**RECOGNITION OF THE SUBSET OF EUKARYOTIC PROMOTERS.**

KONDRAKHIN Y.V., OVERTON G.C.

Center for Bioinformatics, University of Pennsylvania, Philadelphia, PA, USA.

Keywords: recognition, eucaryotic promoters, transcription factors, TBP, Sp1, NF-Y, GATA, TATA-box

Effective recognition of eukaryotic promoters within large regions of unspecified nucleotide sequences is still far from complete.

Recently conducted analysis (J.W.Fickett & A.G.Hartzigeorgiou, Eukaryotic promoter recognition) have indicated the low accuracy of currently available computer tools. Because of the whole set of eukaryotic promoter sequences represents the heterogeneous mixture of different promoter subfamilies we prepose new procedure for recognizing subset of promoters with essentially low rate of false positives. This procedure is based on the local properties of distributions of several individual signals as well as signal pairs.

The main part of considered signals represent the binding sites of well-known transcription factors (proteins) such as TBP, Sp1, NF-Y, GATA. In particular, we extracted the new subclass of TATA-box and discovered the extreme non-uniformity of Sp-1 location upstream the start of transcription. Additionally we took into consideration several unknown signals which were found in the nearest neighbourhood of the start of transcription. New signals as well as all revealed properties were discovered on the base of analysis of promoter sequence data from EPD data base while the accuracy of proposed recognition procedure were estimated with the help of modified set of complete gene sequences