Translating genetic variation to the expression level in the context of a sequence-level gene expression model of the gap gene system in Drosophila development.

V. V. Gursky, K. N. Kozlov, I. V. Kulakovsky, A. Zubair, P. Marjoram, D. Lawrie, M. G. Samsonova, S. V. Nuzhdin

The genotype-phenotype relation and its proper quantitative description provide fundamental biological questions. An important aspect of this relation is to understand the impact of natural genetic variation for gene expression. A necessary step towards addressing these questions concerns development of effective models of gene regulation that link the DNA sequence information with the gene expression level. In this study, we apply a sequence-level model of the gap gene expression in Drosophila, reported in an accompanying paper, to analyze single nucleotide polymorphisms (SNPs) in the data of 216 sequenced Drosophila melanogaster genotypes. We found that the binding sites for transcription factors regulating the gap gene network are essentially less polymorphic than the proximate sequence in the regulatory regions of these genes. By quantifying the regulatory influence of each SNP in the model, one demonstrates how the regulatory effects from multiple SNPs nonlinearly combine exhibiting clusterization. We report a small negative correlation between the SNP regulatory influence and its frequency in the population, indicating possible action of selection. In contrast, no significant correlation was found between the SNP regulatory influence and the binding affinity of corresponding polymorphic binding sites. Taken together, these results provide a system level view of how genetic variation translates to the gene expression level.