Poster L-17

Global analysis of antisense transcription in intronic regions of the human genome

Authors:
Helder I. Nakaya (*Departamento de Bioquímica, Instituto de Química, USP*)
Rodrigo Louro (*Departamento de Bioquímica, Instituto de Química, USP*)
André Lopes (*Departamento de Bioquímica, Instituto de Química, USP*)
Tarik El-Jundi (*Departamento de Bioquímica, Instituto de Química, USP*)
Paulo P. Amaral (*Departamento de Bioquímica, Instituto de Química, USP*)
Angela Fachel (*Departamento de Bioquímica, Instituto de Química, USP*)
Aline M. da Silva (*Departamento de Bioquímica, Instituto de Química, USP*)
Eduardo M. Reis (*Departamento de Bioquímica, Instituto de Química, USP*)
Sergio Verjovski-Almeida (*Departamento de Bioquímica, Instituto de Química, USP*)

Short Abstract: Mapping of all EST sequences to the human genome revealed that 86% of all spliced RefSeq-genes have transcriptionally active intronic regions. These messages were non-coding long unspliced transcripts, and were preferentially found in 5' introns of RefSeq-genes. A tissue signature of antisense intronic transcription was revealed using a custom oligoarray.

Long Abstract:
Here we report the in silico identification of a large set of intronic transcriptionally active regions throughout the human genome, when a detailed genome mapping analysis of 184,451 complete and partial mRNAs (22,458 RefSeq and 161,993 mRNAs), as well as nearly 5 million human EST sequences was performed. Interestingly, we found that 86 % of all human RefSeq genes have transcriptionally active intronic regions. 1,073 sites of active intronic transcription are already defined by well-characterized RefSeq transcripts, and 67% of these are expressed antisense to the RefSeq transcript encoded in the opposite strand. In addition, we found approximately 55,000 totally intronic (TIs) contigs, each one assembled by merging all ESTs that map to a given intronic locus. TI messages were found to be predominantly non-coding long unspliced transcripts, 3.5 times longer than the average size of exons of protein-coding genes. TIs were preferentially found in 5´ introns, and its presence was associated to genes with a higher number of alternative splicing isoforms. We also identified about 5,000 antisense partially intronic transcripts with overlap to exons (TOEs) of well-defined genes encoded in the opposite sense strand. To provide independent evidence for genome-wide intronic transcription in the human genome, a custom Agilent oligoarray was designed containing ~42,000 probes that allows the selective detection of sense and antisense transcripts from TI and TOE messages, as well as from a sense exonic counterpart from the same gene. Expression profiles obtained from three different human tissues, namely liver, prostate and kidney confirmed the ubiquitous expression of antisense intronic messages and showed a tissue signature of antisense intronic transcription. Interestingly, in each tissue the Gene Ontology category "Regulation of transcription" was found to be significantly enriched among the protein coding genes where the top 30 % most highly expressed antisense TIs map. Also, comparison of normal and tumor renal tissue revealed a number of antisense TIs that are differentially expressed in renal cancer, and possibly

modulate gene expression by affecting the abundance and/or the usage of alternative exons in cognate protein coding transcripts. Taken together these results provide compelling evidence that long intronic antisense messages represent a new class of RNA originated from widespread intronic transcription activity in the human genome, which may play a fundamental role in the control of gene expression.