BE(2)-M17 **(D)** and SK-N-SH **(E)** cells. Controls were transfected with empty pcDNA3. Protein (top) and RNA (bottom) levels were analyzed by Western blot with anti-DJ-1 antibody and by qRT-PCR with SINEUP (white) and target-specific (gray) primers, respectively. SINEUP activity was calculated as increase in protein levels relative to empty control samples (fold changes are shown). In all conditions, DJ-1 mRNA quantities were stable ($p > 0.05$). Data indicate mean ± standard deviation and are representative of three independent experiments.

cellular compartment of a selected neuronal cell type increasing specificity.

Here we show that a synthetic SINEUP against the endogenous mRNA for DJ-1, a gene involved in familial PD, is able to increase its protein synthesis. This is important since, to our knowledge, the mammalian genomic DJ-1 locus does not present a natural SINEUP. This experiment thus proves SINEUPs can potentially act on protein coding transcripts of mammalian cells whether or not they are under an endogenous SINEUP-mediated translational control.

Second, considering their effect on translation, SINEUPs may find applications in protein manufacturing. More than 130 therapeutic proteins are currently in use and many more are under development, including antibodies (Leader et al., 2008). Most production strategies have concentrated efforts in optimizing culture conditions and transcription of recombinant genes leaving room for improvement at post-transcriptional level. Recently, large-scale manufacturing platforms have been developed using transiently transfected cells, mainly CHO and HEK293 (Bandaranayake and Almo, 2014). The data presented here and elsewhere (Cotella et al., submitted) support the feasibility of SINEUPs and their potential to be integrated in existing platforms.

Several aspects regulating SINEUP efficacy with selected targets have still to be elucidated. Here we found that, despite identical sequence in the overlapping region, SINEUP-FLAG failed to up-regulate FLAG-TTRAP, whilst being the most effective with FLAG-TRAF6. Furthermore, the very same BD directed to a single mRNA species can increase protein levels with different efficacies according to the host cellular type. Finally, BD of different lengths can act unlike. In this context the role of the secondary structure of the target mRNAs around AUG remains unclear, although different levels of protein increase may be also accounted by turnover rates specific for each protein

or cellular context. While testing synthetic SINEUPs against a large repertory of endogenous mRNAs, we have found that a major cause of apparent lack of activity is due to the wrongful assumption that cells express the Refseq mRNA isoform of the gene. As shown by genome-wide analysis of TSS usage in mammalian cells, the complexity of alternative 5′ ends of mRNAs is staggering. Therefore, when available, we routinely interrogate the FANTOM5 dataset of CAGE libraries (Forrest et al., 2014) of the very same cells used in the experiments taking advantage of the online tool Zenbu (http://fantom.gsc.riken.jp/5/) to identify the correct AUG-surrounding region of the mRNA of interest. Importantly, we observed that high number of cell passages negatively influence SINEUP activity (data not shown). Global CAP-dependent translation is maintained through mTOR activity (Laplante and Sabatini, 2012) and is reduced in the majority of stress conditions (Holcik and Sonenberg, 2005; Sonenberg and Hinnebusch, 2009). We have previously showed that inhibition of mTOR with rapamycin induced an increase in UchL1 protein level dependent on the activity of the natural SINEUP AS Uchl1 (Carrieri et al., 2012). This occurs by triggering shuttling of AS Uchl1 RNA from the nucleus to the cytoplasm and the consequential increased association of Uchl1 mRNA to heavy polysomes for efficient translation. Here we showed that rapamycin and other stressors do not influence the amount of protein increase triggered by synthetic SINEUPs. We may hypothesize that by overexpression we saturate the cytoplasmic content of SINEUP RNA and/or the quantity of target mRNAs that can be associated to heavy polysomes. Further dissection of the cellular pathways that control SINEUP function and identification of SINEUP-binding proteins will provide fundamental insights to improve experimental design and answer these fundamental biological questions.

Finally, lncRNAs represent a new frontier in drug-development. The field of RNA therapeutics has emerged for its great potentials and is set to increase the number of targets beyond initial expectations (Kole et al., 2012). RNA-based drugs have been developed in the past two decades often relying on short non-coding molecules, ASOs or siRNAs, to degrade mRNA or miRNAs of interest. Most recently, lncRNAs appear as therapeutic targets of ASO technology (Modarresi et al., 2012), although the field is still at an early phase. The use of lncRNAs as tools to modulate gene expression is vastly unexplored. This suffers from the limited knowledge of lncRNAs' structure/function relationship and from major obstacles in delivering long RNA molecules. In this context, miniaturization of lncRNAs represents a prerequisite toward applicability in therapeutics. Here we demonstrate that SINEUPs' modular architecture can be employed to construct a miniSINEUP that maintains full-length activity with a length within the small RNA range. Knowing the ED tridimensional structure may provide further insights to help SINEUPs design and optimize activity with minimal length requirements.

In current medical practice there are several unmet therapeutic needs for increasing protein levels *in vivo*. Among them, haploinsufficiency is a condition that arises when the normal phenotype requires the protein product of both alleles, and reduction to 50% or less of gene function results in an abnormal phenotype. This is the cause of a wide spectrum of diseases including ataxias and intellectual and cognitive disabilities. An efficient SINEUP activity specific for the gene of interest would be in principle curative. Furthermore, in many complex and metabolic diseases the increase of pro-survival factors and dysregulated enzymes may impact the well being of patients. As an example, augmented production of neurotrophic factors has been proposed as therapeutic treatments for the majority of neurodegenerative diseases. Therefore SINEUP molecules specific for the transcripts selectively expressed at the site of injury may potentially slow or stop disease progression. Furthermore, it may avoid the unwanted side effects of unregulated expression in the brain that have halted many clinical trials in the past. Our ability to increase endogenous levels of the PD-associated DJ-1 protein represents the first formal prove that synthetic SINEUPs may be a new class of nucleic acid-based drugs.

We are conscious that any potential application of SINEUPs in therapy will strictly depend on their efficient delivery. To this purpose we are currently exploring the preservation of their activity with different delivery systems and chemical modifications.

In summary, here we show first evidences that synthetic SINEUPs may represent a scalable platform for manipulating gene expression of single mRNA species within their physiological range in an array of cell lines.

SINEUPs may thus become a new tool for laboratory experiments, for protein manufacturing and for potential therapeutic intervention *in vivo*.

Acknowledgments

We are indebted to all the members of the SG lab and to the SINEUP network (SISSA, University of Eastern Piedmont, RIKEN) for thought-provoking discussions. We thank Cristina Leonesi for technical support. We are grateful to SISSA technical and administrative staff, especially to J. Franzot, H. Krmac, Monica Sirk and Annalisa Sulli. This work was supported by the Italian Ministry of Education, University and Research (FIRB grant prot. RBAP11FRE9) to SG and FP.

Supplementary Material

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fncel.2015.00174/abstract

Supplementary Figure 1 | Treatment with stressful stimuli does not increase SINEUP activity in transfected cells. HEK 293T/17 cells were transfected with pEGFP in combination with SINEUP-GFP (+SINEUP) or control plasmid (-SINEUP). After transfection, cells were treated with rapamycin or doxorubicin as indicated. Lysates were probed anti-GFP antibody. β-actin was used as loading control. Fold-induction was calculated on Western blot images normalized to β-actin and relative to empty control samples.

Supplementary Figure 2 | SINEUP RNA is detected in the cytoplasm of transfected cells. HEK 293T/17 cells were transfected with pEGFP in combination with SINEUP-GFP (+SINEUP). RNA was purified from separated

nuclear and cytoplasmic fractions. RNA was reverse transcribed and probed for SINEUP RNA and GFP mRNA, as indicated. Purity of nuclear and cytoplasmic fractions was monitored by qRT-PCR on precursor rRNA. Data were normalized to the level of GAPDH in each fraction and analyzed with the ΔΔ Ct method. RNA levels in the cytoplasm were set to 1. Data indicate mean ± standard deviation and are calculated on 3 independent replicas.

Supplementary Figure 3 | SINEUP-increased targets can be detected with target-specific antibodies. HEK 293T/17 cells were transfected with pcDNA3-2XFLAG-TRAF6 in combination with SINEUP-FLAG (+SINEUP) or control plasmid (-SINEUP). Lysates were probed anti-TRAF6 antibody.

Supplementary Figure 4 | SINEUP-FLAG does not increase FLAG-TTRAP protein levels. HEK 293T/17 cells were transfected with pcDNA3-2XFLAG-TTRAP in combination with SINEUP-FLAG (+SINEUP) or control plasmid (-SINEUP). Lysates were probed anti-FLAG antibody. SINEUP RNA and TTRAP mRNA were quantified by qRT-PCR with specific primers. Data indicate average ± stdev and are representative of $n = 3$ independent experiments.

References

Angelini, M., Cannata, S., Mercaldo, V., Gibello, L., Santoro, C., Dianzani, I., et al. (2007). Missense mutations associated with Diamond-Blackfan anemia affect the assembly of ribosomal protein S19 into the ribosome. *Hum. Mol. Genet.* 16, 1720–1727. doi: 10.1093/hmg/ddm120

Baj, G., Leone, E., Chao, M. V., and Tongiorgi, E. (2011). Spatial segregation of BDNF transcripts enables BDNF to differentially shape distinct dendritic compartments. *Proc. Natl. Acad. Sci. U.S.A.* 108, 16813–16818. doi: 10.1073/pnas.1014168108

Bandaranayake, A. D., and Almo, S. C. (2014). Recent advances in mammalian protein production. *FEBS Lett.* 588, 253–260. doi: 10.1016/j.febslet.2013.11.035

Biagioli, M., Pinto, M., Cesselli, D., Zaninello, M., Lazarevic, D., Roncaglia, P., et al. (2009). Unexpected expression of alpha- and beta-globin in mesencephalic dopaminergic neurons and glial cells. *Proc. Natl. Acad. Sci. U.S.A.* 106, 15454–15459. doi: 10.1073/pnas.0813216106

Carrieri, C., Cimatti, L., Biagioli, M., Beugnet, A., Zucchelli, S., Fedele, S., et al. (2012). Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat. *Nature* 491, 454–457. doi: 10.1038/nature11508

Carrieri C., Forrest A. R., Santoro C., Persichetti F., Carninci P., Zucchelli, S. et al. (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. doi: 10.3389/fncel.2015.00114

Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., 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. doi: 10.1101/gr.132159.111

Djebali, S., Davis, C. A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., et al. (2012). Landscape of transcription in human cells. *Nature* 489, 101–108. doi: 10.1038/nature11233

Forrest, A. R., Kawaji, H., Rehli, M., Baillie, J. K., De Hoon, M. J., Haberle, V., et al. (2014). A promoter-level mammalian expression atlas. *Nature* 507, 462–470. doi: 10.1038/nature13182

Foti, R., Zucchelli, S., Biagioli, M., Roncaglia, P., Vilotti, S., Calligaris, R., et al. (2010). Parkinson disease-associated DJ-1 is required for the expression of the glial cell line-derived neurotrophic factor receptor RET in human neuroblastoma cells. *J. Biol. Chem.* 285, 18565–18574. doi: 10.1074/jbc.M109.088294

Herrera, F. E., Zucchelli, S., Jezierska, A., Lavina, Z. S., Gustincich, S., and Carloni, P. (2007). On the oligomeric state of DJ-1 protein and its mutants associated with Parkinson Disease. A combined computational and *in vitro* study. *J. Biol. Chem.* 282, 24905–24914. doi: 10.1074/jbc.M701013200

Holcik, M., and Sonenberg, N. (2005). Translational control in stress and apoptosis. *Nat. Rev. Mol. Cell Biol.* 6, 318–327. doi: 10.1038/nrm1618

Katayama, S., Tomaru, Y., Kasukawa, T., Waki, K., Nakanishi, M., Nakamura, M., et al. (2005). Antisense transcription in the mammalian transcriptome. *Science* 309, 1564–1566. doi: 10.1126/science.1112009

Kole, R., Krainer, A. R., and Altman, S. (2012). RNA therapeutics: beyond RNA interference and antisense oligonucleotides. *Nat. Rev. Drug Discov.* 11, 125–140. doi: 10.1038/nrd3625

