RNA Biology 12:8, 771–779; August 2015; © 2015 Taylor & Francis Group, LLC

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

S Zucchelli[1,2], D Cotella[2], H Takahashi[3], C Carrieri[1], L Cimatti[1], F Fasolo[1], MH Jones[4], D Sblattero[2], R Sanges[5], C Santoro[2], F Persichetti[2], P Carninci[3], and S Gustincich[1,*]

[1]Area of Neuroscience; SISSA; Trieste, Italy; [2]Dipartimento di Scienze della Salute; Universita' del Piemonte Orientale; Novara, Italy; [3]RIKEN Center for Life Science Technologies; Division of Genomic Technologies; Yokohama, Japan; [4]Cell Guidance Systems; Cambridge, UK; [5]Animal Physiology and Evolution Laboratory; Stazione Zoologica Anton Dorhn; Napoli, Italy

Over the past 10 years, it has emerged that pervasive transcription in mammalian genomes has a tremendous impact on several biological functions. Most of transcribed RNAs are lncRNAs and repetitive elements. In this review, we will detail the discovery of a new functional class of natural and synthetic antisense lncRNAs that stimulate translation of sense mRNAs. These molecules have been named SINEUPs since their function requires the activity of an embedded inverted SINEB2 sequence to UP-regulate translation. Natural SINEUPs suggest that embedded Transposable Elements may represent functional domains in long non-coding RNAs. Synthetic SINEUPs may be designed by targeting the antisense sequence to the mRNA of choice representing the first scalable tool to increase protein synthesis of potentially any gene of interest. We will discuss potential applications of SINEUP technology in the field of molecular biology experiments, in protein manufacturing as well as in therapy of haploinsufficiencies.

Keywords: antisense transcript, embedded repetitive elements, long non-coding RNA, protein production, RNA therapeutics, SINEUP, translation

*Correspondence to: S Gustincich; Email: gustinci@sissa.it

Submitted: 04/02/2015

Revised: 06/03/2015

Accepted: 06/04/2015

http://dx.doi.org/10.1080/15476286.2015.1060395

Introduction

Large genomic projects such as ENCODE[1] and FANTOM[2] have shown that the majority of the mammalian genome is transcribed, thus generating a previously underestimated complexity in gene regulatory networks. A vast repertoire of different classes of transcripts includes protein-encoding mRNAs, non-coding RNAs of different size and RNAs of Transposable Elements (TEs), such as LINE (long interspersed nuclear element) and SINE (short interspersed nuclear element).[3-6]

Protein encoding genes present a large diversity of alternative Transcription Start Sites (TSSs) that may drive transcription in a cell type-specific manner.[7] Different 5′UTRs may contain information for mRNA sorting to subcellular compartments as well as for stimulus-dependent translation.

In addition to these 25000 genes encoding for proteins, long non-coding RNA (lncRNA) genes seem to represent the majority of cellular transcriptional output. In this context, LNCipedia v3.1 (2015) database contains >90000 human annotated lncRNAs, with many gene *loci* generating multiple transcripts.[8,9] LncRNAs exhibit a structure and biogenesis not dissimilar from mRNAs and their expression is under exquisite control in time and space. By definition, lncRNAs lack information to code for proteins. However, a threshold is set for open reading frames shorter than 100 aminoacids leaving room for coding potential of short peptides. They can be polyadenylated or polyA$^-$ and may operate in either nuclear and/or cytoplasmic fractions. Based on their genomic location, lncRNAs can be intergenic (referred to as long intergenic non-coding RNAs or lincRNAs) or may intersect other genes in exonic, intronic or fully overlapping configuration.

Representative examples of lncRNAs seem to be organized according to modules that bind to different protein complexes functioning as molecular scaffolds. lncRNAs can also take advantage of the selectivity of their RNA-DNA and