IJBPAS, March, 2015, 4(3): 1678-1685

ISSN: 2277–4998

International Journal of Biology, Pharmacy and Allied Sciences (IJBPAS)

'A Bridge Between Laboratory and Reader'

www.ijbpas.com

IN SILICO MICROARRAY DATA ANALYSIS OF ST8SIA2 GENE RESPONSIBLE FOR SCHIZOPHRENIA

KUMARI A[*1, 2], MOHAN A[1], GANGWAR P[1]

1: Bio-EGICORE Department of Life Sciences, EGICORE, Lucknow (U.P.), India

2: Department of Biotechnology, Amity University, Lucknow (U.P), India

***Corresponding Author: E Mail: bioegicore@gmail.com; Mob.: Number: +91-9621500909**

ABSTRACT

Schizophrenia is a chronic, debilitating mental illness. Patients with schizophrenia have cognitive decay and abnormal electromagnetic sensory gating. Schizophrenia is characterized by development of psychopathology. Early stage of illness is characterized by disruption of oscillatory dynamics. This is followed by deficient memory updating, information sampling and higher cognitive functioning. Concept of schizophrenia has been changing regularly influenced mainly by the scientific paradigms. It begins in early adulthood with men having the tendency of developing it slightly earlier than women. Course and etiopathology have significant heterogeneity. ST8SIA2 is our target gene. This gene codes for ST8 Alpha-N-Acetyl-Neuraminide Alpha-2,8-Sialyltransferase 2, which is a type II membrane protein that is thought to catalyze the transfer of sialic acid from CMP-sialic acid to N-linked oligosaccharides and glycoproteins. Co regulated genes may share similar expression profiles, may be involved in related functions or regulated by common regulatory elements. There are different approaches to analyse the large-scale gene expression data in which the essence is to identify gene clusters. This approach has been used to determine expression profiles of previously described genes where raw data of patients was used as sample. The novel regulated genes identified were analyzed by using *in silico* approaches such as Clustering (HCL, KMC) and phylogenetic analysis. It was found that the genes CTIF and CELF5 were closely related to our gene of interest ST8SIA2.

Keywords: Clustering, Expression Profiles, Gene Expression, *in silico*, Phylogenetic Analysis, Schizophrenia

INTRODUCTION

The definition of schizophrenia has evolved through the six editions of the Diagnostic and Statistical Manual of Mental Disorders **[1]**. Three major roots are reflected in all definitions: a) the Kraepelinian emphasis on avolition, chronicity and poor outcome **(Kraepelin, 1971)**; b) incorporation of the Bleulerian view that dissociative pathology is primary and fundamental and accent on negative symptoms **(Bleuler, 1950)**; and c) the Schneiderian stress on reality distortion or positive symptoms **(Schneider, 1959)**. Schizophrenia is a chronic, debilitating mental illness. Patients with schizophrenia have cognitive decay and abnormal electromagnetic sensory gating. Schizophrenia is characterised by development of psychopathology. Early stage of illness is characterised by disruption of oscillatory dynamics. This is followed by deficient memory updating, information sampling and higher cognitive functioning. Concept of schizophrenia has been changing regularly influenced mainly by the scientific paradigms. Multiple symptoms of this disease involve various aspects of cognition and emotion. The strategy to target the neural substrates involves identifying a single brain region associated with a single symptom. With the increasing number of evidence the vital role of individual characteristics in insight and illness state have been established. The most commonly used definition of insight includes a multidimensional concept **[2-5]** (1) awareness of having a mental disorder, (2) awareness of the need of treatment, (3) understanding the social consequences of the disorder, (4) awareness of specific signs and symptoms of the disorder, and (5) attribution of symptoms to the disorder **[5, 6]**. Lack of this insight has been considered a core symptom of schizophrenia, a psychological coping mechanism aimed at preserving emotional well-being, **[7, 8]** a result of some cognitive or psychological dysfunction, **[9, 10]** a kind of neurological mechanism similar to anosognosia, **[11, 12]** or a combination of the above. Lack of awareness of illness, also termed poor insight in clinical contexts, refers to a patient's seeming inability to recognize his/her own illness or injury. The inexplicable quality of poor insight, the denial of good commonsense, and the dismissal of the obvious all apply to lack of awareness of illness whether seen in neurological or psychiatric conditions. The phenomenon is particularly common in schizophrenia where it affects an estimated 80%–97% of patients **[2, 3]** and impacts negatively on treatment compliance, **[7, 13]** course of illness and

prognosis, **[4]** and response to vocational rehabilitation **[7, 9]**.

