Poster

A genome-wide in-silico motif scan to identify functional clustering

Atreyi Chakraborty, Madhusudhan M.S

Indian Institute of Science Education and Research, Pune
hakraborty.atreyi@students.iiserpune.ac.in

Interacting proteins usually bind 5-6 base pairs of DNA. We looked at distributions of 5/6-mer DNA motifs, in whole chromosomes and in smaller sections, across the whole genome to get insights into how proteins read genomic sequences. The distribution of motifs in the genome is non-random as established by their observed to expected (OE) ratios at all examined length scales. We correlated the motif distributions in promoter regions of genes to one another and found 5 chromosomes implicated in Robertsonian translocations to be strongly linked to one another. Correlations of OE ratios of motifs in gene promoters gave insights into gene regulation and function. We also showed how one could build a possible network of interactions considering 19 genes all transcribed by Jun/Fos. Further, comparing our findings with Hi-C data showed that in >75% of the cases, spatial proximity also implied a high motif correlation. In general, correlating genomic regions by motif distribution comparisons alone is rife with functional information.