Semi- supervised class discovery using quantitative phenotypes – CVD as a case study

Israel Steinfeld[1], Roy Navon[1], Ardigò D[2], Zavaroni I[2] , Zohar Yakhini[1]

1 Agilent Laboratories, Tel Aviv, Israel
[2] Department of Internal Medicine and Biomedical Sciences, Parma University, Italy

Abstract

Despite significant advances in treatment, cardiovascular disease (CVD) is a major source of morbidity in the western world. A fundamental goal in current research is the integration of genomic and clinical information to support understanding of CVD molecular mechanisms. In particular, underlying processes involved in atherosclerosis that lead to CVD events.

Genomic studies typically focus on comparing disease to healthy population. However – the association of genomic processes to variation in normal populations is also highly informative, especially for risk assessment.

In our work, healthy volunteers were screened and stratified for their cardiovascular risk based on clinical, biochemical and instrumental parameters, including IMT (Carotid Intima-Media Thickness) and FMD (Flow-Mediated Vasodilation), fundamental technologies in the field of CVD risk assessment. In addition, gene expression profiling was performed on peripheral blood mononuclear cells (PBMC) isolated from the patients, using oligonucleotide microarrays. Our data therefore contains rich quantitative annotation and expression profiles for all subjects.

Here we present a method that utilizes the availability of rich quantitative annotation of the samples. We develop methods for statistically assessing the differential expression in patterns based on quantitative phenotypes. We further present a semi-supervised class discovery, constraining the search space to patterns that respect an order induced by the measurement values. We show that our method is robust enough to detect known clinical parameters with accordance to expected values (e.g. high LDL level). We also use our method to elucidate putative risk factors, for example: identification of PBMC gene expression signature in heavy smoking patients.