BITS :: Call for Abstracts 2022 - Oral communication

Type	Oral communication
Session	Molecular Evolution analysis
Title	Identification of RNA modules in human lncRNAs
All Authors	Francesco Ballesio(1), Gerardo Pepe(1), Gabriele Ausiello(1), Pier Federico Gherardini(1), Manuela Helmer-Citterich(1)

Affiliation

(1) Dept Biology, University of Rome Tor Vergata, Italy

Motivation

Eukaryotic genes are generally composed of exons and introns. After transcription, they are processed and the messenger RNA is formed as the union of the exonic sequences. Eukaryotic proteins are often composed of domains, that are modularly associated in different numbers and types. It is well known that a strong association is present between exon and domain sequences, which is also the basis for the exonic theory for the origin of eukaryotic genes. In this project we aim to investigate whether long non-coding RNA genes possess a similar modular organization of their exonic structure. Such an architecture may also suggest the existence of independent functional units in lncRNAs. Since the path and pressure of gene evolution should be similar for coding and non-coding genes, we want to identify lncRNA exons corresponding to potential RNA functional units, that could be repeated in different number and type in the non-coding transcriptome, as shown in Fig 1.

Methods

We performed an all vs all comparison between all known lncRNA exons, with the aim of identifying highly similar pairs embedded in a dissimilar sequence context that may constitute independent modules in the architecture of lncRNA genes. The comparison was used to define candidate exon modules and leveraged both sequence and secondary structure global alignments, using the Needleman–Wunsch algorithm for the former and the BEAGLE algorithm for the latter. BEAGLE is a structural alignment method developed in our laboratory based on the BEAR alphabet for RNA secondary structure representation. We also calculated a conservation score for candidate exon modules by using the length normalized BLAST bit score from their alignment against the genomes of 4 primates closely related to humans. To assess the level of interspecific conservation we also investigated whether candidate exon modules are located in syntenic regions across the same primate species. Additionally we calculated the frequency of SNPs in these regions using the data from the ALFA allele frequency aggregator. Finally, since one of the hypothesized mechanisms for module shuffling is transposon-mediated exon shuffling, we looked for traces of these events using the hidden markov model of transposase domains obtained from Pfam.

Results

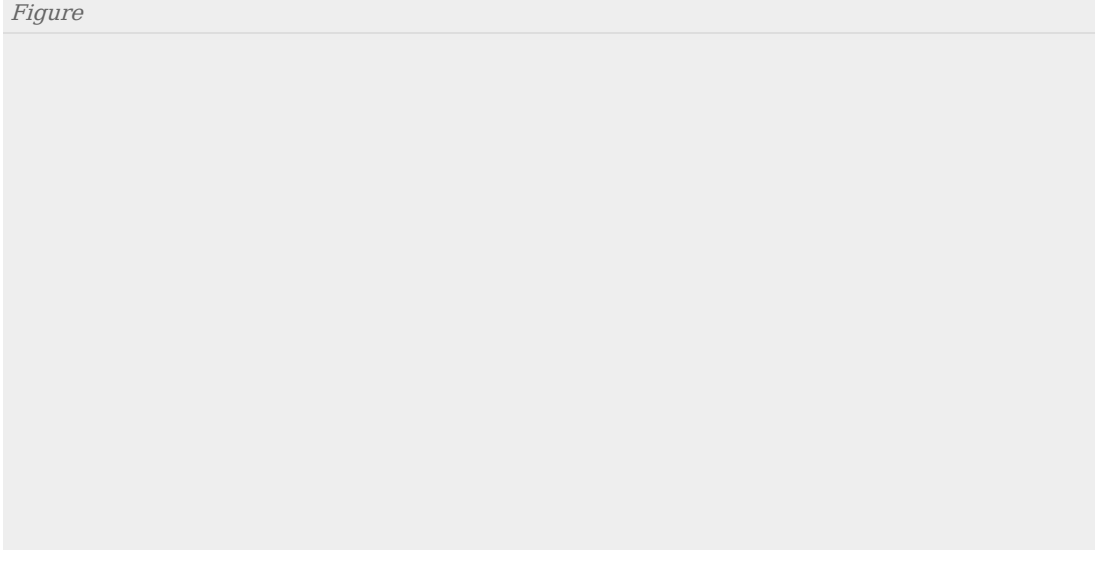
We identified ~157 clusters of RNA modules that share high sequence and/or secondary structure similarity, while being embedded in a dissimilar sequence context. Evolutionary conservation highlights a potential functional role for these exons. In particular we observed a higher conservation score in the exons that we identified as candidate modules, a higher degree of synteny in their genomic location and a lower frequency of SNPs compared to control sequences. Furthermore, we observed a significantly higher presence of transposase domains in the genes containing candidate exon modules, which might indicate a possible mechanism for shuffling. Overall these results indicate that lncRNA genes may have a similar modular organization as protein coding genes, and suggest a potential mechanism for the evolution of this architecture. These observations could also impact the way we define and annotate human non-coding genes.

