Vitamin D receptor initiation codon polymorphism influences genetic susceptibility to type 1 diabetes mellitus in the Japanese population

ABSTRACT

Our research indicates that the vitamin D receptor initiation codon polymorphism is a factor in the genetic susceptibility of Japanese T1DM. This polypheny is also associated with GAD65-Ab-positive T1.

INTRODUCTION

This paper introduces the concept of 'tree harvesting', which is a general method for supervised learning from gene expression data. It begins with real-valued expression measurements taken across thousands of genes, spread across x samples (typically 50 or 100), but will increase its size in coming years. We have a set of outcomes for every sample, such as survival time or cancer class. Our aim is to understand the link between the genes and the outcome. The problem of forecasting an outcome measure from dozens of features is called supervised learning. If the result is quantitative, the term ‘regression’ is used; for categorized outcome, ’classification’. There are several supervised learning methods, including linear regression, discriminant analysis, neural networks, support vector machines, and boosting. However, they are unlikely to work 'off the shelf', as expression data presents unique challenges. The problem is that there are many inputs (genes) to which there is a high correlation when the number of samples is small. Hastie et al. have described 1-2 simple ways to solve this problem, with constructing progressively more ambitious models that take into account gene interactions. The first step in our approach is to cluster the genes based on hierarchical criteria, followed by analyzing the average expression profiles of each cluster in the dendrogram as inputs to our prediction model. This has two advantages: it simplifies the process of predicting global population distribution and increases the likelihood of predictions being made when multiple variables are used as models. The use of hierarchical clustering as a standard descriptive tool for expression data makes its components easy to interpret. Furthermore, by using clusters as inputs, we can bias the inputted datasets towards correlated sets of genes, thereby reducing the rate at which the model overfits. Rather than selecting individual samples, we prioritize larger clusters, as explained further down. The basic technique is explained in the following section for a quantitative output and squared error. Next, it is expanded to include other settings like survival data and qualitative responses. The method of tree harvesting is exemplified with two practical instances and a simulation model is presented to evaluate its effectiveness. Furthermore, we extend the analysis to nonlinear expression effects.

CONCLUSION

We have developed an algorithm for mapping the positively selected mutations of viral quasispecies using sequence data. This has also been used to map the positive-selected variants (i.e., influenza A HA, HIV-1 RT, and HIV–120) of the human immunodeficiency pathogen hepatitis C virus as well as the FMD virus. The most enlightening application of selection mapping is likely to be in the comparison of viral subpopulations under different selective pressures, such as selection mashing HIV isolates with various cellular tropisms for positive-selected mutations that are positively selected in response to the host cell type. Furthermore, we could use selection mapping to examine HIV breakthrough infections to see if the

vaccinations prevented the HIV quasispecies from living in normally favorable regions of the quasiexperimental sequence space. Our proposal is to include the positively selected viral variants in future vaccines that are highly multivalent and designed to compensate for B-cell-selected antigenic drift.