International Journal of Basic and Applied Biology
Print ISSN: 2349 – 5820; Online ISSN: 2349 – 5839; Volume. 2 No. 2, November 2014. pp. 45 - 49

http://www.krishisanskriti.org/ijbab.html

Microarray Analysis of Gene Expression in Primary and Sporadic Lateral Sclerosis

Prerna Agrawal[1], Deepshikha Sharma[2], Shivani Pandey[3], Ruchi Yadav[4]

[1,2,3,4] AMITY Institute of Biotechnology, AMITY University Uttar Pradesh, Lucknow
[1]Prerna5661272@gmail.com
[1]Mobile no- 9451891430

Abstract **: Amyotrophic lateral sclerosis (ALS) is a paralytic disorder caused by motor neuron degeneration. Primary lateral sclerosis is associated with loss of upper motor neurons and a more benign disease course up to 17yrs, amyotrophic lateral sclerosis is caused by loss of both upper and lower motor neurons and has an average disease course of 2-3 year. The majority of cases are sporadic, thereby limiting the availability of cellular models for investigating pathogenic disease mechanisms. In current study microarray data is retrieved from Gene Expression Omnibus database (GSE56808) in this experiment fibroblasts is derived from skin biopsies taken from sporadic amyotrophic lateral sclerosis and primary lateral sclerosis neurologically and normal human controls. There are eighteen samples out of which six samples are of fibroblast control, five samples of Primary Lateral Sclerosis(PLS),six of sporadic ALS and one of UMND. Over this data we have performed differential gene expression analysis using R and Bioconductor packages further significant genes are annotated using David software and Pathway analysis is done using GENMAP-CS to analyze genes involved in different biological pathways. These findings give insight towards the novel target identification that can be used further for identification of potential inhibitors and in docking studies. This research gives insight into pathway analysis of genes expressed in Amyotrophic lateral sclerosis disease and in its varying functions.**

Keywords: Amyotrophic Lateral Sclerosis, PLS, SALS, GEO database, R & Bioconductor, Metabolic Pathway, Differential Gene Expression.

INTRODUCTION

Amyotrophic **l**ateral **s**clerosis (ALS), often referred to as "Lou Gehrig's Disease," is a progressive neurodegenerative disease that affects nerve cells in the brain and the spinal cord.. The progressive degeneration of the motor neurons in ALS eventually leads to their death. When the motor neurons die, the ability of the brain to initiate and control muscle movement is lost. With voluntary muscle action progressively affected, patients in the later stages of the disease may become totally paralyzed.

In this paper we have used R software to analyse differentially expressed genes. R is an integrated suite of software facilities for data manipulation, calculation and graphical display . Among other things it has an effective data handling and storage facility, a suite of operators for calculations on arrays, in particular matrices, graphical facilities for data analysis and display either directly at the computer or on hardcopy.

Bioconductor provides tools for the analysis and comprehension of throughput genomic data. Bioconductor uses the R statistical programming language, and is open source and open development. It has more than 460 packages and is actively used by millions of user worldwide. There are various packages in Bioconductor for Microarray analysis like Affy, LIMMA that can be used for microarray data analysis. This paper describes the methodology to how to write expression values for differentially expressed gene and various visualization plots using R and Bioconductor for expression file . miRNA expression file (CEL file) is downloaded from the GEO database and is analyzed using R and Bioconductor.

