

ABSTRACT:

Hybridization of cDNA to tiled genomic microarrays has allowed the empirical identification of all regions of genomic transcription for several genomes. Even though the *Drosophila* genome is one of the best annotated higher eukaryotic genomes, we have shown through tiling array experiments that a large number of transcripts have been missed using classic annotation strategies such as deep sequencing of cDNA libraries and computational predictions of genes. For example, we identified a large number of unannotated exons linked to known protein coding genes, in particular 5' exons representing alternative transcription start sites. In addition, by predicting and then overlapping the location of novel 5' exons with the location of known mutations that cause developmental defects or death, we were able to map mutations to the genes they affect. We are now employing this strategy to hunt for disease-causing mutations in humans. Turning to *Daphnia pulex*, a genome where annotations are less well defined, we have used the tiling array strategy to improve the annotation quality. This strategy has identified 76,000 TARs (Transcriptionally Active Regions) that confirm or correct gene models as well as 94,000 TARs that identify novel exons of known genes as well as new genes structures. The methodology for generating these data will be discussed.