Kordasiewicz, H. B., Stanek, L. M., Wancewicz, E. V., Mazur, C., Mcalonis, M. M., Pytel, K. A., et al. (2012). Sustained therapeutic reversal of Huntington's disease by transient repression of huntingtin synthesis. *Neuron* 74, 1031–1044. doi: 10.1016/j.neuron.2012.05.009

Laplante, M., and Sabatini, D. M. (2012). mTOR signaling in growth control and disease. *Cell* 149, 274–293. doi: 10.1016/j.cell.2012.03.017

Leader, B., Baca, Q. J., and Golan, D. E. (2008). Protein therapeutics: a summary and pharmacological classification. *Nat. Rev. Drug Discov.* 7, 21–39. doi: 10.1038/nrd2399

Lunardi, A., Chiacchiera, F., D'este, E., Carotti, M., Dal Ferro, M., Di Minin, G., et al. (2009). The evolutionary conserved gene C16orf35 encodes a nucleo-cytoplasmic protein that interacts with p73. *Biochem. Biophys. Res. Commun.* 388, 428–433. doi: 10.1016/j.bbrc.2009.08.027

Modarresi, F., Faghihi, M. A., Lopez-Toledano, M. A., Fatemi, R. P., Magistri, M., Brothers, S. P., et al. (2012). Inhibition of natural antisense transcripts *in vivo* results in gene-specific transcriptional upregulation. *Nat. Biotechnol.* 30, 453–459. doi: 10.1038/nbt.2158

Murayama, A., Ohmori, K., Fujimura, A., Minami, H., Yasuzawa-Tanaka, K., Kuroda, T., et al. (2008). Epigenetic control of rDNA loci in response to intracellular energy status. *Cell* 133, 627–639. doi: 10.1016/j.cell.2008.03.030

Qureshi, I. A., and Mehler, M. F. (2012). Emerging roles of non-coding RNAs in brain evolution, development, plasticity and disease. *Nat. Rev. Neurosci.* 13, 528–541. doi: 10.1038/nrn3234

Schmittgen, T. D., and Livak, K. J. (2008). Analyzing real-time PCR data by the comparative C(T) method. *Nat. Protoc.* 3, 1101–1108. doi: 10.1038/nprot.2008.73

Sonenberg, N., and Hinnebusch, A. G. (2009). Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell* 136, 731–745. doi: 10.1016/j.cell.2009.01.042

Valen, E., Pascarella, G., Chalk, A., Maeda, N., Kojima, M., Kawazu, C., et al. (2009). Genome-wide detection and analysis of hippocampus core promoters using DeepCAGE. *Genome Res.* 19, 255–265. doi: 10.1101/gr.084541.108

Van Bokhoven, H. (2011). Genetic and epigenetic networks in intellectual disabilities. *Annu. Rev. Genet.* 45, 81–104. doi: 10.1146/annurev-genet-110410-132512

Vilotti, S., Biagioli, M., Foti, R., Dal Ferro, M., Lavina, Z. S., Collavin, L., et al. (2012). The PML nuclear bodies-associated protein TTRAP regulates ribosome biogenesis in nucleolar cavities upon proteasome inhibition. *Cell Death Differ.* 19, 488–500. doi: 10.1038/cdd.2011.118

Wang, Y., Zhu, W., and Levy, D. E. (2006). Nuclear and cytoplasmic mRNA quantification by SYBR green based real-time RT-PCR. *Methods* 39, 356–362. doi: 10.1016/j.ymeth.2006.06.010

Yu, D., Pendergraff, H., Liu, J., Kordasiewicz, H. B., Cleveland, D. W., Swayze, E. E., et al. (2012). Single-stranded RNAs use RNAi to potently and allele-selectively inhibit mutant huntingtin expression. *Cell* 150, 895–908. doi: 10.1016/j.cell.2012.08.002

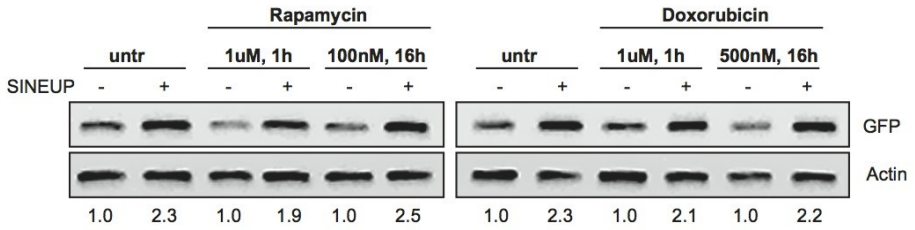
Zucchelli, S., Codrich, M., Marcuzzi, F., Pinto, M., Vilotti, S., Biagioli, M., et al. (2010). TRAF6 promotes atypical ubiquitination of mutant DJ-1 and alpha-synuclein and is localized to Lewy bodies in sporadic Parkinson's disease brains. *Hum. Mol. Genet.* 19, 3759–3770. doi: 10.1093/hmg/ddq290

Zucchelli, S., Marcuzzi, F., Codrich, M., Agostoni, E., Vilotti, S., Biagioli, M., et al. (2011). Tumor necrosis factor receptor-associated factor 6 (TRAF6) associates with huntingtin protein and promotes its atypical ubiquitination to enhance aggregate formation. *J. Biol. Chem.* 286, 25108–25117. doi: 10.1074/jbc.M110.187591

Zucchelli, S., Vilotti, S., Calligaris, R., Lavina, Z. S., Biagioli, M., Foti, R., et al. (2009). Aggresome-forming TTRAP mediates pro-apoptotic properties of Parkinson's disease-associated DJ-1 missense mutations. *Cell Death Differ.* 16, 428–438. doi: 10.1038/cdd.2008.169

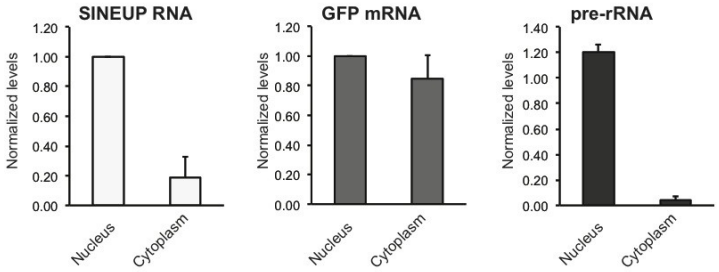
Conflict of Interest Statement: Stefano Gustincich, Piero Carninci, Claudio Santoro, Michael H. Jones and Silvia Zucchelli declare competing financial interests as co-founders and members of TransSINE Technologies (www.transsine.com). Stefano Gustincich, Piero Carninci and Silvia Zucchelli are named inventors in patent issued in the US Patent and Trademark Office on SINEUPs and licensed to TransSINE Technologies. Michael H. Jones is CEO of Cell Guidance Systems, a company distributing SINEUPs as laboratory reagents. Stefano Gustincich and Piero Carninci are co-founders of PARKscreen, an Italian SME aimed to use and develop therapeutic SINEUPs.

Supplementary Figure 1



Supplementary Figure 1. Treatment with stressful stimuli does not increase SINEUP activity in transfected cells. HEK 293T/17 cells were transfected with pEGFP in combination with SINEUP-GFP (+SINEUP) or control plasmid (-SINEUP). After transfection, cells were treated with rapamycin or doxorubicin as indicated. Lysates were probed anti-GFP antibody. β-actin was used as loading control. Fold-induction was calculated on Western blot images normalized to β-actin and relative to empty control samples.

Supplementary Figure 2

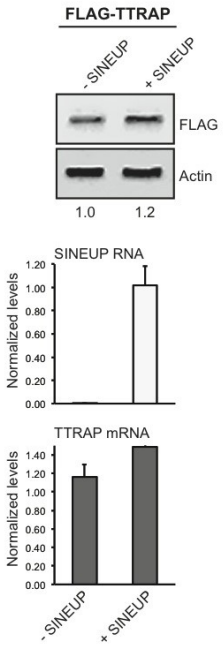


Supplementary Figure 2. SINEUP RNA is detected in the cytoplasm of transfected cells. HEK 293T/17 cells were transfected with pEGFP in combination with SINEUP-GFP (+SINEUP). RNA was purified from separated nuclear and cytoplasmic fractions. RNA was reverse transcribed and probed for SINEUP RNA and GFP mRNA, as indicated. Purity of nuclear and cytoplasmic fractions was monitored by qRT-PCR on precursor rRNA. Data were normalized to the level of GAPDH in each fraction and analyzed with the ΔΔCt method. RNA levels in the cytoplasm were set to 1. Data indicate mean ± standard deviation and are calculated on 3 independent replicas.

Supplementary Figure 3

Supplementary Figure 3. SINEUP-increased targets can be detected with target-specific antibodies. HEK 293T/17 cells were transfected with pcDNA3-2XFLAG-TRAF6 in combination with SINEUP-FLAG (+SINEUP) or control plasmid (-SINEUP). Lysates were probed anti-TRAF6 antibody.

Supplementary Figure 4



Supplementary Figure 4. SINEUP-FLAG does not increase FLAG-TTRAP protein levels. HEK 293T/17 cells were transfected with pcDNA3-2XFLAG-TTRAP in combination with SINEUP-FLAG (+SINEUP) or control plasmid (-SINEUP). Lysates were probed anti-FLAG antibody. SINEUP RNA and TTRAP mRNA were quantified by qRT-PCR with specific primers. Data indicate average ± stdev and are representative of n = 3 independent experiments.

DISCUSSION

Recent advances in sequencing technologies have radically changed our view of RNA. The existence of thousands of non-coding transcripts suggests that RNA is indeed more than just a messenger between genes and proteins. The discovery of regulatory RNAs, such as miRNAs, siRNAs and lncRNAs revealed a previously underestimated complexity underlying gene expression regulatory networks. At the same time, it paved the way for potential uses of such RNAs as tools to modulate gene expression *ad hoc in vitro* and *in vivo.*

Among ncRNAs, lncRNAs represent the major transcriptional output of mammalian cells. Even if the quest for function of lncRNAs is still on its way, a number of examples support their contribution to gene expression regulation through an array of different mechanisms (Huarte and Marín-Béjar, 2015). However, a common grammar seems to stand at the basis of their mode of action. Evidence suggests that lncRNAs work as flexible "modular scaffolds", recruiting and coordinating different effectors through discrete RNA domains (Guttman and Rinn, 2012). RNA folding is now believed to provide functional cues to lncRNA domains, as these display poor conservation within their primary sequence.

The use of lncRNAs as tools to modulate gene expression is vastly unexplored, partially due to limited knowledge of lncRNAs' structure/function relationship. Although a big challenge, unveiling the specific mechanisms ruling the activity of regulatory ncRNAs represents a crucial point for their application in diverse context, from biotechnology to therapy.

In this study we investigated the molecular mechanisms underlying biological activity of natural and synthetic SINEUPs. These are AS lncRNAs able to increase translation of partially overlapping transcripts. Their biological activity relies on the presence of an embedded retrotransposon of the SINEB2 type, which represents SINEUPs ED. In particular, we demonstrated the structural basis for the activation of protein synthesis mediated by the inverted SINEB2. Furthermore, we provided evidence about the ED working as interaction interface with protein partner ILF3, which is involved in regulation of SINEUPs localization. Finally, we showed that, two embedded TEs (inverted SINEB2 and Alu) together serve as SINEUPs nuclear retention signal. In particular, these findings add further layer of complexity to the

role of the ED which, besides conferring translation activation power, contributes to tune SINEUPs action by regulating their location.
After investigating the molecular rules of SINEUP activity, we focused on validating and optimizing synthetic SINEUPs as RNA tools to increase protein production. We eventually propose SINEUPs as the first scalable tool to increase synthesis of chosen proteins, with important applications in molecular biology experiments, protein manufacturing as well as in therapy.