METHODS

Microarray data retrieval from GEO database

The microarray data is retrieved from the GEO database and AFFYEXPRESS. Input file was made with 3 columns in Excel column Name as: **Name, File Name, Target and saved as "target.txt" format.**

	Name	FileName	Target
1	sample 1	NORMALCASE.CEL	control

2	sample2	normal case1.CEL	control
3	sample3	normal case 2.CEL	control
4	sample4	normal case11.CEL	control
5	sample5	normal case 13.CEL	control
6	sample6	normal case 14.CEL	control
7	sample7	pls case15.CEL	PLS
8	sample8	pls case16.CEL	PLS
9	sample9	pls case 22.CEL	PLS
10	sample10	pls case 25.CEL	PLS
11	sample11	pls case 36.CEL	PLS
12	sample12	sals case18.CEL	sALS
13	sample13	sals case21.CEL	sALS
14	sample14	sals case23.CEL	sALS
15	sample15	sals case26.CEL	sALS
16	sample16	sals case27.CEL	sALS
17	sample17	sals case31.CEL	sALS

Target file: - The target file contains the name of the 17 types of the disease with the filename and the specific target.

Quality control analysis through "affyQCReport" package

We have used " affyQCReport" package to quickly access the data quality of Affymetrix GeneChip. It generates many QC plots and printable pdf file is created.
source("http://bioconductor.org/biocLite.R")
biocLite("affyQCReport")

```
>  library(affyQCReport)
>  library(affydata)
> madata= ReadAffy(widget= TRUE)
> madata
> affyQAReport(madata)
> QCReport(madata, file="als.pdf")
```

Generate the Expression file

Normalization method is rma(robust mean analysis)eset= rma(madata)
write.exprs(eset,file="als.txt") by using this command expression file will be generated.
package affylmGUI:
source("http://bioconductor.org/biocLite.R")
biocLite("affylmGUI")
>library("affylmGUI")

RESULT:

Visualization Of The Expression Files

Boxplot- In descriptive statistics, a **box plot** or **boxplot** (also known as a **box-and-whisker diagram** or **plot**) is a convenient way of graphically depicting groups of numerical data through their five-number summaries: the smallest observation (sample minimum), lower quartile (Q1), median (Q2), upper quartile (Q3), and largest observation (sample maximum). A boxplot may also indicate which observations, if any, might be considered outliers.

Boxplots display differences between populations without making any assumptions of the underlying statistical distribution: they are non-parametric. The spacings between the different parts of the box help indicate the degree of dispersion (spread) and skewness in the data, and identify outliers. Boxplots can be drawn either horizontally or vertically.(Figure-1)

MA Plot- DNA microarrays consist of an arrayed series of thousands of microscopic spots of DNA which allow comparisons between two samples of RNA or DNA (target samples), which can provide data on relative gene expression levels (in the case of RNA) or gene copy number (for DNA). The data obtained from two-color DNA microarrays come in the form of fluorescent Red (Cy5) and Green (Cy3) dye intensities. One-color oligonucleotide arrays use only a single fluorescent dye. **M=log2R-log2G and A=1/2*(log2R+log2G)**

M is, therefore, the intensity ratio and A is the average intensity for a dot in the plot. The MA-plot is a plot of the distribution of the red/green intensity ratio ('M') plotted by the average intensity.(figure2,figure-3)

