Detection and analysis of spliced chimeric mRNAs in sequence databanks - (session: Novel Algorithms for Bioinformatics)

Antonello Romani, Marco Trerotola, Emanuela Guerra, Andrew Emerson, Elda Rossi, Agnieszka Bronowska and Saverio Alberti

Mario Negri Sud

A databank screening procedure, the In Silico Trans-splicing Retrieval System (ISTReS), was developed to identify chimeric mRNAs originating from chromosomal translocations, mRNA trans-splicing and multi-locus transcription. A parsing algorithm to screen cDNA-vs-genome Blast outputs was implemented. Key filtering criteria were Blast scores of >= 300, match lengths of >= 95% of the query sequences, junction of the two partners at exon-exon borders and concordant Œsense / sense, reading orientation. ISTReS was validated by the successful identification of bona fide chromosomal translocation-derived fusion transcripts in the HGI and RefSeq databanks. The performance of ISTReS was verified against recently identified chimeric antisense transcripts. Analysis of the UNIGENE database revealed 21742 chimeric sequences overall, that correspond to ~ 1% of the database transcripts. Novel FOP-Rho GAP and methionyl tRNA synthetase-advillin chimeric mRNAs with the canonical features of trans-spliced-transcripts were identified among 246 chimeras from the RefSeq databank. This suggests a frequency of canonically-spliced chimeras of approximately 1% of all the hybrid sequences in current databanks. These findings demonstrate the efficiency of ISTReS and the overall feasibility of sequence/structure-based strategies to search for chimeric mRNAs candidate to derive from the splicing of heterologous transcripts.