A microarray is an array of DNA molecules that permit many hybridization experiments to be performed simultaneously. It can monitor expression levels of thousands of genes at one time. Microarray emerged in late 90s as a high-throughput technology for gene expression analysis. It has become a powerful tool for biomedical research **[14]**. In just a few years, microarrays have gone from obscurity to being almost ubiquitous in biological research. Microarray Data Analysis is one of the best and most widely used techniques in bioinformatics to study gene expression, disease diagnosis, target identification, gene screening, marker mapping and other developmental biology. Raw data of this microarray work was available at Gene Expression Omnibus (GEO) Database. At the same time, the statistical methodology for microarray analysis has progressed from simple visual assessments of results to a weekly deluge of papers that describe various novel algorithms for analysing changes in gene expression **[15]**. Various microarray gene clustering algorithms like hierarchical clustering, self organizing maps and k-means are found useful for discovering groups of correlated or co-expressed genes potentially co-regulated or associated to the disease or conditions under investigation.

Microarray Experiment

Microarray is a novel technology that facilitates the simultaneous measurement of thousands of gene expression levels **[16]**. The DNA microarray is made out of a glass, plastic or silicon chip. This chip has many microscopic DNA spots on its surface which form the array. The DNA spots are known as probes, because they are probing the sample which is hybridized to the chip. The sample which contains cDNA is called the target since the probes match these targets. Microarray experiments are typically made to compare two or more samples which represent two or more conditions, one being for example a cell that has mutated into a cancer cell and the other a normal cell **[17]**. We identified several genes not properly expressed in the schizophrenia patients at the mRNA level. In this report we have performed gene clustering of the raw microarray data sets provided by Gene Expression Omnibus (GEO) database at NCBI.

Gene ontology gives a further insight about cellular and molecular processes governing expression of ST8SIA2. Using GENESIS's HCL (hierarchical clustering) and K-means analysis, we found 24 harmful genes. The

expression and function of these genes were studied using GeneCard and Gene (NCBI). Further phylogenetic analysis was done for the harmful genes using MEGA 6.0. The genes which were found to be closely related to our target gene ST8SIA2 were CELF5 and CTIF. These closely related genes can be considered as potential novel targets to design drugs for schizophrenic patients in near future of biomedical research.

MATERIALS AND METHODOLOGY

Data Retrieval

The Gene Expression Omnibus is a publicly accessible repository at http://www.ncbi.nlm.nih.gov/geo . The high-throughput molecular abundance data (mainly gene expression data), which is generated by DNA microarray technology is archived and freely distributed by GEO. Fully annotated information is supported in MIAME (Minimum Information about a Microarray Experiment) infrastructure. A huge range of biological phenomena is addressed by about a billion individual gene expression measurements stored at GEO. 13 million gene expression profiles, for over 100 microorganisms , submitted by almost 1,500 laboratories are available for analysis. The GEO Datasets Database stores curated gene expression Datasets, as well as original series and platform records in the Gene Expression Omnibus (GEO) repository. Dataset records contain additional resources including cluster tools and differential expression queries. The raw data for microarray data analysis was obtained. The dataset record was GDS3938. This data was for *Homo sapiens.* **Citation**: Brennand K J , Simone A , Jou J ..(2011). **Title**: Induced pluripotent stem cell-derived neurons from Schizophrenia patients. The reference series was GSE25673 and the sample count was 24. For cluster analysis Dataset soft file was downloaded. The downloaded and saved raw data was then copied to excel sheet. The total number of genes present in raw data were 33298, which also included data for control.

Normalisation

The average value for our gene of interest i.e ST8SIA2 was found to be 7.08322. The raw data was sorted and scaled by taking logarithm at base 2 of R/G normalised (mean) ratio. The data was normalised by the rule that if missing value in a row is more than 80% then the row was deleted. After normalisation the number of genes left were 5029.

