Application of EMDomics to Identify Age-Associated Expression and Treatments in Cancer

Malladi, Sadhika

Differential gene expression analysis is commonly used to associate gene expression with a biological or clinical subtype of patients. However, patients rarely divide into homogeneous groups, especially in a varied disease such as cancer. Current tools used for differential gene expression analysis, such as Significance Analysis of Microarrays (SAM), do not account for intra-class heterogeneity. To address this shortcoming, we developed EMDomics, a novel method that repurposes the common Earth Mover's Distance (EMD) used in image analysis for the purpose of genomic analyses. Specifically, EMDomics uses histograms of gene expression data to compute the significance of the observed differences in expression between two or more pre-defined classes of patients. We have published EMDomics on Bioconductor for use by other bioinformaticians. We test the robustness of EMDomics using several simulation experiments, and then we apply it to conduct the first known large-scale analysis to identify age-associated changes in expression in patients of 11 different cancers. Particularly, gliomas exhibited highly age-influenced transcriptomic behavior. Furthermore, we discovered biological pathways that may be enriched variably with age, particularly that the extracellular matrix was significantly age- driven in enrichment. Finally, we identified perturbagens that may be more effective in older or younger patients. Across the three cancers with the greatest transcriptomic disparities, we predicted that camptothecin-related compounds have greater efficacy in younger patients. Our findings support the viability of EMDomics and provide an understanding of age-influenced mechanisms in cancer patients.