SVMRFE based approach for prediction of most discriminatory gene target for type II diabetes

Atul Kumar, D. Jeya Sundara Sharmila, Sachidanand Singh

PII: S2213-5960(16)30185-4
DOI: doi: 10.1016/j.gdata.2017.02.008
Reference: GDATA 644

To appear in: *Genomics Data*

Received date: 29 November 2016
Revised date: 7 February 2017
Accepted date: 15 February 2017

Please cite this article as: Atul Kumar, D. Jeya Sundara Sharmila, Sachidanand Singh , SVMRFE based approach for prediction of most discriminatory gene target for type II diabetes. The address for the corresponding author was captured as affiliation for all authors. Please check if appropriate. Gdata(2017), doi: 10.1016/j.gdata.2017.02.008

SVMRFE based Approach for Prediction of Most Discriminatory Gene Target for Type II Diabetes

Atul Kumar[1], D. Jeya Sundara Sharmila[2], Sachidanand Singh[1]
[1](Department of Biotechnology and Health Sciences, Karunya University, Coimbatore, Tamil Nadu, India)
[2](Department of Nanosciences and Technology, Tamil Nadu Agriculture University, Coimbatore, Tamil Nadu, India)

Abstract

Type II diabetes is a chronic condition that affects the way our body metabolizes sugar. The body's important source of fuel is now becoming a chronic disease all over the world. It is now very necessary to identify the new potential targets for the drugs which not only control the disease but also can treat it. Support vector machines are the classifier which has a potential to make a classification of the discriminatory genes and non-discriminatory genes. SVMRFE a modification of SVM ranks the genes based on their discriminatory power and eliminate the genes which are not involved in causing the disease. A gene regulatory network has been formed with the top ranked coding genes to identify their role in causing diabetes. To further validate the results pathway study was performed to identify the involvement of the coding genes in type II diabetes. The genes obtained from this study showed a significant involvement in causing the disease, which may be used as a potential drug target.

Keyword: Type II diabetes, SVMRFE, Microarray, protein-protein interaction, t-test.

INTRODUCTION

Support Vector Machine (SVM), a machine learning technique implied in the area of time series prediction and classification (Ng and Mishra 2007; Rice et al. 2005) has widely been applied in the life science fields, especially in Bioinformatics. It can handle nonlinear classification tasks efficiently by mapping the samples into a higher dimensional feature space by using a nonlinear kernel function. Since the SVM approach is data-driven and model-free, it has important discriminating power for classification. This characteristic of SVM is obvious in cases where the sample sizes are negligible and numerous variables are involved (high-dimensional space).