Functional exploration of antisense long non-coding RNAs containing transposable elements: a bioinformatics approach

A thesis submitted to the Open University of London for the degree of

Doctor of Philosophy

by

Veerendra P Gadekar

Master of Science in Bioinformatics, Manipal University,

Manipal- Karnataka, India

Stazione Zoologica Anton Dhorn, Naples, Italy

The Open University, London, United Kingdom

Director of studies: Dr. Remo Sanges, Ph.D.

External Supervisor: Dr. Stefano Gustincich, Ph.D.

Co-supervisor: Dr. Paolo Sordino, Ph.D.

January, 2016

Abstract

Long non-coding RNA (lncRNAs) show a wide range of regulatory functions at the transcriptional and post-transcriptional levels both in the nucleus and cytoplasm. Recently, antisense lncRNAs (ASlncRNAs) were reported to up-regulate protein synthesis post-transcriptionally through a mechanism depending on an embedded inverted SINE B2 and 5' overlap to the target mRNAs. Such ASlncRNAs are also referred as SINEUPs. Synthetic SINEUPs with identical modular organization were also demonstrated to exert the same activity suggesting a functional relationship between SINE repetitive elements and ASlncRNAs. In order to gain a broader insight on the contribution of transposable elements (TEs) in the sequence composition of ASlncRNAs, I have developed a bioinformatic pipeline that can identify and characterize transcripts containing TEs and analyze TEs coverage for different classes of coding/non-coding sense/antisense (S/AS) pairs. I aimed at identifying if the functional activity of SINEUPs could be a widespread phenomenon across multiple similar natural ASlnRNAs in the transcriptomes of the extensively studied model organisms that have a well annotated catalog of lncRNAs. From my initial analysis I identified human and mouse are the two species that showed a significant coverage enrichment of SINE repeats among ASlncRNAs. I further performed several functional enrichment analysis for the sense coding genes overlapping to ASlncRNAs taking into consideration of different characteristics of the 5' binding domain and the 3' embedded SINE repetitive elements. This permitted me to identify the effect of these modular features over the functional associations of sense coding genes. The results of the analysis showed that the products of coding genes associated to ASlncRNAs containing SINEs are significantly enriched for mitochondrial localization. Further, to determine if these ASlncRNAs could exert SINEUP-like activity during stress, I analyzed the

data from a published custom microarray experiment study, that were associated to the polysome fractions of MRC5 cell lysates in control and oxidative stress condition. The results revealed that the ASlncRNA carrying inverted or direct SINE repeats and their corresponding sense coding genes do not show any significant differential polysome loading in stress with respect to normal conditions, which is not a desired characteristic of a potential SINEUP. However, ASlncRNAs with inverted and direct SINE repeats corresponding to high translating polysome fractions showed a significantly higher ratio of means for RNA levels in stress over control, in contrast to noASlncRNA. This suggests that the ASlncRNA containing SINE elements are the key RNA molecules that are active during stress, although to determine if they are also involved in the increased polysome loading of their respective sense coding mRNAs, there is a need of further experimentation and exploration. Altogether, the work presented in this thesis provides a novel bioinformatics approach to study transcriptome-wide ASlncRNAs containing TEs and their functional association over the sense coding genes, and discover new significant functional features of ASlncRNA to be biologically validated.