Genesis

Genesis was developed for the analysis of gene expression at a large scale. It is a platform independent, easy to use and versatile Java suite. Clustering algorithms

present in genesis include hierarchical clustering, self-organizing maps, k-means, principal component analysis, and support vector machines. Various tools for analysis of the data like normalisation and visualisation tools including filters are also present. The transparent nature of the results obtained after clustering for all implemented methods enables thorough analysis of the outcome of different parameters and algorithms. For further investigations of mechanisms that control transcription and to enhance promoter analysis mapping of gene expression data onto chromosomal sequences was implemented

Clustering for data

Hierarchical clustering (HCL) is a method of cluster analysis in which hierarchy of clusters is made **(Eisen *et al.*, 1998)**. It can be either divisive or agglomerative. The cluster was obtained as output by using microarray gene expression data .The tree obtained as output contained all the given data in hierarchical form. In accordance with gene expression value co-expressed genes will come under same cluster. With the help of different correlation it was found that the correlation which is suitable for this type of clustering and gives the most appropriate output is centred correlation **[18]**. After performing manual sub-clustering 27 different clusters were obtained. Then k means clustering was done. K-means **(MacQueen, 1967)** is one of the simplest unsupervised learning algorithms that solve the well known clustering problem. The procedure followed was that the number of clusters (assume k clusters) fixed a priori were the number of clusters obtained by manual clustering. The parameters defined were that the number of clusters were 27 and the maximum iteration value of 2000 was selected.

Comparison of clusters

Each and every cluster of HCL was compared with the clusters of k-means (27 clusters).It was found that our gene of interest ST8SIA2 was present in cluster number 6 of HCL and cluster number 9 of KMC. All the genes were studied independently. The gene list for the clusters having common genes was prepared. After manual study of all the genes, potentially harmful genes were also screened out.

Phylogenetic Analysis

Phylogenetic analysis helps in inferring evolutionary relationship between the important genes screened out. Evolutionarily related genes can have similar sequences i.e. can be similar for various mechanisms at molecular level. This property can be exploited for identifying potential drug targets

against the proteins encoded by these harmful genes.

RESULTS AND DISCUSSIONS

Prediction of co-expresses genes

Manual analysis of K means cluster and HCL cluster showed that one cluster was found which contained same genes by both HCL and KMC.

Common genes in clusters

The genes of these selected clusters 'seed cluster' can be used for further analysis. Both cluster contain different number of genes which show same gene expression. By evaluating clustered genes in genesis and then further analysing it was found that 24 harmful genes were similar. For the genes which appear in these clusters it can be stated that they may also be responsible for Schizophrenia. Through phylogenetic analysis we have seen that CTIF and CELF5 genes are closest to our target gene ST8SIA2 **(Figure 1)**.