SINEUPs as a paradigm for lncRNAs structure and function

The mouse genome contains approximately 350,000 SINEB2 sequences as independent transcriptional units or embedded in RNA polymerase II transcripts (Kramerov and Vassetzky, 2011a). Like all TEs, SINE elements have been originally presented as genomic parasites (Doolittle and Sapienza, 1980), due to the disruptive effects on gene expression caused by their insertions into coding or regulatory sequences in genomes. However, conversely to LINEs', SINEs insertions within coding regions seem to be more tolerated and have often resulted in the rise of functional elements of transcripts (Lander et al., 2001). In particular, “embedded” or “exonized” SINEs, as well as other TEs, have been recently proposed to work as portable domains responsible for lncRNAs regulatory functions (Johnson and Guigó, 2014; Kapusta and Feschotte, 2014; Zucchelli et al., 2015a; Carrieri et al., 2012). This is supported by many examples of repeat-rich, functional lncRNAs (Johnson and Guigó, 2014). A major issue to the “embedded domain” hypothesis is represented by the poor sequence conservation of lncRNAs during evolution. Different studies have reported that lncRNAs present highly conserved secondary and tertiary structures, which have been often shown to be related to their functions (Li et al., 2016). Theoretically, this concept may be applied to embedded TEs as well. Of notice, SINE elements have a highly conserved secondary structure that derives from the RNA species from which SINEs took their origin, that is tRNAs.
Carrieri and colleagues showed that mouse lncRNA AS Uchl1 is capable of enhancing translation of sense protein-coding Uchl1 mRNA through an embedded retrotransposon of the SINE type (Carrieri et al., 2012). In this context, we exploited the modular organization and the well-defined biological function of AS Uchl1 to

address the issue of structural/functional relationship in lncRNA domains. Previous data of chemical footprinting showed that the inverted SINEB2 mostly folds into helical structures, with several internal loops and hairpins. Similar structures have been reported to provide interfaces for specific recognition by molecular partners. For example, SRA (steroid receptor RNA activator), a breast cancer-linked lncRNA, which co-activates several nuclear receptors and proteins, is reported to have highly conserved helices, terminal loops, and bulges in many species (Novikova et al., 2012). Moreover, as regards SINEs, an extended stem-loop structure, with terminal loops and internal bulges, has been previously observed for other SINE RNAs (Sun et al., 2007), positioned either at the 5' end, as for BC1 (Rozhdestvensky et al., 2001), or at the 3' terminus, as in the case of salmon SmaI SINE (Kawagoe-Takaki et al., 2006).

The deletion of the SL1 structural motif abolished the ability of AS Uchl1 to increase endogenous UCHL1 protein levels. This supports a role of SL1 as a structural determinant of AS Uchl1 activity. This model was further confirmed by the validation of a structure/function pattern in synthetic SINEUP AS GFP. Hence, common mechanisms of action are likely to underlie the biological activity of natural and synthetic SINEUPs. However, further studies are needed to understand whether SL1 motif is the unique ED portion determining biological activity of SINEUPs.

According to mFOLD prediction (**Appendix Figure 1**), mutations characterizing AS Uchl1 delta SL1 exclusively result in disruption of the hairpin structure, without affecting the overall structural stability of the inverted SINEB2. This is in line with chemical footprinting data that indicate a stable "basal" structure of the molecule with a more flexible region at its "apical" part. Altogether, these results support a model of the invSINEB2 as an independent folding unit acting as ED thanks to a terminal stem loop structure.

Recently Shein and collegues identified a lncRNA antisense to the human protein phosphatase 1 regulatory subunit 12A (PPP1R12A), named as R12A-AS1, that functions as SINEUPs in human cells. R12A-AS1 overlaps the 5′ UTR and first coding exon of the PPP1R12A mRNA and contains a free right Alu monomer repeat element within its 3′ end. These domains are required for R12A-AS1-mediated PPP1R12A mRNA translation enhancement, coherently with what expected for

SINEUPs. Interestingly, a human SINE element -Alu monomer repeat- was for the first time found to provide for the domain conferring translation activity to SINEUPs, suggesting that different TEs can work as EDs (Schein et al., 2016). It will be interesting to verify if similar structural determinants confer biological activity to both mouse and human embedded SINEs or, alternatively, if different structural determinants are associated with the same functional output.

Phage display selection to identify SINEUP-binding proteins

The modular architecture of natural and synthetic SINEUPs strongly support a role of embedded TEs as functional units of lncRNAs. Interestingly, a number of studies have demonstrated that TEs found in lncRNAs sequences can provide modules for association with molecular partners. For example, an embedded TE (Alu repeat) modulates ANRIL activity by recruiting protein components of the Polycomb Repressor Complex (Holdt et al., 2013). Binding of dsRBP Staufen and subsequent STAUFEN-mediated degradation is triggered by the formation of dsRNA following hybridization between mRNAs and lncRNAs containing complementary Alu fragments (Gong and Maquat, 2011).

Previous data of chemical footprinting on the inverted SINEB2 from AS Uchl1 revealed the presence of structural domains providing ideal interfaces for association with protein partners. One of these (SL1) was shown to be required for biological activity of natural and synthetic SINEUPs. Given the well established role of protein partners in modulating lncRNAs function, we asked whether the embedded inverted SINEB2 could provide the recognition site for specific protein complexes modulating SINEUP activity. To profile SINEUPs interactors, we applied the RIDome pipeline (Patrucco et al., 2015). This *in vitro* selection methodology combines phage display with NGS, thus allowing an unbiased, high-throughput characterization of RBDs interacting with desired target RNAs. Indeed, as an *in vitro* approach, it allows carrying out serial screenings on different targets. These can be either full-length RNAs or specific domains of interest. We decided to focus our attention on the inverted SINEB2, as the core domain of SINEUPs biological activity. In parallel, we carried out a wider investigation of SINEUP-binding proteins by selecting the library on AS Uchl1 Δ5' RNA. This represents the common

backbone on which synthetic SINEUPs are built (Carrieri et al., 2012; Zucchelli et al., 2015b). Particularly, it contains the ED in an embedded format.

Phage libraries can faithfully represent full-length proteomes or domainomes of cells, with the advantage of coupling phenotype to genotype identification. Of notice, when "filtered" ORFs domain libraries are employed for *in vitro* selection of targets, specific functional domains involved in bait-binding can be directly identified.

The outcome of all selections consisted of a set of several hundreds genes. It is empirically known that most of them form the "noise" of the phage selection, as a large portion of these is represented by a reduced number of reads. Curiously, enrichment analysis revealed the presence of a clone by far more abundant than the others, in both inverted SINEB2 and AS Uchl1 Δ5'-driven selections. Similar results were observed when employing slightly different selection protocols with the same bait (i.e. ssDNA or tRNA as competitor), supporting the robustness of the selection procedure and the reliability of such result. The "top" clone corresponded to ILF3 dsRBM2. Preliminary biochemical validation of three differentially enriched RBPs (ILF3, SRSF5, HNRNPA3) commonly identified in different selections confirmed that differential enrichment did correspond to different binding specificity. Among these, ILF3 was the unique one specifically binding to the inverted SINEB2. This strengthens the importance of ELISA-based biochemical validation as first "specificity check point" along the pipeline. In this way, false positives arising from either technical limits of *in vitro* selection methodologies or possible artifacts introduced by NGS can thus be easily unmasked and excluded from subsequent validation steps.

ILF3 in SINEUP biology

ILF3 is a ubiquitously expressed dsRNA and DNA-binding protein. Firstly identified as a transcription factor in the IL-2 promoter-binding complex (Corthésy and Kao, 1994; Kao et al., 1994), it has been later found to be involved in diverse processes besides transcription (splicing, translation, etc.) and, more generally, in RNA metabolism (transport, localization, stability, etc.). Although the specific biological functions of ILF3 are not precisely defined, the protein appears to be essential for cellular development and integrity (Castella et al., 2015).

Different protein isoforms are generated by a combination of alternative splicing and different polyadenylation events (Castella et al., 2015). The most abundant splicing variants are known as nuclear factor (NF) 90 and NF110, of 90 and 110 kDa, respectively. These proteins are basically identical, except for the presence of an additional region of ~200 aa within NF110 C-terminal. Their common region contains a predicted nuclear localization signal, two dsRBMs and an RGG-rich sequence, possibly interacting with a single stranded RNA (ssRNA) or a ssDNA. ILF3 RNA-binding capability mainly relies on the two dsRBMs, referred to as dsRBM1 and dsRBM2. Interestingly, individual alignment of ILF3 reads relative to both inverted SINEB2 and AS Uchl1 Δ5'-selected libraries showed dominant mapping on dsRBM2. Sequencing of the starting (non-selected) library confirmed that such enrichment was exclusively observed after selections.

RNA recognition through dsRBDs is achieved by both sequence and shape determinants. In particular, hairpins apical loops provide "consensus" structures for dsRBPs (Masliah et al., 2013). A similar structural motif has been identified (Podbevšek et al., submitted) within the ED of SINEUPs, providing an optimal candidate domain mediating interactions with dsRBDs-containing proteins. Interestingly, a recent study reported that, together with structures, specific sequences in dsRNAs may affect target recognition by ILF3 dsRBMs, similarly to what happens for adenosine-to-inosine editing enzyme, ADAR2 (Jayachandran et al., 2016). In particular, ADAR activity is triggered by the presence of long dsRNA stretches. In human, transcripts containing exonized Alus are extensively edited, as their high frequency results in the formation of dsRNA due to base-pairing between consecutive Alus IRAlus (Athanasiadis et al., 2004; Kim et al., 2004; Levanon et al., 2004). Although the functional implications of ILF3/ADAR2 similarity are not clear, the hypothesis of ILF3 regulating ADAR2 activity by competing in binding with RNA seems reasonable. The association of ILF3 with a murine SINE would be coherent in such a scenario, even if the editing level in mouse SINEs is known to be much lower (Neeman et al., 2006).

Validation of binding between clone dsRBM2 and SINEUP baits *in vitro* pointed out higher ELISA signal for AS Uchl1 Δ5'. We speculate that this might be linked to specific folding features of the inverted SINEB2 when present in an embedded

format. For example, the formation of extra-dsRNA regions can be triggered, improving binding affinity. As a further proof of the specific nature of ILF3 dsRBM2 interaction with AS Uchl1, we stress that, *in vitro,* no binding has been observed with the other dsRBD present in ILF3 (dsRBM1). However, we cannot exclude the requirement of this domain in binding *in vivo*. Biochemical approaches provide in fact simplified models of the complex molecular dynamics characterizing interactions in cellular environments. Therefore, a crucial point was validation of AS Uchl1/ILF3 interaction *in vivo.* In particular, we were interested in assessing the requirement of the embedded inverted SINEB2 as domain mediating binding with ILF3 *in vivo.* Interestingly, we observed a reproducible enrichment AS Uchl1 RNA in endogenous ILF3 immunoprecipitates, which was remarkably reduced when the embedded inverted SINEB2 was removed. Coherently with what learnt from *in vitro* selection and validation, we thus demonstrated that binding with ILF3 is likely enabled by the inverted SINEB2. These findings strengthen the role of embedded TEs as functional domains mediating lncRNAs contacts with molecular partners (Johnson and Guigó, 2014; Carrieri et al., 2012; Zucchelli et al., 2015a).

AS Uchl1 is predominantly found in nuclei of cells upon physiologic conditions, even if low levels of endogenous SINEUP RNA are always detected in the cytoplasm (Carrieri et al., 2012). Notably, even when overexpressed, AS Uchl1 keeps a predominant nuclear localization. Curiously, AS Uchl1-derived synthetic SINEUPs showed a similar distribution in cells (Zucchelli et al., 2015b), suggesting a common mechanisms determining subcellular fate for natural and synthetic SINEUPs. On the other hand, ILF3 is enriched in nuclei, even if it can shuttle to the cytoplasm during specific phases of cell cycle or upon certain stress stimuli (Parrott et al., 2005; Matsumoto-Taniura et al., 1996; Kuwano et al., 2008). We found that AS Uchl1 localization is affected by ILF3 silencing, with a 10-20% increase of AS Uchl1 RNA detected in the cytoplasmic fraction. More interestingly, removal of the inverted SINEB2 phenocopied the distribution observed upon ILF3 knock-down. Hence, the inverted SINEB2 directs AS Uchl1 localization to ILF3-containing nuclear complexes, and is required to trap AS Uchl1 within cell nuclei. The ED is thus involved in regulating SINEUPs localization, likely providing an interface for association with ILF3-containing nuclear protein complexes. Similarly, even if not a

TE, a repetitive sequence named repeating RNA domain (RRD) in FIRRE lncRNA represents the binding site for nuclear protein hnRNP U. Conversely to ILF3, this protein is sufficient for sequestration of FIRRE in the nucleus. By engaging the RRD motif, the protein drives FIRRE focal localization at the site of transcription, where the lncRNA directs the formation of trans-chromosomal contacts (Hacisuleyman et al., 2014). Although both regulating localization, hnRNP U and ILF3 have opposite effects on their partners' s activity, which is enabled and inhibited, respectively.

