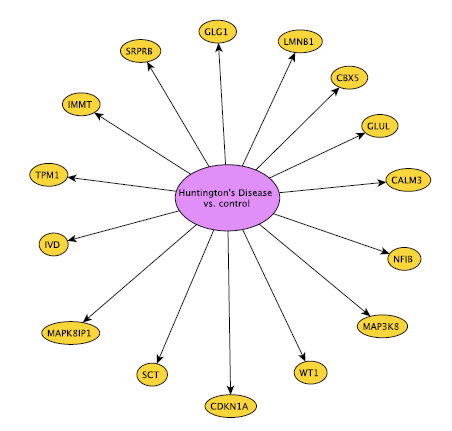


Previous research:

1. *An Automated Bayesian Framework for Integrative Gene Expression Analysis and Predictive Medicine*
   1. We are looking at GEO for gene expression data.
   2. Bayesian approach since it’s been shown to be good. We start with a *prior distribution* and we build a *posterior distribution* using the training data.
   3. In the multinet classifiers, the dataset is first partitioned by classes, and a single Bayesian network is constructed on each partition. Then the data is classified to the class that maximizes the posterior probability. Compared to a simple network, the multinet aims to more precisely model underlying pattern of dependency between features as the structure of the classifier is not forced to be static across classes. Therefore, the main difference for constructing singly structure Bayesian network and multinets is in treating the samples considered in the trial. For constructing multinets we keep the samples of each GEO experiment separate and we integrate the information content of each experiment through Bayesian inference step.
   4. Basically, automation of the whole process of integration GEO data in regard to **a single disease** although it could be any disease. Nothing to do with genes in common between diseases.
   5. Diseases processed: Huntington’s Disease, Obesity, Leukemia, and Lymphoma.
   6. Figure 3: Constructed Bayesian multinet showing the interactions among genes related to Huntington’s Disease and controls.



* 1. Figure 4: Constructed Bayesian single net showing the genes contributing to differentiating patients with Huntington’s disease from controls. 

1. *Genetic-Demographic Dissection of Alcoholism:*
   1. More than GEO: demographic, clinically ascertainable, and genetic factors
   2. Measure for predictive power : area under the receiver operating characteristic curve
2. *Predictive Probabilistic Analysis Reveals Multiscale Epistasis in Nicotine Dependence*
   1. Used multi‐level Bayesian framework

We want to employ Bayesian network built based on Gene Expression, Gene sequence, demographic info, etc. to find the common factors in between different mental diseases as well as different addictions and further, to find commonalities between mental diseases and addictions. We want to have a fully automated pipeline to feed as input GEO data, dbGaP, demographics, etc., and build Bayesian networks and compare the networks and as an output get the common factors between diseases. We have two similar pipelines already designed but they are not complete and particularly they lack the ability to compare Bayesian networks

* The duplications in survey information should be pruned out manually.
* We can use “knime”
* The current Bayesian network doesn’t have weights for edges; it’s good to add that feature.