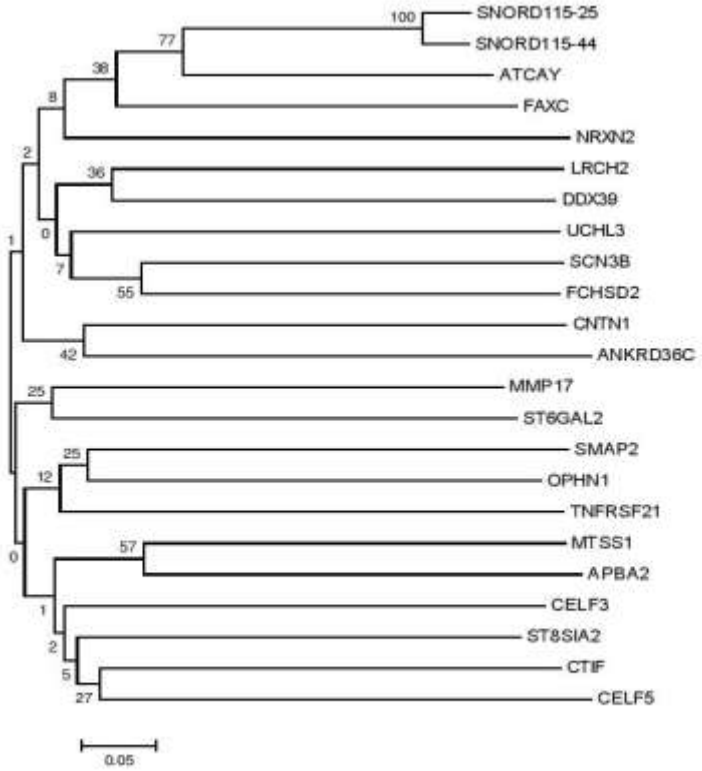

Figure 1: Phylogenetic tree of most important genes

CONCLUSION

The results of the clustering obtained from HCL and KMC algorithms indicate that genes common in specific cluster (KMC 9 and HCL 6) have similar expression patterns. Thus, it can be concluded that common genes of both clusters are differentially co-expressed. Comparative analysis can then be used as

basis to draw the conclusion that co-expression is present within the genes of same clusters. CTIF nad CELF5 genes were found to be most closely related to ST8SIA2 gene which is responsible for Schizophrenia.

REFERENCES

[1] DSM-I, DSM-II, DSM-III, DSM-III-R, DSM-IV, and DSM-IV-TR; American Psychiatric Association, 1952, American Psychiatric Association, 1968, American Psychiatric Association, 1980, American Psychiatric Association, 1987, American Psychiatric Association, 1994 and American Psychiatric Association, 2000).

[2] Amador XF, Flaum M, Andreasen NC, Awareness of illness in schizophrenia and chizo affective and mood disorders. Arch Gen Psychiatry. 1994; 51: 826–836.

[3] Amador XF, Strauss DH, Yale SA, Assessment of insight in psychosis. Am J sychiatry. 1993; 150: 873–879.

[4] Aleman A, Agrawal N, Morgan KD, David AS. Insight in psychosis and neuropsychological function: meta-analysis.Br J Psychiatry. 2006;189:204–212.

[5] Amador XF, Gorman JM. Psychopathologic domains and insight in schizophrenia. Psychiatr Clin North Am. 1998; 21: 27–42.

[6] Markova IS, Berrios GE. The meaning of insight in clinical psychiatry. Br J Psychiatry. 1992; 160: 850–860.

[7] Cuesta MJ, Peralta V. Lack of insight in schizophrenia. Schizophr Bull. 1994;20:359–366.

[8] Carpenter WT, Jr., Strauss JS, Bartko JJ. Flexible system for the diagnosis of schizophrenia: report from the WHO International Pilot Study of Schizophrenia. Science. 1973; 182: 1275–1278.

[9] Lysaker PH, Bryson GJ, Lancaster RS, Evans JD, Bell MD.Insight in schizophrenia: associations with executive function and coping style. Schizophr Res. 2003;59:41–47.

[10] Martens WH. The role of self-complexity in reducing harmful insight among persons with schizophrenia. Theoretical and therapeutic implications. Am J Psychother. 2009; 63: 53–68.

[11] Arango C, Adami H, Sherr JD, Thaker GK, Carpenter WT,Jr. Relationship of awareness of dyskinesia in schizophrenia to insight into mental illness. Am J Psychiatry. 1999; 156: 1097–1099.

[12] Gilleen J, Greenwood K, David A. Anosognosia in schizophrenia and other neuropsychiatric disorders: similarities and differences. In: Prigatano G, ed. Advances in the Study of Anosognosia. New York, NY: Oxford Psychiatric Press; 2010: 255–290.

[13] Schwartz-Stav O, Apter A, Zalsman G. Depression, suicidal behavior and insight in adolescents with schizophrenia. Eur Child Adolesc Psychiatry. 2006;15: 352–359.

[14] Module 10: Microarray Data Analysis I YanCui (ycui2@utmem.edu http://compbio.utmem.edu/MSCI814/Module10.html

[15] David B. Allison, Xiangqin Cui, Grier P. Page & Mahyar Sabripour Nature Reviews Genetics 7, 55-65 (January 2006).

[16] Bradley Efron, Robert Tibshirani, John D Storey andVirginia Tusher, Empirical Bayes Analysis of a Microarray Experiment, Journal of the American Statistical Association, Volume 96, Issue 456, 2001 pages 1151-1160

[17] David J. Lockhart & Elizabeth A. Winzeler, Genomics, gene expression and DNA arrays, Nature, Vol. 405, 15 June 2000.

[18] Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci., 95: 14863-8.