

Automatic Annotation Tool and Browser for Whole-Genome Tiling-Array Data Analysis

Yoshiki Mochizuki, Katsura Hirosawa, Yoshikazu Hasegawa, Naohiko Heida and Tetsuro Toyoda

Knowledge Base Research Team, Genomic Sciences Center, RIKEN, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

URL and Availability: <http://omicspace.riken.jp>

Abstract

Tiling arrays of high-density oligonucleotide probes spanning the entire genome are powerful tools for the discovery of new genes. However, it is difficult to determine the structure of the spliced product of a structurally unknown gene from noisy array signals only. Here we introduce a statistical method that estimates the precise splicing points and the exon/intron structure of a structurally unknown gene by maximizing the likelihood of observed array-signal intensities and nucleic-acid sequences based on the combined model of a threshold-based intensity likelihood, a splice-point likelihood by a bi-directional Markov model and length likelihood of exons and introns. Our method predicted more accurately the gene structures than the simple threshold-based method, and more correctly estimated the expression values of structurally unknown genes than the window-based method. It was observed that the Markov model contributed to the precision of splice-points, and that the statistical significance of expression (P value) well represented the reliability of the estimated gene structure and expression value. We have implemented the method as a program ARTADE and applied it to the *Arabidopsis thaliana* whole-genome array data analysis. The predicted results are integrated and browsed through our original genome browser GPS. The ARTADE program and GPS are available at <http://omicspace.riken.jp>.