ILF3 is not sufficient to retain AS Uchl1 in the nucleus, suggesting that other nuclear proteins could be directly responsible for AS Uchl1 bias towards nuclear localization. In this context, we showed that ILF3 interacts with $p54^{nrb}$, one of the core components of paraspeckles (Fox and Lamond, 2010). These are subnuclear bodies whose function is still unclear. They have been proposed to be involved in nuclear sequestration of specific factors, including repeat-containing edited RNAs (Prasanth et al., 2005). This would argue for a link between ILF3, editing and nuclear retention. So far these remain speculations, and further experiments should be carried out to investigate such hypothesis. Some evidence however exists in literature of ILF3 interacting with $p54^{nrb}$, even if not in the context of paraspeckles (Yamauchi et al., 2012).

TEs as nuclear localization signals of SINEUPs

A considerable portion of AS Uchl1 deleted of the inverted SINEB2 was still found in the nuclear compartment. We thus investigated the presence of other sequences involved in regulation of SINEUPs location, focusing on the partial Alu repeat present immediately downstream the inverted SINEB2. Strikingly, combined removal of inverted SINEB2 and Alu significantly perturbed AS Uchl1 distribution across the cell, with about 70% shuffling to the cytoplasmic compartment. Recently, IRAlus contained in lincRNA-p21 have been shown to fold into specific structures required for nuclear localization. Indeed, mutations disrupting such secondary structures resulted in altered lincRNA-p21 distribution (Chillón and Pyle, 2016). It would be tempting to hypothesize a similar mechanism underlying cytoplasmic location of AS Uchl1 lacking both TEs, SINEB2 and Alu. Tandem inverted SINEB2 and Alu elements would in this context provide “unconventional” inverted repeats

dictating SINEUPs nuclear localization. Coherently, heterodimeric structures of SINEs belonging to different families have been reported previously (i.e. 7SL RNA/tRNA) (Kramerov and Vassetzky, 2011b), even if no specific function have been assigned to these yet.

Embedded TEs of synthetic SINEUPs are of murine origin; it is interesting to notice that they work as nuclear localization signal even in human cell lines (HEK 293T/17). A different situation occurs with FIRRE RRD, which has been shown to work as species-specific nuclear localization signal in Sox2 mRNA chimeras, in a hnRNP U-dependent fashion (Hacisuleyman et al., 2016).

As a consequence of the absence of the ED, AS Uchl1 mutants lacking TEs are predominantly found in the cytoplasm; however, they are not functional. This suggests that the cytoplasmic availability of functional SINEUPs is likely under strict control. Hence, it might exist a pool of cellular nuclear-retained RNAs normally kept latent in the nucleus, that are ready to respond to acute stimuli by changing their subcellular localization. An example fulfilling this model is represented by mouse CTN-RNA. This is retained in the nucleus but, following stress-induced post-transcriptional cleavage of its 3' end, is efficiently exported and translated. Noteworthy, three inverted repeats of SINE origin are present within its 3' end and are likely involved in CTN-RNA sequestration (Prasanth et al., 2005). As CTN-RNA has been proposed as non-coding nuclear reservoir of coding mRNAs, we can infer that natural SINEUPs are stored in nuclei as “inactive” molecules, but they can rapidly be transferred to translation sites within the cytoplasm in order to face specific cell needs. However, if we partially unveiled some of the mechanisms regulating nuclear retention, the molecular dynamics underlying SINEUPs export still await further exploration.

SINEUPs as versatile tools to increase translation of selected target mRNAs

The potential use of AS Uchl1-derived lncRNAs as enhancers of target mRNAs translation opens the way to interesting biotechnological and therapeutical applications. As shown previously (Carrieri et al., 2012), by exploiting the characteristic modular architecture of natural SINEUPs, it is indeed possible to engineer synthetic SINEUPs directed against desired targets by solely manipulating the BD. Hence, SINEUPs represent a powerful tool to modulate gene expression *in vitro* and *in vivo*. In this context, two are the main advantages of SINEUP technology: 1) no stable genomic changes are introduced and 2) a gentle increase of selected proteins, within physiological ranges, is achieved. These features determine the applicability of SINEUPs as reagents for molecular biology, when increasing the amount of a specific protein is required. SINEUPs may be designed to protein tags commonly used in eukaryotic expression systems, with a single SINEUP theoretically being able to target several N-terminally tagged proteins. We showed this ability with SINEUP-FLAG and other groups provide similar experimental evidence with SINEUPs against HA and GFP tags (Yao et al., 2015).

A crucial point was the demonstration that synthetic SINEUPs are capable of acting on protein coding transcripts of mammalian cells that do not present endogenous SINEUPs. To this, we design a synthetic SINEUP targeting endogenous Parkinson's disease-associated DJ-1. This proved to be efficient in manipulating DJ-1 expression within physiological ranges, thus paving the way for employment of SINEUPs in therapy. As the main limitation in application of SINEUPs as naked RNA therapeutics is delivery of long molecules, we aimed at synthesizing the shortest functional miniSINEUP, in which BD and ED represent the only retained sequences. MiniSINEUPs keep biological activity in a miniaturized version within the range of small RNAs length. In particular, the activity of miniSINEUPs, where BD and ED are isolated from the rest of the sequence, represent the formal prove that SINEUPs domains work as independent units presenting their own structure and function.

RNA therapeutics provides a huge potential in increasing the range of targets beyond the scope of existing pharmacological drugs (Gustincich et al., 2016). In this context, SINEUPs and miniSINEUPs molecules may represent an ideal tool for increasing

gene expression *in vivo* for therapy. In particular, SINEUPs would provide ideal therapeutic tools for treatment of haploinsufficiency, as well as many complex and metabolic diseases where the increase of pro-survival factors and dysregulated enzymes may impact the well being of patients. In principle, SINEUPs would act on the target mRNA under its physiological regulation *in vivo*, limiting adverse influences on the tissue. Even if *in vivo* application of SINEUPs is still in its infancy, two recent studies reported the efficacy of SINEUPs in modulating gene expression *in vivo*, encouraging their use as RNA drugs for the future (Indrieri et al., 2016; Long et al., 2016).

Eventually, considering their effect on translation, SINEUPs may find applications in protein manufacturing (Zucchelli et al., 2016; Patrucco et al., 2015b). A number of therapeutic proteins are currently in use and many more are under development, including antibodies (Leader et al., 2008). In this field, optimization of culture yield represents a main issue (Bandaranayake and Almo, 2014). We and other groups (Patrucco et al., 2015b) demonstrated that SINEUP technology can be exploited, in combination with existing platforms, to improve the efficiency of such processes.

LIST OF PUBLICATIONS

SINEUPs are modular antisense long non-coding RNAs that increase synthesis of target proteins in cells.

Zucchelli S*, Fasolo F*, Russo R, Cimatti L, Patrucco L, Takahashi H, Jones MH, Santoro C, Sblattero D, Cotella D, Persichetti F, Carninci P, Gustincich S.

* These authors contributed equally to this work

Front Cell Neurosci. 2015 May 13;9:174. doi: 10.3389/fncel.2015.00174.eCollection 2015.

SINEUPs: A new class of natural and synthetic antisense long non-coding RNAs that activate translation.

Zucchelli S, Cotella D, Takahashi H, Carrieri C, Cimatti L, Fasolo F, Jones MH, Sblattero D, Sanges R, Santoro C, Persichetti F, Carninci P, Gustincich S.

RNA Biol. 2015;12(8):771-9. doi: 10.1080/15476286.2015.1060395. Review.

Engineering mammalian cell factories with SINEUP noncoding RNAs to improve translation of secreted proteins.

Patrucco L, Chiesa A, Soluri MF, Fasolo F, Takahashi H, Carninci P, Zucchelli S, Santoro C, Gustincich S, Sblattero D, Cotella D.

Gene. 2015 Sep 15;569(2):287-93. doi: 10.1016/j.gene.2015.05.070. Epub 2015 Jun 2.

Structural determinants of the SINEB2 element embedded in the long non-coding RNA activator of translation *AS Uchl1*.

Peter Podbevšek, Francesca Fasolo, Carlotta Bon, Laura Cimatti, Sabine Reißer, Piero Carninci, Giovanni Bussi, Silvia Zucchelli, Janez Plavec* and Stefano Gustincich*.

Submitted.

Identification of functional features of synthetic SINEUPs, antisense lncRNAs that specifically enhance protein translation.

Hazuki Takahashi, Ana Kozhuharova, Harshita Sharma, Masakazu Hirose, Takako Ohyama, Francesca Fasolo, Toshio Yamazaki, Diego Cotella, Claudio Santoro, Silvia Zucchelli, Stefano Gustincich and Piero Carninci.

Submitted.

The interleukin enhancer-binding Factor 3 (ILF3) binds to the embedded SINEB2 element of SINEUP AS Uchl1 long non-coding RNA influencing its nuclear localization.

Fasolo F, Patrucco L, Santoro C, Carninci P, Zucchelli S, Sblattero D, Cotella D* and Gustincich S*.

Manuscript in preparation.

BIBLIOGRAPHY

Allen, T.A., Von Kaenel, S., Goodrich, J.A., and Kugel, J.F. (2004). The SINE-encoded mouse B2 RNA represses mRNA transcription in response to heat shock. Nat. Struct. Mol. Biol. *11*, 816–821.

An, H.J., Lee, D., Lee, K.H., and Bhak, J. (2004). The association of Alu repeats with the generation of potential AU-rich elements (ARE) at 3' untranslated regions. BMC Genomics *5*, 97.

Athanasiadis, A., Rich, A., and Maas, S. (2004a). Widespread A-to-I RNA editing of Alu-containing mRNAs in the human transcriptome. PLoS Biol. *2*, e391.

Aviv, T., Lin, Z., Ben-Ari, G., Smibert, C.A., and Sicheri, F. (2006). Sequence-specific recognition of RNA hairpins by the SAM domain of Vts1p. Nat. Struct. Mol. Biol. *13*, 168–176.

Baltz, A.G., Munschauer, M., Schwanhäusser, B., Vasile, A., Murakawa, Y., Schueler, M., Youngs, N., Penfold-Brown, D., Drew, K., Milek, M., et al. (2012). The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. Mol. Cell *46*, 674–690.

Bandaranayake, A.D., and Almo, S.C. (2014). Recent advances in mammalian protein production. FEBS Lett. *588*, 253–260.
Bartel, D.P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. Cell *116*, 281–297.

Battle, D.J., and Doudna, J.A. (2001). The stem-loop binding protein forms a highly stable and specific complex with the 3' stem-loop of histone mRNAs. RNA N. Y. N *7*, 123–132.

Beckmann, B.M., Castello, A., and Medenbach, J. (2016). The expanding universe of ribonucleoproteins: of novel RNA-binding proteins and unconventional interactions. Pflugers Arch. *468*, 1029–1040.

Bejerano, G., Lowe, C.B., Ahituv, N., King, B., Siepel, A., Salama, S.R., Rubin, E.M., Kent, W.J., and Haussler, D. (2006). A distal enhancer and an ultraconserved exon are derived from a novel retroposon. Nature *441*, 87–90.

Blow, M., Futreal, P.A., Wooster, R., and Stratton, M.R. (2004). A survey of RNA editing in human brain. Genome Res. *14*, 2379–2387.

