ISSN: 0975-8585

Research Journal of Pharmaceutical, Biological and Chemical Sciences

Prediction of Epigenetic Variations in Alzheimer's disease Identification of Ethnic Variants through Pharmacogenomic Approach.

Sanjay Kumar P[1], S Karthikeyan[1], Preethi M Iyer[2], and P K Krishnan Namboori[1]*.

[1]Computational Chemistry Group (CCG), Amrita School of Engineering, Coimbatore, Amrita Vishwa Vidyapeetham, Amrita University, India.
[2]Dept. of Electronics and Communications Engineering, Amrita School of Engineering, Coimbatore, Amrita Vishwa Vidyapeetham, Amrita University, India.

ABSTRACT

To analyze the clinical trial data corresponding to proneness of Alzheimer's disease, to identify the epigenetic attributes helpful in making the prediction for different ethnic groups and to establish a correlation between the epigenetic factors and ethnicities towards AD responsiveness. The proneness of Alzheimer's disease varies among different ethnicities due to the variations in epigenetic, metagenomic and environmental factors. The epigenetic variants are highly contributing towards individual responsiveness. In the present work, clinical information from four major ethnic groups, African-American, Asian, Caucasian and Latino have been used for the analysis. The predictions have been made through machine learning approach using the kernel magic of 'support vector machine'. The behavior has been compared with the global population to quantify the influence of different ethnic groups. It has been found that the AD proneness can be effectively predicted through the RBF kernel of SVM with specific bias offset parameters. These parameters have been identified as unique computational markers of each ethnic group.

Keywords: Alzheimer's disease, Neurodegenerative diseases, Epigenetics, Ethnic variations

Corresponding author

ISSN: 0975-8585

INTRODUCTION

Alzheimer's disease (AD) is a type of neurodegenerative disease which is slow in progression and gradually results in impairment of memory, disturbances in reasoning etc.[1]. The proneness towards AD shows difference for various ethnicities concerned. It has been found that, the African-American and Hispanic (Latino) populations are more vulnerable towards late onset AD (LOAD) than whites (Caucasian). The incidence of LOAD is much less among Native Americans when compared to whites while the susceptibility rate was similar for Asian-Americans [2]. A recent study conducted at Australia proved that ethnic minorities there are at risk of having early onset AD (EOAD) than the majority Caucasians. The AD risk factor gene ApoE Є2 has been observed frequently among ethnic minorities like African-Americans, Hispanic, Native Alaskans and Hawaiians leading in to EOAD [3].

Any inherited phenotypic change in a multicellular organism without the involvement of genotypic change is defined as epigenetics. The gene regulation and structural organization are the consequences of DNA methylation process at the cytosine nucleotides catalyzed by DNA methyltransferases [4]. The methylation process is expected to be associated with a natural demethylation process, reverting back the variation without which it can become mutagenic and unstable [5]. Recent genetic studies have proved that the epigenetic modifications on DNA of neurons can bring about neuroendocrine, neurophysiologic and neurodegenerative consequences leading into AD [6].

METHODS

In the current study, data from clinical trials conducted by *Jason et al.* on patients with AD extracted from 'the national institute on aging genetics of Alzheimer's disease data storage (NIAGADS) has been used for the analysis. The data includes multi-ethnic exome array study of neurodegenerative diseases such as AD with the influensive frequency coding variants, probability of mutation of genes both in general neuro degenerative (discovery p value) and ethnic specific to AD (replication p value) . The ethnic identities considered in the study are African American, Latino, Caucasian and Asian. The p values from the clinical research have been considered for computing the gene vulnerability towards the disease. The genes with a higher p value ($p>0.05$) has been considered TRUE and the ones with lower ($p<0.05$) has been taken as FALSE. In the present analysis, 59 genes from the clinical research database with p value>0.05 (TRUE) and more associated with causing AD have been reviewed.

Involvement of AD genes causing disease varies from place to place; hence each gene has to be studied in a comprehensive manner. It has been found that the epigenetic variations regulate the gene expression leading into different levels of mutations. For establishing the importance of various attributes contributing to epigenetic variations, 847 promising attributes coming into 10 groups, DNA sequence, DNA structure, Repetitive DNA, Chromosome Organization, evolutionary history, population variation, Genes, regulatory regions, transcriptome, and epigenome and chromatin structure have been identified with the help of 'Epigraph', a web based SVM tool. The details of these attributes are listed below.

- DNA Sequence-The presence of thymine-adenine-thymine-adenine pattern, CpG island patern and cytosine content are the major components of concern.
- DNA Structure –The twists in DNA helix and solvent accessibility
- Repetitive DNA predicting the methylation possibility due to overlap with Alu elements, LINE elements and Tandem repeats.
- Chromosome Organization considering the possibility of mutation due to overlap with chromosome bands and isochores.
- Evolutionary history predicting the possibility of mutation due to overlap of evolutionary conserved regions.
- Population Variation signifying the possibility of mutation due to density of SNPs and overlap with specific SNP types
- Genes corresponding to the possibility of mutation due to overlap with annotated genes, pseudo genes and predicted microRNA genes.
- Regulatory regions signifying the possibility of overlap with CpG islands and predicting transcription factor binding sites causing mutation.