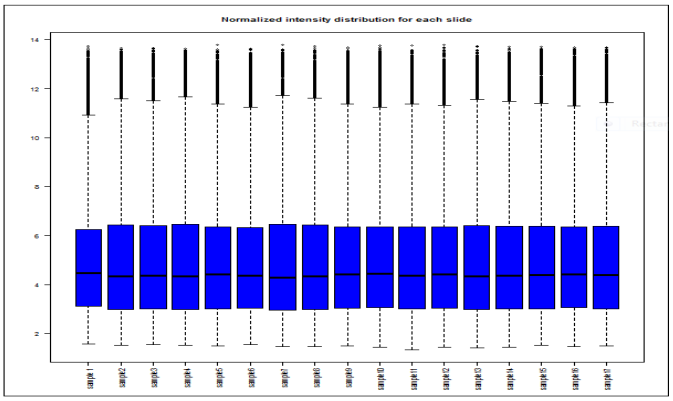

Figure 1: Normalized box plot

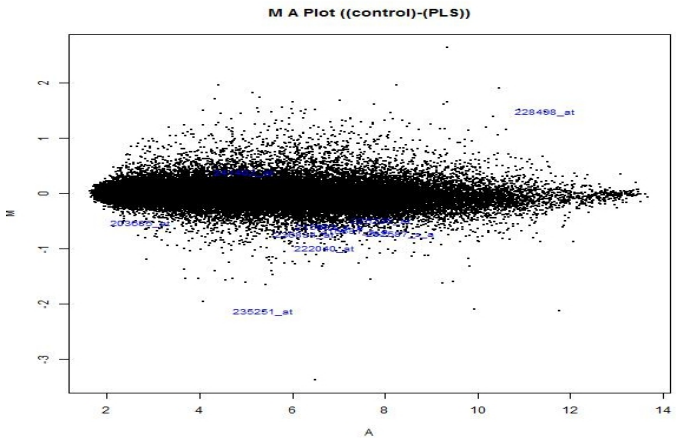

Figure 2- M A plot (control- pls)

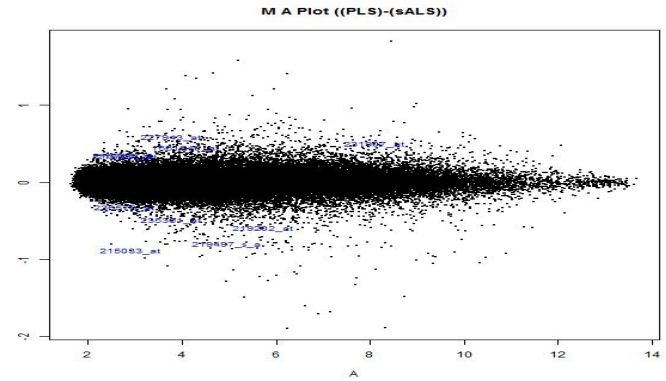

Figure 3- M A plot (PLS-sALS)

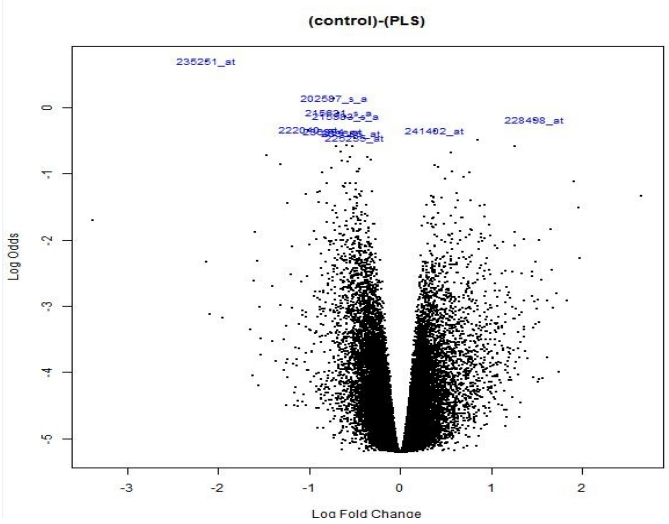

Figure4- log out plot (control-PLS)