Boria, I., Boatti, L., Pesole, G., and Mignone, F. (2013). NGS-Trex: Next Generation Sequencing Transcriptome profile explorer. BMC Bioinformatics *14 Suppl 7*, S10.

Borodulina, O.R., and Kramerov, D.A. (2001). Short interspersed elements (SINEs) from insectivores. Two classes of mammalian SINEs distinguished by A-rich tail structure. Mamm. Genome *12*, 779–786.

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.

Bourque, G., Leong, B., Vega, V.B., Chen, X., Lee, Y.L., Srinivasan, K.G., Chew, J.-L., Ruan, Y., Wei, C.-L., Ng, H.H., et al. (2008). Evolution of the mammalian transcription factor binding repertoire via transposable elements. Genome Res. *18*, 1752–1762.

Brown, C.J., Hendrich, B.D., Rupert, J.L., Lafrenière, R.G., Xing, Y., Lawrence, J., and Willard, H.F. (1992). The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. Cell *71*, 527–542.

Buckley, P.T., Lee, M.T., Sul, J.-Y., Miyashiro, K.Y., Bell, T.J., Fisher, S.A., Kim, J., and Eberwine, J. (2011). Cytoplasmic intron sequence-retaining transcripts can be dendritically targeted via ID element retrotransposons. Neuron *69*, 877–884.

Cabili, M.N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., and Rinn, J.L. (2011). Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev. *25*, 1915–1927.

Cabili, M.N., Dunagin, M.C., McClanahan, P.D., Biaesch, A., Padovan-Merhar, O., Regev, A., Rinn, J.L., and Raj, A. (2015). Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution. Genome Biol. *16*, 20.

Carlevaro-Fita, J., Rahim, A., Guigó, R., Vardy, L.A., and Johnson, R. (2016). Cytoplasmic long noncoding RNAs are frequently bound to and degraded at ribosomes in human cells. RNA *22*, 867–882.

Campbell, Z.T., Bhimsaria, D., Valley, C.T., Rodriguez-Martinez, J.A., Menichelli, E., Williamson, J.R., Ansari, A.Z., and Wickens, M. (2012). Cooperativity in RNA-protein interactions: global analysis of RNA binding specificity. Cell Rep. *1*, 570–581.

Carmen, S., and Jermutus, L. (2002). Concepts in antibody phage display. Brief. Funct. Genomic. Proteomic. *1*, 189–203.

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.

Castella, S., Bernard, R., Corno, M., Fradin, A., and Larcher, J.-C. (2015). Ilf3 and NF90 functions in RNA biology: Functions of Ilf3 and NF90 in RNA biology. Wiley Interdiscip. Rev. RNA *6*, 243–256.

Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B.M., Strein, C., Davey, N.E., Humphreys, D.T., Preiss, T., Steinmetz, L.M., et al. (2012). Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. Cell *149*, 1393–1406.

Castello, A., Fischer, B., Hentze, M.W., and Preiss, T. (2013). RNA-binding proteins in Mendelian disease. Trends Genet. TIG *29*, 318–327.

Cesana, M., Cacchiarelli, D., Legnini, I., Santini, T., Sthandier, O., Chinappi, M., Tramontano, A., and Bozzoni, I. (2011). A Long Noncoding RNA Controls Muscle Differentiation by Functioning as a Competing Endogenous RNA. Cell *147*, 358–369.

Chen, L.-L. (2016). Linking Long Noncoding RNA Localization and Function. Trends Biochem. Sci. *41*, 761–772.

Chen, L.-L., and Carmichael, G.G. (2008). Gene regulation by SINES and inosines: biological consequences of A-to-I editing of Alu element inverted repeats. Cell Cycle *7*, 3294–3301.

Chen, X., Yan, C.C., Zhang, X., and You, Z.-H. (2016). Long non-coding RNAs and complex diseases: from experimental results to computational models. Brief. Bioinform. Bbw060.

Chen, J., Wang, R., Zhang, K., and Chen, L.-B. (2014). Long non-coding RNAs in non-small cell lung cancer as biomarkers and therapeutic targets. J. Cell. Mol. Med. *18*, 2425–2436.

Cheng, J., Kapranov, P., Drenkow, J., Dike, S., Brubaker, S., Patel, S., Long, J., Stern, D., Tammana, H., Helt, G., et al. (2005). Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. Science *308*, 1149–1154.

Chillón, I., and Pyle, A.M. (2016). Inverted repeat *Alu* elements in the human lincRNA-p21 adopt a conserved secondary structure that regulates RNA function. Nucleic Acids Res. Gkw599.

Collins, K. (2008). Physiological assembly and activity of human telomerase complexes. Mech. Ageing Dev. *129*, 91–98.

Cook, K.B., Hughes, T.R., and Morris, Q.D. (2015). High-throughput characterization of protein-RNA interactions. Brief. Funct. Genomics *14*, 74–89.

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

Corthésy, B., and Kao, P.N. (1994). Purification by DNA affinity chromatography of two polypeptides that contact the NF-AT DNA binding site in the interleukin 2 promoter. J. Biol. Chem. *269*, 20682–20690.

Crick, F.H. (1958). On protein synthesis. Symp. Soc. Exp. Biol. *12*, 138–163.
D'Angelo, S., Velappan, N., Mignone, F., Santoro, C., Sblattero, D., Kiss, C., and Bradbury, A.R. (2011). Filtering "genic" open reading frames from genomic DNA samples for advanced annotation. BMC Genomics *12*, S5.

D'Angelo, S., Mignone, F., Deantonio, C., Di Niro, R., Bordoni, R., Marzari, R., De Bellis, G., Not, T., Ferrara, F., Bradbury, A., et al. (2013). Profiling celiac disease antibody repertoire. Clin. Immunol. *148*, 99–109.

Danner, S., and Belasco, J.G. (2001). T7 phage display: A novel genetic selection system for cloning RNA-binding proteins from cDNA libraries. Proc. Natl. Acad. Sci. *98*, 12954–12959.

Darnell, R.B. (2010a). RNA Regulation in Neurologic Disease and Cancer. Cancer Res. Treat. *42*, 125.

Darnell, R.B. (2010b). HITS-CLIP: panoramic views of protein-RNA regulation in living cells. Wiley Interdiscip. Rev. RNA *1*, 266–286.

Dawkins, R., and Krebs, J.R. (1979). Arms Races between and within Species. Proc. R. Soc. B Biol. Sci. *205*, 489–511.

DeCerbo, J., and Carmichael, G.G. (2005). SINEs point to abundant editing in the human genome. Genome Biol. *6*, 216.

Deininger, P.L., Tiedge, H., Kim, J., and Brosius, J. (1996). Evolution, expression, and possible function of a master gene for amplification of an interspersed repeated DNA family in rodents. Prog. Nucleic Acid Res. Mol. Biol. *52*, 67–88.

Deragon, J.-M., and Zhang, X. (2006). Short interspersed elements (SINEs) in plants: origin, classification, and use as phylogenetic markers. Syst. Biol. *55*, 949–956.

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.

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

Di Niro, R., Sulic, A.-M., Mignone, F., D'Angelo, S., Bordoni, R., Iacono, M., Marzari, R., Gaiotto, T., Lavric, M., Bradbury, A.R.M., et al. (2010). Rapid interactome profiling by massive sequencing. Nucleic Acids Res. *38*, e110.

DiDonato, M., Deacon, A.M., Klock, H.E., McMullan, D., and Lesley, S.A. (2004). A scaleable and integrated crystallization pipeline applied to mining the Thermotoga maritima proteome. J. Struct. Funct. Genomics *5*, 133–146.

van Dijk, E.L., Chen, C.L., d'Aubenton-Carafa, Y., Gourvennec, S., Kwapisz, M., Roche, V., Bertrand, C., Silvain, M., Legoix-Né, P., Loeillet, S., et al. (2011). XUTs are a class of Xrn1-sensitive antisense regulatory non-coding RNA in yeast. Nature *475*, 114–117.

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.

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

Duszczyk, M.M., Wutz, A., Rybin, V., and Sattler, M. (2011). The Xist RNA A-repeat comprises a novel AUCG tetraloop fold and a platform for multimerization. RNA *17*, 1973–1982.

Eddy, S.R. (2001). Non-coding RNA genes and the modern RNA world. Nat. Rev. Genet. *2*, 919–929.

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

Ellington, A.D., and Szostak, J.W. (1990). In vitro selection of RNA molecules that bind specific ligands. Nature *346*, 818–822.

ENCODE Project Consortium (2004). The ENCODE (ENCyclopedia Of DNA Elements) Project. Science *306*, 636–640.

Fan, X.C., and Steitz, J.A. (1998). HNS, a nuclear-cytoplasmic shuttling sequence in HuR. Proc. Natl. Acad. Sci. *95*, 15293–15298.

FANTOM Consortium and the RIKEN PMI and CLST (DGT), Forrest, A.R.R., Kawaji, H., Rehli, M., Baillie, J.K., de Hoon, M.J.L., Haberle, V., Lassmann, T., Kulakovskiy, I.V., Lizio, M., et al. (2014). A promoter-level mammalian expression atlas. Nature *507*, 462–470.

Fatica, A., and Bozzoni, I. (2013). Long non-coding RNAs: new players in cell differentiation and development. Nat. Rev. Genet. *15*, 7–21.

Faulkner, G.J., Kimura, Y., Daub, C.O., Wani, S., Plessy, C., Irvine, K.M., Schroder, K., Cloonan, N., Steptoe, A.L., Lassmann, T., et al. (2009). The regulated retrotransposon transcriptome of mammalian cells. Nat. Genet. *41*, 563–571.

Fica, S.M., Tuttle, N., Novak, T., Li, N.-S., Lu, J., Koodathingal, P., Dai, Q., Staley, J.P., and Piccirilli, J.A. (2013). RNA catalyses nuclear pre-mRNA splicing. Nature *503*, 229–234.

Fields, S., and Song, O. (1989). A novel genetic system to detect protein-protein interactions. Nature *340*, 245–246.

Foti, R., Zucchelli, S., Biagioli, M., Roncaglia, P., Vilotti, S., Calligaris, R., Krmac, H., Girardini, J.E., Del Sal, G., and Gustincich, S. (2010). Parkinson disease-associated DJ-1 is required for the expression of the glial cell line-derived neurotrophic factor receptor RET in human neuroblastoma cells. J. Biol. Chem. *285*, 18565–18574.

Fox, A.H., and Lamond, A.I. (2010). Paraspeckles. Cold Spring Harb. Perspect. Biol. *2*, a000687–a000687.

Gerstberger, S., Hafner, M., and Tuschl, T. (2014). A census of human RNA-binding proteins. Nat. Rev. Genet. *15*, 829–845.

Glisovic, T., Bachorik, J.L., Yong, J., and Dreyfuss, G. (2008). RNA-binding proteins and post-transcriptional gene regulation. FEBS Lett. *582*, 1977–1986.

Gogolevsky, K.P., Vassetzky, N.S., and Kramerov, D.A. (2008). Bov-B-mobilized SINEs in vertebrate genomes. Gene *407*, 75–85.

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. (2011a). lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu elements. Nature *470*, 284–288.

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

Gustincich, S., Zucchelli, S., and Mallamaci, A. (2016). The Yin and Yang of nucleic acid-based therapy in the brain. Prog. Neurobiol.

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

Guttman, M., Amit, I., Garber, M., French, C., Lin, M.F., Feldser, D., Huarte, M., Zuk, O., Carey, B.W., Cassady, J.P., et al. (2009). Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature *458*, 223–227.

Guttman, M., Donaghey, J., Carey, B.W., Garber, M., Grenier, J.K., Munson, G., Young, G., Lucas, A.B., Ach, R., Bruhn, L., et al. (2011). lincRNAs act in the circuitry controlling pluripotency and differentiation. Nature *477*, 295–300.

Hacisuleyman, E., Goff, L.A., Trapnell, C., Williams, A., Henao-Mejia, J., Sun, L., McClanahan, P., Hendrickson, D.G., Sauvageau, M., Kelley, D.R., et al. (2014). Topological organization of multichromosomal regions by the long intergenic noncoding RNA Firre. Nat. Struct. Mol. Biol. *21*, 198–206.

