A Novel Approach to Gene Expression Analysis of Ethnicities

Naidu, Prathik

DNA and RNA microarrays are important tools to measure the gene expression levels in cells. However, these technologies have major limitations, such as their inability to detect genetic aberrations. A computational method was developed and implemented to analyze gene expression more efficiently than current technologies. This approach was applied to understand the gene expression in the American European, Great British, and Yoruba African ethnic groups. The computational method aligned the RNA reads to the reference genome, assembled the transcripts, and preformed the gene expression analysis. These steps were streamlined into a software package named DGEST, which expedites the discovery of differentially expressed genes using RNA-sequencing. Statistical tests, implemented in the R programming language, filtered the large gene set from the DGEST analysis into a smaller group of differentially expressed genes. The aligned reads and protein isoforms were then visualized in a genome browser and the Protein Atlas database. Initial results showed that 74 genes with the lowest q-value and the largest log(FPKM) expression value were the most differentially expressed. Protein modeling for the isoforms of these genes indicated differences primarily in cellular processes. The CCL4L1 gene was chosen for further analysis, and indicated higher expression in the Yoruba population, suggesting greater affinity for HIV inhibition. In addition, the computational method yielded 95% accuracy in the RNA read mapping, illustrating the efficacy of this approach for gene expression analysis. The results gleaned from this method can assist in the development of targeted cancer therapeutics and optimization of drug design for ethnicities.

Awards Won: