In silico discovery of de novo structured RNAs in genomic and transcriptomic sequence

Jan Gorodkin[1]

[1] Center for non-coding RNA in Technology and Health, IKVH, University of Copenhagen

Abstract. Given that the protein coding sequence of the mammalian genome takes up only about a percent of the total genomic sequence, the mammalian genome holds great potential for harboring numerous non-coding RNAs (ncRNAs). In addition, the recent ENCODE project has revealed that most of the human genome is transcribed. Many ncRNAs have a profound secondary structure which is essential for their function and while this makes in silico screen for structured RNAs possible, it is also what makes the screen relatively computational expensive. However, to conduct such screens it is necessary to make use of corresponding (orthologous) sequences from multiple genomic organisms. Results from computational search strategies on genomic sequence including methodologies that re-align existing sequence based alignments will be presented along with strategies for downstream analysis of transcriptomic sequences, taking the profiles from mapped reads of small RNAseq data into account. The strategies mentioned can also lead to the discovery of differential processing patterns, such as microRNA arm switching.