Hacisuleyman, E., Shukla, C.J., Weiner, C.L., and Rinn, J.L. (2016). Function and evolution of local repeats in the Firre locus. Nat. Commun. *7*, 11021.
Hamilton, A.J., and Baulcombe, D.C. (1999). A species of small antisense RNA in posttranscriptional gene silencing in plants. Science *286*, 950–952.

Harada, K., Martin, S.S., and Frankel, A.D. (1996). Selection of RNA-binding peptides in vivo. Nature *380*, 175–179.

Häsler, J., and Strub, K. (2006). Alu elements as regulators of gene expression. Nucleic Acids Res. *34*, 5491–5497.

Häsler, J., Samuelsson, T., and Strub, K. (2007). Useful “junk”: Alu RNAs in the human transcriptome. Cell. Mol. Life Sci. *64*, 1793–1800.

Havecker, E.R., Gao, X., and Voytas, D.F. (2004). The diversity of LTR retrotransposons. Genome Biol. *5*, 225.

van Heesch, S., van Iterson, M., Jacobi, J., Boymans, S., Essers, P.B., de Bruijn, E., Hao, W., MacInnes, A.W., Cuppen, E., and Simonis, M. (2014). Extensive localization of long noncoding RNAs to the cytosol and mono- and polyribosomal complexes. Genome Biol. *15*, R6.

Heger, A., and Holm, L. (2003). Exhaustive enumeration of protein domain families. J. Mol. Biol. *328*, 749–767.

Hogan, D.J., Riordan, D.P., Gerber, A.P., Herschlag, D., and Brown, P.O. (2008). Diverse RNA-Binding Proteins Interact with Functionally Related Sets of RNAs, Suggesting an Extensive Regulatory System. PLoS Biol. *6*, e255.

Holdt, L.M., Hoffmann, S., Sass, K., Langenberger, D., Scholz, M., Krohn, K., Finstermeier, K., Stahringer, A., Wilfert, W., Beutner, F., et al. (2013). Alu elements

in ANRIL non-coding RNA at chromosome 9p21 modulate atherogenic cell functions through trans-regulation of gene networks. PLoS Genet. *9*, e1003588.

Huang, C.R.L., Burns, K.H., and Boeke, J.D. (2012). Active Transposition in Genomes. Annu. Rev. Genet. *46*, 651–675.

Huarte, M., and Marín-Béjar, O. (2015). Long noncoding RNAs: from identification to functions and mechanisms. Adv. Genomics Genet. 257.
Huarte, M., Guttman, M., Feldser, D., Garber, M., Koziol, M.J., Kenzelmann-Broz, D., Khalil, A.M., Zuk, O., Amit, I., Rabani, M., et al. (2010). A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. Cell *142*, 409–419.

Hung, T., Wang, Y., Lin, M.F., Koegel, A.K., Kotake, Y., Grant, G.D., Horlings, H.M., Shah, N., Umbricht, C., Wang, P., et al. (2011). Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. Nat. Genet. *43*, 621–629.

Incarnato, D., Neri, F., Anselmi, F., and Oliviero, S. (2014). Genome-wide profiling of mouse RNA secondary structures reveals key features of the mammalian transcriptome. Genome Biol. *15*, 1.

Indrieri, A., Grimaldi, C., Zucchelli, S., Tammaro, R., Gustincich, S., and Franco, B. (2016). Synthetic long non-coding RNAs [SINEUPs] rescue defective gene expression in vivo. Sci. Rep. *6*, 27315.

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.

Iyer, M.K., Niknafs, Y.S., Malik, R., Singhal, U., Sahu, A., Hosono, Y., Barrette, T.R., Prensner, J.R., Evans, J.R., Zhao, S., et al. (2015). The landscape of long noncoding RNAs in the human transcriptome. Nat. Genet. *47*, 199–208.

Jacob, F., and Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. J. Mol. Biol. *3*, 318–356.

Jayachandran, U., Grey, H., and Cook, A.G. (2016). Nuclear factor 90 uses an ADAR2-like binding mode to recognize specific bases in dsRNA. Nucleic Acids Res. *44*, 1924–1936.

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.

Jones, S., Daley, D.T., Luscombe, N.M., Berman, H.M., and Thornton, J.M. (2001). Protein-RNA interactions: a structural analysis. Nucleic Acids Res. *29*, 943–954.

Kao, P.N., Chen, L., Brock, G., Ng, J., Kenny, J., Smith, A.J., and Corthésy, B. (1994). Cloning and expression of cyclosporin A- and FK506-sensitive nuclear factor of activated T-cells: NF45 and NF90. J. Biol. Chem. *269*, 20691–20699.

Kapitonov, V.V., and Jurka, J. (2003). A novel class of SINE elements derived from 5S rRNA. Mol. Biol. Evol. *20*, 694–702.

Kaplan, G., and Gershoni, J.M. (2012). A general insert label for peptide display on chimeric filamentous bacteriophages. Anal. Biochem. *420*, 68–72.

Kapusta, A., and Feschotte, C. (2014a). Volatile evolution of long noncoding RNA repertoires: mechanisms and biological implications. Trends Genet. TIG *30*, 439–452.

Kapusta, A., Kronenberg, Z., Lynch, V.J., Zhuo, X., Ramsay, L., Bourque, G., Yandell, M., and Feschotte, C. (2013). Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs. PLoS Genet. *9*, e1003470.

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.

Kato, Y., Kaneda, M., Hata, K., Kumaki, K., Hisano, M., Kohara, Y., Okano, M., Li, E., Nozaki, M., and Sasaki, H. (2007). Role of the Dnmt3 family in de novo methylation of imprinted and repetitive sequences during male germ cell development in the mouse. Hum. Mol. Genet. *16*, 2272–2280.

Kawagoe-Takaki, H., Nameki, N., Kajikawa, M., and Okada, N. (2006). Probing the secondary structure of salmon SmaI SINE RNA. Gene *365*, 67–73.

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

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.

Kim, D.D.Y., Kim, T.T.Y., Walsh, T., Kobayashi, Y., Matise, T.C., Buyske, S., and Gabriel, A. (2004). Widespread RNA editing of embedded alu elements in the human transcriptome. Genome Res. *14*, 1719–1725.

Kim, E.Z., Wespiser, A.R., and Caffrey, D.R. (2016). The domain structure and distribution of Alu elements in long noncoding RNAs and mRNAs. RNA N. Y. N *22*, 254–264.

Kim, T.-K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Laptewicz, M., Barbara-Haley, K., Kuersten, S., et al. (2010). Widespread transcription at neuronal activity-regulated enhancers. Nature *465*, 182–187.

Kino, T., Hurt, D.E., Ichijo, T., Nader, N., and Chrousos, G.P. (2010). Noncoding RNA Gas5 Is a Growth Arrest- and Starvation-Associated Repressor of the Glucocorticoid Receptor. Sci. Signal. *3*, ra8-ra8.

Koh, Y.Y., and Wickens, M. (2014). Identifying proteins that bind a known RNA sequence using the yeast three-hybrid system. Methods Enzymol. *539*, 195–214.

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.

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.

Krayev, A.S., Kramerov, D.A., Skryabin, K.G., Ryskov, A.P., Bayev, A.A., and Georgiev, G.P. (1980). The nucleotide sequence of the ubiquitous repetitive DNA sequence B1 complementary to the most abundant class of mouse fold-back RNA. Nucleic Acids Res. *8*, 1201–1215.

Krayev, A.S., Markusheva, T.V., Kramerov, D.A., Ryskov, A.P., Skryabin, K.G., Bayev, A.A., and Georgiev, G.P. (1982). Ubiquitous transposon-like repeats B1 and B2 of the mouse genome: B2 sequencing. Nucleic Acids Res. *10*, 7461–7475.

Kuwano, Y., Kim, H.H., Abdelmohsen, K., Pullmann, R., Martindale, J.L., Yang, X., and Gorospe, M. (2008). MKP-1 mRNA stabilization and translational control by RNA-binding proteins HuR and NF90. Mol. Cell. Biol. *28*, 4562–4575.

Kuwano, Y., Pullmann, R., Marasa, B.S., Abdelmohsen, K., Lee, E.K., Yang, X., Martindale, J.L., Zhan, M., and Gorospe, M. (2010). NF90 selectively represses the translation of target mRNAs bearing an AU-rich signature motif. Nucleic Acids Res. *38*, 225–238.

Lagos-Quintana, M., Rauhut, R., Lendeckel, W., and Tuschl, T. (2001). Identification of novel genes coding for small expressed RNAs. Science *294*, 853–

858.
Laird-Offringa, I.A., and Belasco, J.G. (1996). In vitro genetic analysis of RNA-binding proteins using phage display libraries. Methods Enzymol. *267*, 149–168.

Lambert, N., Robertson, A., Jangi, M., McGeary, S., Sharp, P.A., and Burge, C.B. (2014). RNA Bind-n-Seq: quantitative assessment of the sequence and structural binding specificity of RNA binding proteins. Mol. Cell *54*, 887–900.

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

Lasa, I., Toledo-Arana, A., and Gingeras, T.R. (2012). An effort to make sense of antisense transcription in bacteria. RNA Biol. *9*, 1039–1044.

Leader, B., Baca, Q.J., and Golan, D.E. (2008). Protein therapeutics: a summary and pharmacological classification. Nat. Rev. Drug Discov. *7*, 21–39.

Lee, J.Y., Ji, Z., and Tian, B. (2008). Phylogenetic analysis of mRNA polyadenylation sites reveals a role of transposable elements in evolution of the 3'-end of genes. Nucleic Acids Res. *36*, 5581–5590.

Lee, R.C., Feinbaum, R.L., and Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell *75*, 843–854.

Leung, E.K.Y., Suslov, N., Tuttle, N., Sengupta, R., and Piccirilli, J.A. (2011). The mechanism of peptidyl transfer catalysis by the ribosome. Annu. Rev. Biochem. *80*, 527–555.

Levanon, E.Y., Eisenberg, E., Yelin, R., Nemzer, S., Hallegger, M., Shemesh, R., Fligelman, Z.Y., Shoshan, A., Pollock, S.R., Sztybel, D., et al. (2004). Systematic identification of abundant A-to-I editing sites in the human transcriptome. Nat. Biotechnol. *22*, 1001–1005.

Lev-Maor, G., Ram, O., Kim, E., Sela, N., Goren, A., Levanon, E.Y., and Ast, G. (2008). Intronic Alus influence alternative splicing. PLoS Genet. *4*, e1000204.

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*, 702.

Liu, B., Sun, L., Liu, Q., Gong, C., Yao, Y., Lv, X., Lin, L., Yao, H., Su, F., Li, D., et al. (2015). A cytoplasmic NF-κB interacting long noncoding RNA blocks IκB phosphorylation and suppresses breast cancer metastasis. Cancer Cell *27*, 370–381.

Long, H., Yao, Y., Jin, S., Yu, Y., Hu, X., Zhuang, F., Zhang, H., and Wu, Q. (2016). RNAe in a transgenic growth hormone mouse model shows potential for use in gene therapy. Biotechnol. Lett.

Lunde, B.M., Moore, C., and Varani, G. (2007). RNA-binding proteins: modular design for efficient function. Nat. Rev. Mol. Cell Biol. *8*, 479–490.

Lunyak, V.V., Prefontaine, G.G., Núñez, E., Cramer, T., Ju, B.-G., Ohgi, K.A., Hutt, K., Roy, R., García-Díaz, A., Zhu, X., et al. (2007). Developmentally regulated activation of a SINE B2 repeat as a domain boundary in organogenesis. Science *317*, 248–251.

Lustig, A.J. (2004). Telomerase RNA: a flexible RNA scaffold for telomerase biosynthesis. Curr. Biol. CB *14*, R565-567.

Lynch, V.J., May, G., and Wagner, G.P. (2011). Regulatory evolution through divergence of a phosphoswitch in the transcription factor CEBPB. Nature *480*, 383–386.

