

Abstract

Gene expression arrays are used to assess expression of exons and genes of organisms. The design of expression arrays is based on a genome of laboratory strains of model organisms. The most frequent summarization algorithms used to process data from measurements are gcRMA, PLER and IterPLIER. When using expression arrays to research free living species, the measured values are influenced by differences in genomes of free living and model organisms. We propose a method to improve the results by removing parts of genomes influenced by known differences between species from the summarization. Removing influenced parts can improve summarization, especially on exon level.