International Indexed,Referred Research Journal, ISSN-2250-2629, January,2013

Research Paper - Bioinformatics

Prediction of Splice Sites in Drosophila melanogaster

*** Ms. Deepshikha ** Dr. Manvender Singh**

*** Research Scholar, Singhania University (Rajasthan)**
****Assistant Professor, MDU, Rohtak**

A B S T R A C T

Splice sites are the key signal sequences that determine the boundaries of exons. A sequence of eight nucleotides is highly conserved at the 5'ss and a sequence of 4 nucleotides is highly conserved at the 3'ss, preceded by a pyrimidine-rich region.

Keywords :-Splicing, NNSPLICE, Donor, Acceptor

Introduction

Almost all of the nuclear genes coding for proteins in eukaryotes are split into coding (exon) and non-coding (intron) sequences. The intron sequences are precisely spliced out of the initial gene transcript before the mRNA is transported to the cytoplasm for translation. A sequence of eight nucleotides is highly conserved at the boundary between an exon and an intron, the 5' splice site (5'ss). The boundary between an intron and an exon, the 3' splice site (3'ss) also exhibits a highly conserved sequence of 4 nucleotides, preceded by a pyrimidine-rich region. These conserved sequences are an essential part of the process of exon splicing and provide a specific molecular signal for the RNA splicing machinery to identify the precise splice points.

Materials and methods

We extracted a dataset of 50 sequences from BDGP. Sequence set comprised of full insert sequences of the cDNA clones comprising the DGC. Splice sites were predicted using BDGP Splice site Predictor. This server runs the NNSPLICE 0.9 version of the splice site predictor. It considers genes that have constraint consensus splice sites, i.e., GT' for the 5' and AG' for the 3' splice site. Threshold was set as 0.4. The output of the network is a score between 0 and 1 for a potential splice site.

Results and Discussion

We analyze the structure of donor and acceptor sites using a separate neural network recognizer for each site. A backpropagation feedforward neural network with one layer of hidden units was trained to recognize donor and acceptor sites, respectively; using a novel optimized representative data set. Only those genes are considered that have constraint consensus splice sites, i.e., GT' for the donor and AG' for the acceptor site. The sites recognized by HSF have the following structure:

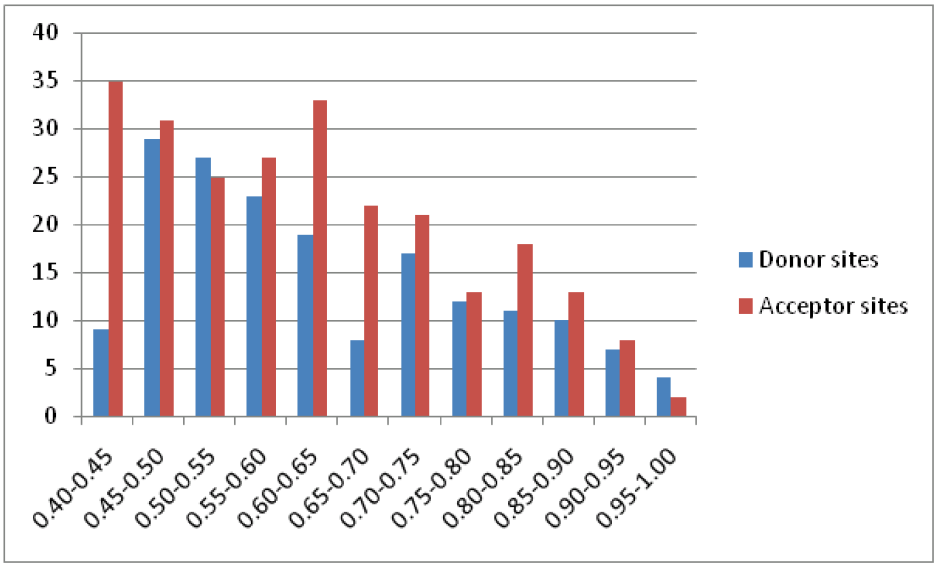
* Donor site: 7 bases in the exons and 8 bases in the intron, ie AAGCGAGgtaagcaa
* Acceptor site: 20 bases in the exons and 21 bases in the intron, ie tcatcttctgtcccttcccag AAAACC TACCAGGGCAGCTA

Splice motifs found in these sequences were retrieved, along with their scores. The scores generated by NNSPLICE range from 0 to 1. Score value of 1.0 corresponds to a perfect match with a consensus splice site. A threshold of 0.4 is used by NNSPLICE to filter results, so that scores are retrieved only if a motif with a score>0.4 has been found either in the wild-type or

Table1. Scores for predicted donor and acceptor sites at threshold 0.40

Score	0.40-0.45	0.45-0.50	0.50-0.55	0.55-0.60	0.60-0.65	0.65-0.70	0.70-0.75	0.75-0.80	0.80-0.85	0.85-0.90	0.90-0.95	0.95-1.00
Donor sites	09	29	27	23	19	08	17	12	11	10	07	04
Acceptor sites	35	31	25	27	33	22	21	13	18	13	08	02

Graph1. Distribution of predicted donor and acceptor sites in score intervals.



mutant sequence.

We have organized the retrieved data in form of a table (Table1). Maximum (29) donor sites have score values in between 0.45-0.50 and minimum (04) have score values in between 0.95-1.00. Maximum (35) acceptor sites have score values in between 0.40-0.45 and minimum (02) have score values in between 0.95-1.00. This data has been plotted in a graph (Graph1).

See Table 1

See Graph 1

Acknowledgements

We thank the Berkeley Drosophila Genome Project group for providing us with Drosophila genome datasets and splice site prediction server.

REFERENCE

1. Reese MG, Eeckman, FH, Kulp, D, Haussler, D, 1997. ``Improved Splice Site Detection in Genie".J Comp Biol 4(3), 31123.
2. Shapiro and Periannan Senapathy "RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implcations in gene expresson" NAR Volume 15 Number 17 1987 Websites
3. http://www.fruitfly.org/seq_tools/splice.html
4. http://fruitfly.org/data/sequence/ sequence_db/ na_cDNA.dros