# *Decision* tree-based method for integrating gene expression, demographic, and clinical data to determine disease endotypes.

Williams-DeVane CR[1], Reif DM, Hubal EC, Bushel PR, Hudgens EE, Gallagher JE, Edwards SW.

Author information

## Abstract

**BACKGROUND:** Complex diseases are often difficult to diagnose, treat and study due to the multi-factorial nature of the underlying etiology. Large data sets are now widely available that can be used to define novel, mechanistically distinct disease subtypes (endotypes) in a completely data-driven manner. However, significant challenges exist with regard to how to segregate individuals into suitable subtypes of the disease and understand the distinct biological mechanisms of each when the goal is to maximize the discovery potential of these data sets.

**RESULTS:** A multi-step **decision tree-based** method is described for defining endotypes based on gene expression, clinical covariates, and disease indicators using childhood asthma as a case study. We attempted to use alternative approaches such as the Student's t-test, single data domain clustering and the Modk-prototypes algorithm, which incorporates multiple data domains into a single analysis and none performed as well as the novel multi-step **decision** tree method. This new method gave the best segregation of asthmatics and non-asthmatics, and it provides easy access to all genes and clinical covariates that distinguish the groups.

**CONCLUSIONS:** The multi-step **decision** tree method described here will lead to better understanding of complex disease in general by allowing purely data-driven disease endotypes to facilitate the discovery of new mechanisms underlying these diseases. This application should be considered a complement to ongoing efforts to better define and diagnose known endotypes. When coupled with existing methods developed to determine the genetics of gene expression, these methods provide a mechanism for linking genetics and exposomics data and thereby accounting for both major determinants of disease.