Expanding the SnoRNA Interaction Network - Conservation of Guiding Function in Vertebrates

Stephanie Kehr

Abstract

Small nucleolar RNAs (snoRNAs) are one of the most abundant and evolutionary ancient group of small non-coding RNAs. Their main function is to target chemical modifications of ribosomal RNAs (rRNAs) and small nuclear (snRNAs). They fall into two classes, box C/D snoRNAs and box H/ACA snoRNAs, which are clearly distinguished by conserved sequence motifs and the type of modification that they govern. The box H/ACA snoRNAs are responsible for targeting pseudouridylation sites and the box C/D snoRNAs for directing 2'-O-methylation of ribonucleotides. A subclass that localize to the Cajal bodies, termed scaRNAs, are responsible for methylation and pseudouridylation of snRNAs. In addition an amazing diversity of non-canonical functions of individual snoRNAs arose. The modification patterns in rRNAs and snRNAs are retained during evolution making it even possible to project them from yeast onto human. The stringent conservation of modification sites and the slow evolution of rRNAs and snRNAs contradicts the rapid evolution of snoRNA sequences.

Recent studies that incorporate high-throughput sequencing experiments still identify undetected snoRNAs even in well studied organisms as human. The snoRNAbase, which has been the standard database for human snoRNAs has not been updated since 2006 and misses these new data. Along with the lack of a centralized data collection across species, which incorporates also snoRNA class specific characteristics the need to integrate distributed data from literature and databases into a comprehensive snoRNA set arose. Although several snoRNA studies included *pro forma* target predictions in individual species and more and more studies focus on non-canonical functions of sub-classes a systematic survey on the guiding function and especially functional homologies of snoRNAs was not available.

To establish a sound set of snoRNAs a computational snoRNA annotation pipeline, named `snoStrip` that identifies homologous snoRNAs in related species was employed. For large scale investigation of the snoRNA function, state-of-the-art target predictions were performed with our software `RNAsnoop` and `PLEXY`. Further, a new measure the Interaction Conservation Index (ICI) was developed to evaluate the conservation of snoRNA function.

The `snoStrip` pipeline was applied to vertebrate species, where the genome sequence

has been available. In addition, it was used in several ncRNA annotation studies (48 avian, spotted gar) of newly assembled genomes to contribute the snoRNA genes. Detailed target analysis of the new vertebrate snoRNA set revealed that in general functions of homologous snoRNAs are evolutionarily stable, thus, members of the same snoRNA family guide equivalent modifications. The conservation of snoRNA sequences is high at target binding regions while the remaining sequence varies significantly. In addition to elucidating principles of correlated evolution it was possible, with the help of the ICI measure, to assign functions to previously orphan snoRNAs and to associate snoRNAs as partners to known but so far unexplained chemical modifications. As further pattern redundant guiding became apparent. For many modification sites more than one snoRNA encodes the appropriate antisense element (ASE), which could ensure constant modification through snoRNAs that have different expression patterns. Furthermore, predictions of snoRNA functions in conjunction with sequence conservation could identify distant homologies. Due to the high overall entropy of snoRNA sequences, such relationships are hard to detect by means of sequence homology search methods alone.

The snoRNA interaction network was further expanded through novel snoRNAs that were detected in data from high-throughput experiments in human and mouse. Through subsequent target analysis the new snoRNAs could immediately explain known modifications that had no appropriate snoRNA guide assigned before. In a further study a full catalog of expressed snoRNAs in human was provided. Beside canonical snoRNAs also recent findings like AluACAs, sno-lncRNAs and extraordinary short SNORD-like transcripts were taken into account. Again the target analysis workflow identified undetected connections between snoRNA guides and modifications. Especially some species/clade specific interactions of SNORD-like genes emerged that seem to act as *bona fide* snoRNA guides for rRNA and snRNA modifications. For all high confident new snoRNA genes identified during this work official gene names were requested from the HUGO Gene Nomenclature Committee (HGNC)[1] avoiding further naming confusion.

[1] http://www.genenames.org/