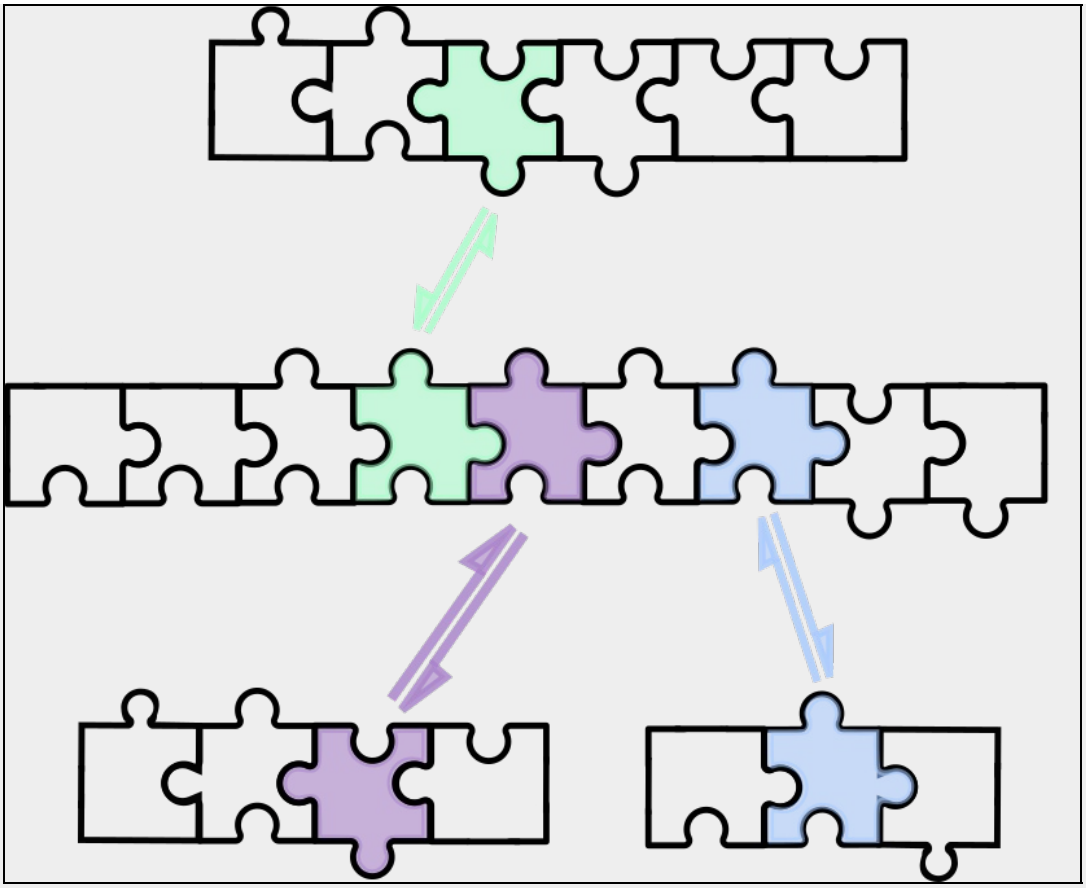
Info

-

filename	fig_BITS2022.png

Figure





Availability	-

Corresponding Author

Name, Surname	Pier Federico, Gherardini
Email	federico.gherardini@gmail.com

Submitted on	11.05.2022