- Transcriptome corresponds to the mutations occurring due to the overlap with expressed sequence tag (ESTs) and m RNA sequence.
- Epigenome and Chromatin Structure signifying the histone modifications causing mutations indicated by overlapping with ChIP seq tags.

The radial kernel based 'Support Vector Machine' has been used as the platform for prediction and generating the regression equation using Rapid Miner Studio [8]. The mean correlation values of the 10 attributes from epigraph, mentioned above are used as attributes in the input dataset along with gene name as id and discovery p-value from clinical trials as label. Applying the magic of SVM kernels, RBF has been identified as the most suitable platform for the data. The radial kernel SVM can be defined as;

$$\exp\left(-\gamma\left\|x-x_i\right\|^2\right)+k$$

Where 'g' is gamma specified by the kernel gamma parameter which is adjustable and must be tuned carefully for the better performance of the kernel.

RESULTS AND DISCUSSIONS

The dataset consisting of 59 genes and 10 attributes has been identified as useful in making the predictions without any mismatch. The RBF kernel has been identified as the most suitable platform providing maximum accuracy. The regression equations for predicting the proneness of AD for the four ethnic groups and for the general population have been generated (Equations 1-5).

Afro-American

$$AA = N\exp\left(-\gamma\left\|x-x_i\right\|^2\right)+0.490 \quad (1)$$

Asian

$$A = N\exp\left(-\gamma\left\|x-x_i\right\|^2\right)+0.514 \quad (2)$$

Caucasian

$$C = N\exp\left(-\gamma\left\|x-x_i\right\|^2\right)+0.565 \quad (3)$$

Latino

$$L = N\exp\left(-\gamma\left\|x-x_i\right\|^2\right)+0.691 \quad (4)$$

General

$$G = N\exp\left(-\gamma\left\|x-x_i\right\|^2\right)+0.529 \quad (5)$$

Where N is the normalization constant, ϒ is the RBF parameter signifying the influence of samples selected as the support vectors, (x – xi) and the constant included is known as the bias offset parameter.

ISSN: 0975-8585

The RBF bias offset constants included in the equations act as fitting parameters specific to the trait and can be used for assessing the viability of the dataset for the predictions (Figure1).

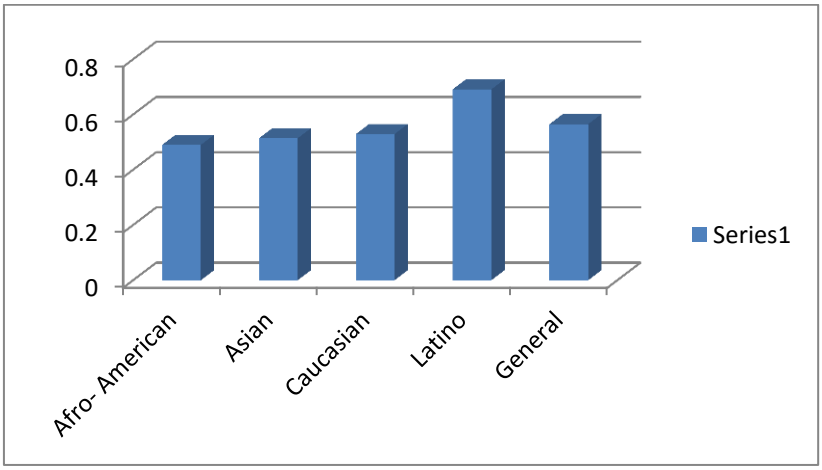

Figure 1: The bar diagram depicts the variation in the measure of bias offset variable of different ethnicities considered.

CONCLUSION

The responsiveness towards AD can be predicted effectively through the epigenetic attributes. However, ethnic variations lead into slight inconsistencies in the predictability. On analysis of the clinical data of the epigenetic attributes and proneness of the disease, it has been concluded that the behavior can be well predicted through regression equations generated in the RBF kernels. The bias offset constant has been identified as a unique ethnicity based marker/point indicating the efficiency of the prediction.

ACKNOWLEDGEMENT

The authors express their thanks to the Chairman, Director and all officials of the 'Coconut Development Board', Government of India, for the financial support in the form of a funded project on Alzheimer's disease.

REFERENCES

[1] Suresh Kumar A, Shilpa S, Anil Kumar NC, Rohith V, and Krishnan Namboori PK. Int J Recent Trends in Engineering and Technology 2010; 4(2):40-43.
[2] Anderson NB, Bulatao RA, and Cohen B. Critical perspectives on racial and ethnic differences in health in late life. National Research Council (US) Panel on Race, Ethnicity, and Health in Later Life. 2004; 4.
[3] Huei Yang Chen and Peter K Panegyres. Journal of Alzheimer's Disease 2016; Preprint:1-9.
[4] Szyf Moshe. Journal of Neurodevelopmental Disorders 2011; 3.3(-):238-249.
[5] C D Allis, T Jenuwein, and D Reinberg. Epigenetics. Cold Spring Harbor Laboratory Press, 2007.
[6] Natacha Coppieters and Mike Dragunow. Current Pharmaceutical Design 2011; 17: 3398-3412
[7] Chen JA, Wang Q, Davis-Turak J, and et al. JAMA Neurology 2015; 72(4):414-422.
[8] Rapidminer is an open source-learning environment for data mining and machine learning. https://rapidminer.com.Accessed: 2016-03-19.