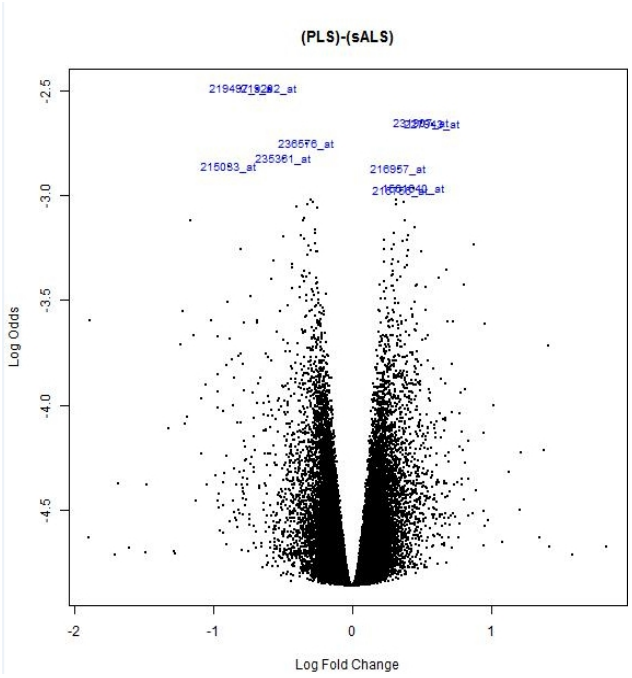

Figure5- log out plot(PLS-sALS)

David Tool- DAVID

(The **D**atabase for **A**nnotation, **V**isualization and **I**ntegrated **D**iscovery) is a free online bioinformatics resource developed by the Laboratory of Immunopathogenesis and Bioinformatics. All tools in the DAVID Bioinformatics Resources aim to provide functional interpretation of large lists of genes derived from genomicstudies, e.g. microarray and proteomics studies.

Gene ID	M	A	t	P.Value	B
235251_at	-2.13062	5.372172	-6.02558	0.660003	0.697703
202587_s_at	-0.73402	8.336269	-5.42314	0.660003	0.13381
215631_s_at	-0.67563	7.339889	-5.21224	0.660003	-0.07816
215983_s_at	-0.60705	6.810038	-5.1509	0.660003	-0.14124
228498_at	1.475662	11.46566	5.102549	0.660003	-0.1914
222040_at	-1.00294	6.702011	-4.96055	0.660003	-0.34099
241402_at	0.374263	4.975647	4.94462	0.660003	-0.35799
236834_at	-0.74692	6.268606	-4.94187	0.660003	-0.36092
203685_at	-0.54229	2.737331	-4.91406	0.660003	-0.3907

Table1- Top 10 Differentially Expressed Genes for (control)-(PLS)

ID	M	A	t	P.Value	B
219497_s_at	-0.8021	4.948415	-4.64645	0.999991	-2.49023
219202_at	-0.60003	5.734705	-4.64498	0.999991	-2.49099
231907_at	0.498047	8.092268	4.3368	0.999991	-2.65525
227943_at	0.579112	3.775054	4.329087	0.999991	-2.65949
236576_at	-0.32743	2.773641	-4.16058	0.999991	-2.75367
235361_at	-0.4949	3.774656	-4.03434	0.999991	-2.82613
215033_at	-0.88956	2.893404	-3.96585	0.999991	-2.8661
216957_at	0.332298	2.836757	3.949304	0.999991	-2.87583
1561640_at	0.441358	4.075699	3.792516	0.999991	-2.96929

Table 2- Top 10 Differentially Expressed Genes for (PLS)-(sALS)

CONCLUSION

These top 10 differentially expressed genes give insight towards the novel target identification that can be used further for identification of potential inhibitors and in docking studies. This research gives insight into pathway analysis of genes expressed in Amyotrophic lateral sclerosis disease and in its varying functions.

REFERENCES

[1] Matthew C Kiernan, Steve Vucic, Benjamin C Cheah, et al. (12 March 2011)."Amyotrophic lateral sclerosis.". Lancet. **377** (9769): 942–55

[2] Richard A. Becker, John M. Chambers and Allan R. Wilks (1988), The New S Language. Chapman& Hall, New York.

[3] Laurent Gautier, Leslie Cope, Benjamin M. Bolstad, and Rafael A. Irizarry. affy–analysis of Affymetrix GeneChip data at the probe level. Bioinformatics, 20(3):307–315, 2004.

[4] Gotkine M, Argov Z. Clinical differentiation between primary lateral sclerosis and upper motor neuron predominant amyotrophic lateral sclerosis. Arch Neurol. 2007;64:1545;