HL111

A Set of Structural Features Defines the Cis-Regulatory Modules of Antenna-Expressed Genes in Drosophila melanogaster.

Yosvany López 1, Alexis Vandenbon 2 and Kenta Nakai 1

1 The University of Tokyo, Japan

2 Osaka University, Japan

Abstract

Deciphering the biological information within regulatory regions has become one of the main focuses of current genomic research. It has been hypothesized that regulatory regions of co-expressed genes (or at least some of them) share a similar architecture. Our study has proved that genes expressed in the same tissue have indeed similar regulatory structures. It would be of obvious interest to the research community to be able to predict genes expressed at specific time points or under physiological conditions by analyzing their regulatory sequence. Previous studies have also shown dependence among motifs, but to the best of our knowledge, no studies have simultaneously examined multiple structural features such as positioning of *cis*-regulatory elements relative to transcription start sites or to each other, order and orientation of regulatory motifs to accurately describe overall *cis*-regulatory structure. This work has used large-scale data for uncovering novel *Drosophila melanogaster* transcription factor–binding motifs and several common structural features of these motifs in *cis*-regulatory modules of antenna-expressed genes. Six potential antenna-related motifs were initially predicted, including three that appeared to be novel. A huge feature set was created with the predicted motifs, a correlation-based filter was employed to remove those irrelevant features, and a genetic algorithm was finally designed to obtain the most informative features. As a result, a set of eight structural features was found in the regulatory regions of antenna-expressed genes. We subsequently used such features to score the *D. melanogaster* regulatory regions for potentially unknown antenna-expressed genes. Validation of our predictions with an independent RNA sequencing dataset showed that 76.7% of genes with high scoring regions were expressed in antenna. We also found that the identified features were highly conserved in regulatory regions of orthologs across the *Drosophila* lineage. Our method showed comparable performance to previous approaches, but uncovered additional interesting features because it considered the order and orientation of regulatory motifs.