

Genome-wide analysis of the sequence region surrounding the transcription start site of human mRNAs - (session: Comparative Genomics and Molecular Evolution)

Michele Iacono, Flavio Mignone and Graziano Pesole

Università degli Studi di Milano

Gene expression is finely regulated at both the transcriptional and post-transcriptional level. Transcriptional control is mediated by transcription factors, RNA polymerase and a series of cis-acting elements located in the DNA. Most important cis-elements are located within the Core Promoter Region, in close proximity to the transcription start site (TSS), and are classified as upstream or downstream promoter elements according to their position with respect to the TSS.

One of the main problems in studying regulation of gene expression is the identification of the motifs that have transcriptional meaning and are functionally important in transcriptional regulation, and the genes each motif regulates.

The recent availability of the draft human genome draft sequence as well as of a very large number of full length transcript sequences now make it possible to carry out an extensive and systematic comparative study of the genomic context of the TSS.

We present here a comprehensive sequence analysis, performed on a Human Core Promoter (HCP) dataset including 3140 sequences. HCP sequences have been extracted from human genome assembly (Release n°30) based on TSS mapping determined after comparison with the DBTSS and Refseq collections of reference transcripts. Putative cis-elements involved in transcription regulation have been identified through the application of pattern discovery algorithms and their presence in orthologous mouse genes has also been investigated.