Maenner, S., Blaud, M., Fouillen, L., Savoye, A., Marchand, V., Dubois, A., Sanglier-Cianférani, S., Van Dorsselaer, A., Clerc, P., Avner, P., et al. (2010). 2-D Structure of the A Region of Xist RNA and Its Implication for PRC2 Association. PLoS Biol. *8*, e1000276.

Mariner, P.D., Walters, R.D., Espinoza, C.A., Drullinger, L.F., Wagner, S.D., Kugel, J.F., and Goodrich, J.A. (2008). Human Alu RNA is a modular transacting repressor of mRNA transcription during heat shock. Mol. Cell *29*, 499–509.

Masliah, G., Barraud, P., and Allain, F.H.-T. (2013). RNA recognition by double-stranded RNA binding domains: a matter of shape and sequence. Cell. Mol. Life Sci. CMLS *70*, 1875–1895.

Matsumoto-Taniura, N., Pirollet, F., Monroe, R., Gerace, L., and Westendorf, J.M. (1996). Identification of novel M phase phosphoproteins by expression cloning. Mol. Biol. Cell *7*, 1455–1469.

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

McHugh, C.A., Russell, P., and Guttman, M. (2014). Methods for comprehensive experimental identification of RNA-protein interactions. Genome Biol. *15*, 203.

Medstrand, P., van de Lagemaat, L.N., and Mager, D.L. (2002). Retroelement distributions in the human genome: variations associated with age and proximity to genes. Genome Res. *12*, 1483–1495.

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

Mitton-Fry, R.M., DeGregorio, S.J., Wang, J., Steitz, T.A., and Steitz, J.A. (2010). Poly(A) tail recognition by a viral RNA element through assembly of a triple helix. Science *330*, 1244–1247.

Moore, M.J. (2005). From birth to death: the complex lives of eukaryotic mRNAs. Science *309*, 1514–1518.

Murayama, A., Ohmori, K., Fujimura, A., Minami, H., Yasuzawa-Tanaka, K., Kuroda, T., Oie, S., Daitoku, H., Okuwaki, M., Nagata, K., et al. (2008). Epigenetic control of rDNA loci in response to intracellular energy status. Cell *133*, 627–639.
Nagai, K. (1996). RNA-protein complexes. Curr. Opin. Struct. Biol. *6*, 53–61.
Neeman, Y., Levanon, E.Y., Jantsch, M.F., and Eisenberg, E. (2006). RNA editing level in the mouse is determined by the genomic repeat repertoire. RNA N. Y. N *12*, 1802–1809.

Nekrutenko, A., and Li, W.H. (2001). Transposable elements are found in a large number of human protein-coding genes. Trends Genet. TIG *17*, 619–621.

Nishihara, H., Terai, Y., and Okada, N. (2002). Characterization of novel Alu- and tRNA-related SINEs from the tree shrew and evolutionary implications of their origins. Mol. Biol. Evol. *19*, 1964–1972.

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.

Noh, J.H., Kim, K.M., Abdelmohsen, K., Yoon, J.-H., Panda, A.C., Munk, R., Kim, J., Curtis, J., Moad, C.A., Wohler, C.M., et al. (2016). HuR and GRSF1 modulate the nuclear export and mitochondrial localization of the lncRNA RMRP. Genes Dev. *30*, 1224–1239.

Novikova, I.V., Hennelly, S.P., and Sanbonmatsu, K.Y. (2012). Structural architecture of the human long non-coding RNA, steroid receptor RNA activator. Nucleic Acids Res. *40*, 5034–5051.

Ohshima, K., and Okada, N. (2005). SINEs and LINEs: symbionts of eukaryotic genomes with a common tail. Cytogenet. Genome Res. *110*, 475–490.

Oie, S., Matsuzaki, K., Yokoyama, W., Tokunaga, S., Waku, T., Han, S.-I., Iwasaki, N., Mikogai, A., Yasuzawa-Tanaka, K., Kishimoto, H., et al. (2014). Hepatic rRNA transcription regulates high-fat-diet-induced obesity. Cell Rep. *7*, 807–820.

Ostertag, E.M., and Kazazian, H.H. (2001). Biology of mammalian L1 retrotransposons. Annu. Rev. Genet. *35*, 501–538.

Parrott, A.M., Walsh, M.R., Reichman, T.W., and Mathews, M.B. (2005). RNA binding and phosphorylation determine the intracellular distribution of nuclear factors 90 and 110. J. Mol. Biol. *348*, 281–293.

Patrucco, L., Peano, C., Chiesa, A., Guida, F., Luisi, I., Boria, I., Mignone, F., De Bellis, G., Zucchelli, S., Gustincich, S., et al. (2015a). Identification of novel proteins binding the AU-rich element of α-prothymosin mRNA through the selection of open

reading frames (RIDome). RNA Biol. *12*, 1289–1300.

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

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

Piriyapongsa, J., Mariño-Ramírez, L., and Jordan, I.K. (2007). Origin and evolution of human microRNAs from transposable elements. Genetics *176*, 1323–1337.

Plath, K., Mlynarczyk-Evans, S., Nusinow, D.A., and Panning, B. (2002). Xist RNA and the mechanism of X chromosome inactivation. Annu. Rev. Genet. *36*, 233–278.

Poliseno, L., Salmena, L., Zhang, J., Carver, B., Haveman, W.J., and Pandolfi, P.P. (2010). A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. Nature *465*, 1033–1038.

Postepska-Igielska, A., Giwojna, A., Gasri-Plotnitsky, L., Schmitt, N., Dold, A., Ginsberg, D., and Grummt, I. (2015). LncRNA Khps1 Regulates Expression of the Proto-oncogene SPHK1 via Triplex-Mediated Changes in Chromatin Structure. Mol. Cell *60*, 626–636.

Prasanth, K.V., Prasanth, S.G., Xuan, Z., Hearn, S., Freier, S.M., Bennett, C.F., Zhang, M.Q., and Spector, D.L. (2005). Regulating Gene Expression through RNA Nuclear Retention. Cell *123*, 249–263.

Ray, D., Kazan, H., Chan, E.T., Peña Castillo, L., Chaudhry, S., Talukder, S., Blencowe, B.J., Morris, Q., and Hughes, T.R. (2009). Rapid and systematic analysis of the RNA recognition specificities of RNA-binding proteins. Nat. Biotechnol. *27*, 667–670.

Rinn, J.L., and Chang, H.Y. (2012). Genome regulation by long noncoding RNAs. Annu. Rev. Biochem. *81*, 145–166.

Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Brugmann, S.A., Goodnough, L.H., Helms, J.A., Farnham, P.J., Segal, E., et al. (2007). Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell *129*, 1311–1323.

Roberts, R.W., and Szostak, J.W. (1997). RNA-peptide fusions for the in vitro selection of peptides and proteins. Proc. Natl. Acad. Sci. U. S. A. *94*, 12297–12302.

Rozhdestvensky, T.S., Kopylov, A.M., Brosius, J., and Hüttenhofer, A. (2001). Neuronal BC1 RNA structure: evolutionary conversion of a tRNA(Ala) domain into an extended stem-loop structure. RNA N. Y. N *7*, 722–730.

Salditt-Georgieff, M., Harpold, M.M., Wilson, M.C., and Darnell, J.E. (1981). Large heterogeneous nuclear ribonucleic acid has three times as many 5' caps as polyadenylic acid segments, and most caps do not enter polyribosomes. Mol. Cell. Biol. *1*, 179–187.

Sánchez, Y., and Huarte, M. (2013). Long Non-Coding RNAs: Challenges for Diagnosis and Therapies. Nucleic Acid Ther. *23*, 15–20.

Savva, Y.A., Rieder, L.E., and Reenan, R.A. (2012). The ADAR protein family. Genome Biol. *13*, 252.

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.

Schmittgen, T.D., and Livak, K.J. (2008). Analyzing real-time PCR data by the comparative C(T) method. Nat. Protoc. *3*, 1101–1108.

Schramm, L., and Hernandez, N. (2002). Recruitment of RNA polymerase III to its target promoters. Genes Dev. *16*, 2593–2620.

Siprashvili, Z., Webster, D.E., Kretz, M., Johnston, D., Rinn, J.L., Chang, H.Y., and Khavari, P.A. (2012). Identification of proteins binding coding and non-coding human RNAs using protein microarrays. BMC Genomics *13*, 633.

Shu, L., Yan, W., and Chen, X. (2006). RNPC1, an RNA-binding protein and a target of the p53 family, is required for maintaining the stability of the basal and stress-induced p21 transcript. Genes Dev. *20*, 2961–2972.

Smith, G.P. (1985). Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. Science *228*, 1315–1317.

Smith, M.A., Gesell, T., Stadler, P.F., and Mattick, J.S. (2013). Widespread purifying selection on RNA structure in mammals. Nucleic Acids Res. *41*, 8220–8236.

Sorek, R., Ast, G., and Graur, D. (2002). Alu-containing exons are alternatively spliced. Genome Res. *12*, 1060–1067.

Spigoni, G., Gedressi, C., and Mallamaci, A. (2010). Regulation of Emx2 expression by antisense transcripts in murine cortico-cerebral precursors. PloS One *5*, e8658.

Spitale, R.C., Tsai, M.-C., and Chang, H.Y. (2011). RNA templating the epigenome: long noncoding RNAs as molecular scaffolds. Epigenetics *6*, 539–543.

Su, M., Han, D., Boyd-Kirkup, J., Yu, X., and Han, J.-D.J. (2014). Evolution of Alu

elements toward enhancers. Cell Rep. *7*, 376–385.

Sun, B.K., Deaton, A.M., and Lee, J.T. (2006). A transient heterochromatic state in Xist preempts X inactivation choice without RNA stabilization. Mol. Cell *21*, 617–628.

Sun, F.-J., Fleurdépine, S., Bousquet-Antonelli, C., Caetano-Anollés, G., and Deragon, J.-M. (2007). Common evolutionary trends for SINE RNA structures. Trends Genet. TIG *23*, 26–33.

Tenenbaum, S.A., Carson, C.C., Lager, P.J., and Keene, J.D. (2000). Identifying mRNA subsets in messenger ribonucleoprotein complexes by using cDNA arrays. Proc. Natl. Acad. Sci. U. S. A. *97*, 14085–14090.

Tichon, A., Gil, N., Lubelsky, Y., Havkin Solomon, T., Lemze, D., Itzkovitz, S., Stern-Ginossar, N., and Ulitsky, I. (2016). A conserved abundant cytoplasmic long noncoding RNA modulates repression by Pumilio proteins in human cells. Nat. Commun. *7*, 12209.

Tripathi, V., Ellis, J.D., Shen, Z., Song, D.Y., Pan, Q., Watt, A.T., Freier, S.M., Bennett, C.F., Sharma, A., Bubulya, P.A., et al. (2010). The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. Mol. Cell *39*, 925–938.

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.

Tuerk, C., and Gold, L. (1990). Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science *249*, 505–510.

Ule, J., Jensen, K.B., Ruggiu, M., Mele, A., Ule, A., and Darnell, R.B. (2003). CLIP identifies Nova-regulated RNA networks in the brain. Science *302*, 1212–1215.

Ulitsky, I., and Bartel, D.P. (2013). lincRNAs: genomics, evolution, and mechanisms. Cell *154*, 26–46.

Vance, K.W., and Ponting, C.P. (2014). Transcriptional regulatory functions of nuclear long noncoding RNAs. Trends Genet. *30*, 348–355.

Vassetzky, N.S., Ten, O.A., and Kramerov, D.A. (2003). B1 and related SINEs in mammalian genomes. Gene *319*, 149–160.

Veniaminova, N.A., Vassetzky, N.S., and Kramerov, D.A. (2007). B1 SINEs in different rodent families. Genomics *89*, 678–686.

Versteeg, R., van Schaik, B.D.C., van Batenburg, M.F., Roos, M., Monajemi, R.,

Caron, H., Bussemaker, H.J., and van Kampen, A.H.C. (2003). The human transcriptome map reveals extremes in gene density, intron length, GC content, and repeat pattern for domains of highly and weakly expressed genes. Genome Res. *13*, 1998–2004.

