

— Keynote Address —

RNAs Everywhere: Genome-Wide Annotation of Structured RNAs

Ivo L Hofacker

ivo@tbi.univie.ac.at

Institute for Theoretical Chemistry, University of Vienna, Währingerstr. 17, 1090 Vienna Austria

Keywords: non-coding RNA, RNA secondary structure, genome annotation

1 Introduction

Starting with the discovery of microRNAs and the advent of genome-wide transcriptomics, non-coding RNA transcripts have moved from a fringe topic to a central field of research in molecular biology. However, in contrast to the fairly reliable and complete annotation of protein coding genes in the human genome, comprehensive information on non-coding RNAs is still lacking. This is in spite of the fact that RNA is computationally easy to handle on the level of secondary structures and thus provides a fruitful field for bioinformatical studies.

2 Method and Results

Many functional RNA molecules depend upon on a characteristic secondary structure that is strongly conserved in evolution. This is exploited by algorithms that predict the consensus structure for a set of related sequences to obtain highly accurate secondary structure predictions [1]. Conversely, since conservation implies function, consensus structure prediction can be used to distinguish functional RNA structures from random ones.

Recent tools for the detection of structural non-coding RNAs, such as our programs AlifoldZ [2] and RNaz [4], evaluate sequences for signatures of structural conservation and exceptional thermodynamic stability, and are fast and accurate enough for genome-wide ncRNA surveys. As an example I will present results from a recent comparative screen of human and other vertebrate genomes [3], where we find more than 30000 structured RNA elements, almost 1000 of which are conserved across all vertebrates. Roughly one third of the candidates is found in introns of known genes, a sixth are potential regulatory elements in UTRs of protein-coding mRNAs, and about half are located far away of any known gene, see figure1. The widespread conservation of secondary structure points to a large number of functional ncRNAs and cis-acting mRNA structures in the human genome. In contrast, the number of detectable ncRNAs in invertebrate genomes, such as urochordates and nematodes, appear to be at least one order of magnitude smaller.

Only a small fraction of the predicted ncRNA has been described previously or can be assigned to any of the known ncRNA families on the basis of sequence similarity. Consequently, the possible function of most ncRNAs remain unknown. For ncRNA classes with a sufficient number of known examples, such as miRNAs and snoRNAs, specialized classifiers based on structure, sequence, and conservation patterns can be used to identify novel members and subfamilies. In some cases novel RNA

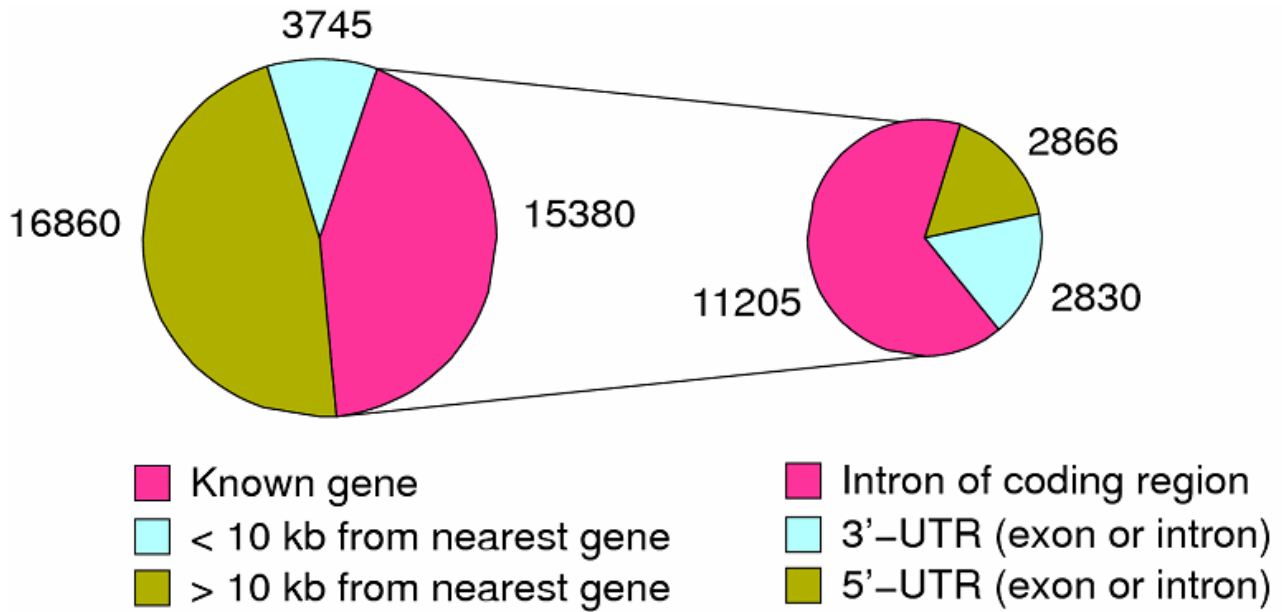


Figure 1: Genomic location of ncRNAs predicted by RNAz in vertebrate genomes [3].

families can be obtained from clustering ncRNA candidates based on structural similarity. Yet another avenue is to assign function by predicting possible targets for RNA-RNA interactions, e.g. between ncRNA and mRNA.

Despite these recent advances reliable functional annotation of non-coding RNAs remains an elusive goal beyond the few cases that can be handled by sequence homology with known ncRNA genes.

References

- [1] Hofacker, I.L., Fekete, M., Stadler, P.F., Secondary structure prediction for aligned RNA sequences. *J. Mol. Biol.*, 319:1059–1066, 2002.
- [2] Washietl, S., Hofacker, I.L., Consensus folding of aligned sequences as a new measure for the detection of functional RNAs by comparative genomics, *J. Mol. Biol.*, 342:19–39, 2004.
- [3] Washietl, S., Hofacker, I.L., Lukasser, M., Hüttenhofer, A., Stadler, P.F., Mapping of conserved RNA secondary structures predicts thousands of functional non-coding RNAs in the human genome, *Nat. Biotechnol.*, 23:1383–1390, 2005.
- [4] Washietl, S., Hofacker, I.L., Stadler, P.F., Fast and reliable prediction of noncoding RNAs, *Proc. Natl. Acad. Sci. USA*, 102:2454–2459, 2005.