

Abstract

The functional annotation and identification of genes involved in the development and progression of complex diseases is a cumbersome and non trivial task. Exploiting the comparisons of the human genome with other genomes at both the distal and proximal evolutionary edges of the vertebrate tree is expected to represent a powerful tool in the puzzle of decoding molecular mechanisms underlying development or disease.

The main objective of this thesis was to develop a comprehensive and efficient bioinformatics platform for large-scale transcriptomic studies. It facilitates comparative analyses of human diseases and corresponding mouse models by integrating gene expression data with genome sequence information.

The specific achievements of the systematic approach represented here are threefold: First, a set of representative transcriptomic datasets describing mouse embryo fibroblasts and human multipotent adipose-derived stem cells during adipocyte differentiation has been produced, annotated, as well as stored in an organized and easily accessible way within a microarray database management system. Second, sophisticated computational tools are provided within a bioinformatics platform for large-scale comparative transcriptomic analyses to distinguish the similar from the dissimilar and to analyze these data in a straightforward, efficient, and reliable way. Several methods are proposed to derive meaningful biological information and distributed high-performance computing is used to facilitate these types of large-scale data analyses in reasonable time. Third, comparative analyses of the human and mouse datasets described above have been conducted with contingent new insights into the universality as well as the specialization between the most important model organism mouse and the designation of all clinical research, the human.

Finally and ultimately these investigations attempt to provide the research community with a markedly improved repertoire of computational tools that facilitate the translation of accumulated information from comparative transcriptomic studies into novel biological insights.

Keywords: comparative genomics, transcriptomics, microarray, adipogenesis, transcriptional profiling, functional annotation, cluster analysis, bioinformatics, high-performance computing.