Viranaicken, W., Gasmi, L., Chaumet, A., Durieux, C., Georget, V., Denoulet, P., and Larcher, J.-C. (2011). L-Ilf3 and L-NF90 Traffic to the Nucleolus Granular Component: Alternatively-Spliced Exon 3 Encodes a Nucleolar Localization Motif. PLoS ONE *6*, e22296.

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.

Wang, K.C., and Chang, H.Y. (2011a). Molecular Mechanisms of Long Noncoding RNAs. Mol. Cell *43*, 904–914.

Wang, K.C., Yang, Y.W., Liu, B., Sanyal, A., Corces-Zimmerman, R., Chen, Y., Lajoie, B.R., Protacio, A., Flynn, R.A., Gupta, R.A., et al. (2011). A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. Nature *472*, 120–124.

Wang, T., Zeng, J., Lowe, C.B., Sellers, R.G., Salama, S.R., Yang, M., Burgess, S.M., Brachmann, R.K., and Haussler, D. (2007). Species-specific endogenous retroviruses shape the transcriptional network of the human tumor suppressor protein p53. Proc. Natl. Acad. Sci. U. S. A. *104*, 18613–18618.

Wang, Y., Zhu, W., and Levy, D.E. (2006). Nuclear and cytoplasmic mRNA quantification by SYBR green based real-time RT-PCR. Methods *39*, 356–362.

Weinberg, R.A., and Penman, S. (1968). Small molecular weight monodisperse nuclear RNA. J. Mol. Biol. *38*, 289–304.

Werner, A. (2005). Natural antisense transcripts. RNA Biol. *2*, 53–62.

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J.L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., et al. (2007). A unified classification system for eukaryotic transposable elements. Nat. Rev. Genet. *8*, 973–982.

Wutz, A., Rasmussen, T.P., and Jaenisch, R. (2002). Chromosomal silencing and localization are mediated by different domains of Xist RNA. Nat. Genet. *30*, 167–174.

Yamauchi, T., Nakamura, N., Hiramoto, M., Yuri, M., Yokota, H., Naitou, M., Takeuchi, M., Yamanaka, K., Kita, A., Nakahara, T., et al. (2012). Sepantronium bromide (YM155) induces disruption of the ILF3/p54(nrb) complex, which is required for survivin expression. Biochem. Biophys. Res. Commun. *425*, 711–716.

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

Yin, Q.-F., Yang, L., Zhang, Y., Xiang, J.-F., Wu, Y.-W., Carmichael, G.G., and Chen, L.-L. (2012). Long Noncoding RNAs with snoRNA Ends. Mol. Cell *48*, 219–230.

Yoon, J.-H., Abdelmohsen, K., Srikantan, S., Yang, X., Martindale, J.L., De, S., Huarte, M., Zhan, M., Becker, K.G., and Gorospe, M. (2012a). LincRNA-p21 suppresses target mRNA translation. Mol. Cell *47*, 648–655.

Yu, W., Gius, D., Onyango, P., Muldoon-Jacobs, K., Karp, J., Feinberg, A.P., and Cui, H. (2008). Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA. Nature *451*, 202–206.

Zacchi, P., Sblattero, D., Florian, F., Marzari, R., and Bradbury, A.R.M. (2003). Selecting open reading frames from DNA. Genome Res. *13*, 980–990.

Zappulla, D.C., and Cech, T.R. (2004). From The Cover: Yeast telomerase RNA: A flexible scaffold for protein subunits. Proc. Natl. Acad. Sci. *101*, 10024–10029.

Zhang, C., and Darnell, R.B. (2011). Mapping in vivo protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data. Nat. Biotechnol. *29*, 607–614.

Zhang, Z., and Carmichael, G.G. (2001). The fate of dsRNA in the nucleus: a p54nrb-containing complex mediates the nuclear retention of promiscuously A-to-I edited RNAs. Cell *106*, 465–476.

Zhang, Y., Zhang, X.-O., Chen, T., Xiang, J.-F., Yin, Q.-F., Xing, Y.-H., Zhu, S., Yang, L., and Chen, L.-L. (2013). Circular Intronic Long Noncoding RNAs. Mol. Cell *51*, 792–806.

Zhao, J., Sun, B.K., Erwin, J.A., Song, J.-J., and Lee, J.T. (2008). Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. Science *322*, 750–756.
Zietkiewicz, E., and Labuda, D. (1996). Mosaic evolution of rodent B1 elements. J. Mol. Evol. *42*, 66–72.

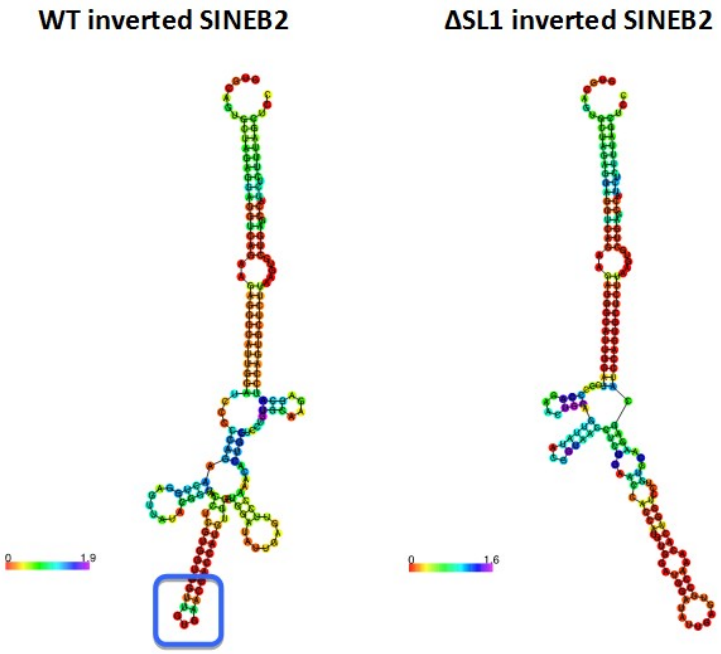
Zucchelli, S., Cotella, D., Takahashi, H., Carrieri, C., Cimatti, L., Fasolo, F., Jones, M., Sblattero, D., Sanges, R., Santoro, C., et al. (2015a). SINEUPs: A new class of natural and synthetic antisense long non-coding RNAs that activate translation. RNA Biol. *12*, 771–779.

Zucchelli, S., Fasolo, F., Russo, R., Cimatti, L., Patrucco, L., Takahashi, H., Jones, M.H., Santoro, C., Sblattero, D., Cotella, D., et al. (2015b). SINEUPs are modular

antisense long non-coding RNAs that increase synthesis of target proteins in cells. Front. Cell. Neurosci. *9*.

Zucchelli, S., Patrucco, L., Persichetti, F., Gustincich, S., and Cotella, D. (2016). Engineering Translation in Mammalian Cell Factories to Increase Protein Yield: The Unexpected Use of Long Non-Coding SINEUP RNAs. Comput. Struct. Biotechnol. J. *14*, 404–410.

APPENDIX



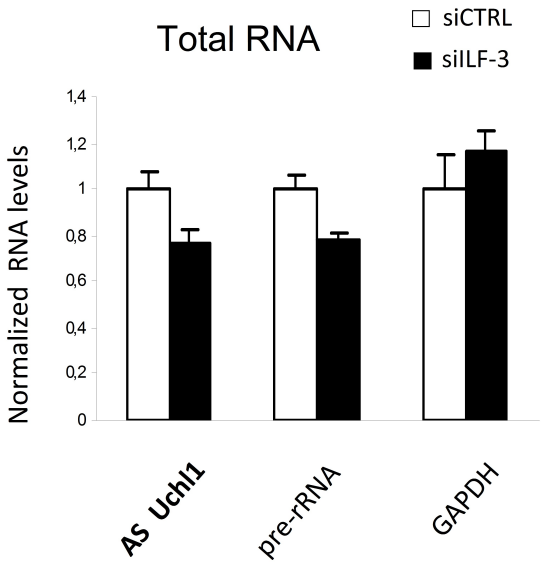
Appendix Figure 1. Prediction of wild type (left) and mutated (right) AS Uchl1 inverted SINEB2 secondary structure. Secondary structure models were generated using mFOLD software. The blue square highlights the stem-loop 1 (SL1).

Intersected Selections	Total Elements	Elements ID
AS Uchl1 Δ5'_tRNA	18	CCDC124
AS Uchl1 Δ5'_ssDNA		DNTTIP1
inverted SINEB2_tRNA		**HNRNPA3**
inverted SINEB2_ssDNA		**ILF3**
		LOC400550
		MTND4P12
		NEAT1
		PAM
		PCBD2
		PPM1G
		RPN1
		SENP3
		SENP3-EIF4A1
		STX1A
		UBXN1
		VAT1
		VPS41
		WIBG

Appendix Table 1. List of clones (alphabetic order) common to AS Uchl1 Δ5'_tRNA, AS Uchl1 Δ5'_ssDNA, inverted SINEB2_tRNA, inverted SINEB2_ssDNA phage display selections. In bold are ORFs chosen for preliminary validation.

Intersected Selections	Total Elements	Elements ID
AS Uchl1 Δ5'_tRNA inverted SINEB2_tRNA inverted SINEB2_ssDNA	16	AP3D1 DAB2 DNAJB2 EBF4 **ILF3** MBTPS1 MDH2 MTND5P11 NARF PEA15 PFKL PHC1 PKN1 SCAMP3 SPON2 **SRSF5**

Appendix Table 2. List of clones (alphabetic order) common to AS Uchl1 Δ5'_tRNA, inverted SINEB2_tRNA, inverted SINEB2_ssDNA phage display selections. In bold are ORFs chosen for preliminary validation.



Appendix Figure 2. Total RNA in ILF3-silenced and control HEK 293T/17. Total RNA levels of AS Uchl1, pre-rRNA and GAPDH were quatified by qRT-PCR. Values were normalizad on beta actin. No significative differences in total RNA content between siCTRL (white bars) and siILF3 (black bars) samples were detected. Data are representative of three independent experiments and indicate mean ± st dev**.**

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Stefano Gustincich, for giving me the opportunity to join his laboratory. These have been years of deep professional and personal growth.

I had the chance to meet and work with brilliant people, who guided and inspired me. Let me start from my first "teacher" Laura C., who has been fantastic in transmitting her passion and knowledge, with an enormous patience. Even if I often teased her for being nerdy, I really appreciated her way of being a smart young scientist.

I would like to express my gratitude to Silvia, who gave me a method and taught me to be focused and strict.

My deepest acknowledgements go to professor Claudio Santoro and his group. In particular, I really wish to thank Diego, for his precious scientific advice and teachings. A sweet thank-you goes to Laura P., who has been the perfect unexpected phage display-colleague and room-mate. We had fun and we learnt a lot together.

I would also like to acknowledge all the collaborators of the TransSINE network for thought-provoking and scientific discussion.

I cannot deny my biggest thanks to Lavinia, Alice and Aurora. They have been wonderful Ph.D. colleagues, and, more importantly, special friends. We will split soon, but I guess only in a physical sense!

Words are not enough to say thank-you to Carlotta and Chiara, who have been two little angels tidying up the mess in my head…and even on my bench! I tried to do my best with them, but the truth is that maybe they taught me much more than I did.

For being always there, with the right smile (Sara), the right lab-tip (Marta C. and Maria) and the perfect hair and make-up (Marta M.), I would like to thank all the Gustincich Lab Girls.

Unexpectedly, also the male gender is progressively populating the laboratory, so I must specify that my acknowledgements are dedicated to all Gustincich Lab Guys too!

I do not want to forget the former Ph.D. students and post-docs (Stefania, Federica, Christina & Co) and all the people who passed by the Neurogenomics Lab (Roberto, Matteo, Dave & Co), as well as the seniors who have become honorary members (Paolo & Francesca). Globally, what I built throughout these years, and not only in

terms of scientific production, is the final big result of having met you-all.
Eventually, I would like to thank Jacopo, my family and my friends, for their continuous support and for their love.