In silico search for natural antisense transcripts in human genome and analysis of their expression patterns.

To be presented by Mikhail SKOBLOV

[1]*Medical Genetics Research Center, Russian Academy of Medical Sciences, Moscow, Russia,*
mskoblov@generesearch.ru

Both mRNA expression in a eukaryotic cell and efficiency of its translation into proteins are controlled by many regulatory levels subsequent to transcription initiation. As mRNA is a single strand molecule, the expression of a complementary antisense strand may alter transcription, elongation, processing, stability, and translation of the template RNA. Functional antisense RNAs have been identified in bacteria, but later were shown to be involved in gene regulation and differentiation in several eukaryotic organisms, including mammals. Natural antisense transcripts (NAT) usually arise via independent transcription initiation on the opposite DNA strand at the same genomic locus as the sense strand. Computational analysis of data from large-scale sequencing projects has revealed a surprising abundance of antisense transcripts in several eukaryotic genomes. As some antisense transcripts have been shown to regulate gene expression, it is possible that antisense transcription might be a common mechanism of regulating gene expression in eukaryotic cells.

We created an algorithm that allows high-throughput mapping of NATs. We used exact coordinates of transcripts and their orientations on the plus/minus chains of the human genome archived at NCBI server (NCBI http://www.ncbi.nlm.nih.gov/). In-house software "Antisense Searcher" was written on C++ and SQL. This program fulfills following tasks: 1) forming EST and mRNA transcripts in clusters on every chain of DNA; 2) retrieving all overlapping pairs of transcripts that are located on different DNA strands with more than 20 nucleotide overlaps; 3) retrieving an intersection of two previous sequence sets. EST clusters that contain only 1 or two ESTs were filtered at the subsequent stage of analysis. By this method we mapped approximately 13,500 NATs.

To study expression patterns of natural antisense pairs we created C++ -based software "Antisense Cluster Filter". This software allowed us to retrieve tissue expression field for all the transcripts from the lists of NATs. We used cDNA library descriptions

available from CGAP website (CGAP http://cgap.nci.nih.gov/) and other sources. By that, our data describing NATs were updated by information of pattern expression of transcripts. NATs were sorted by two criteria: 1) prevalence of expression in tumor or in normal cells 2) tissue specificity. In both cases we found 108 NATs in which only one member of a pair expresses only in tumor cells or in specific tissue. These pairs are currently studied experimentally.

In particular, we experimentally characterized an antisense mRNA *asAFAP* overlapping human *AFAP1* gene that encodes for an actin filament binding protein, which serves as a modificator of actin filament structure and integrity and relay a signal from receptor tyrosine kinases through PKCα to Src protein kinase. To study the intriguing phenomenon of tumor-specific *asAFAP* antisense expression we performed detailed *in silico* analysis of *asAFAP* sequences and experimentally quantified this transcript in normal and tumor human tissues. We also studied an antisense mRNA *asLZK* overlapping human *MAP3K13*/*LZK* gene that is involved in mitogenesis related JNK/SAPK signal transduction pathway. According to the functional annotation of the human genome, *asLZK* transcript (LOC647276) is expressed at the relatively high level and overrepresented in tumor samples. To our surprise, experimental study of human *asLZK* revealed that this sequence is not expressed, but represents a silent pseudogene of ribosomal protein L4 encoding gene *RPL4*. This pseudogene resulted from relatively recent retroposition of *RPL4* mRNA into the first inton of *MAP3K13* gene and does not participate in the regulation of *MAP3K13* expression. This study stresses that, after initial *in silico* mapping efforts, experimental verification of the expression landscape is warranted.