

## NOVEL MECHANISM PROPOSAL FOR INFLUENZA A (H1N1) BY GENE EXPRESSION DATA

Starlin.T<sup>1,3</sup>, Vidya.N<sup>2</sup>, Jannetvennila.J<sup>1</sup>, Gopalakrishnan V.K<sup>3</sup>, Anburaj.J<sup>4</sup>,  
Vimala .B<sup>5</sup>

<sup>1</sup>Department of Bioinformatics, School of Bio-Technology & Health Sciences. Karunya University, Coimbatore TN; <sup>2</sup>Sampark Tech Labs, Bangalore. Karnataka, India; <sup>3</sup>Department of Bioinformatics - Biochemistry, School of life Science Karpagam University Coimbatore TN; <sup>4</sup>Department of Biotechnology, Ayya Nadar Janaki Ammal College, Sivakasi TN; <sup>5</sup> Department of Biotechnology, Dayananda Sagar college of Engineering. Bangalore, KA.

### Abstract:

A basic, yet challenging task in the analysis of microarray gene expression data is the identification of changes in gene expression that are associated with particular biological conditions. The first stage is the normalization stage where data is cleaned from the effects of unwanted experimental biases. Clustering is an important stage in microarray data analysis. Gene clustering approach is to derive the meaningful data to investigate the response of gene with specific pathways from gene microarray data. Those genes will interact to carry out a particular biological function. The R environment package is used for micro array data analysis.

**Key words:** R package, Microarray, ArrayExpress, MeV, Gene pathway, docking.

### Introduction:

Influenza virus is one of the major pathogenic organisms in human and it produces a spectrum of clinical responses ranging from upper respiratory illness to central nervous system involvement. H1N1 virus is a subtype of influenza A virus and the most common cause of influenza (flu) in the case of humans. In 1918 the original H1N1 virus started infecting humans, then jumped to pigs and subsequently was reported in many wild animals since then lived on in pig population (Alan *et al.*, 2001).

The open source programming environment R provides mathematical statistical and graphical facilities that are used in many different field of science for data analysis or development of new analysis methods. It consists of a variety of R packages and interfaces to other software systems, each tackling specific bioinformatics data manipulation and analysis needs. R and Bioconductor together provide a comprehensive and powerful set of tools (Steffen *et al.*, 2009). Bioconductor is an open source and open development software project that provides a wide range of statistical and graphical tools based on R (Ihaka and Gentleman

1996) for the analysis and comprehension of genomic data (Gentleman *et al.*, 2004). These tools are distributed separately as interoperable packages, each specializing in different subareas of analysis such as the 'affy' package is to normalize Affymetrix chip data and the 'graph' package is to handle graph data structures. BioMart (<http://www.ebi.ac.uk/biomart>) is a simple federated query system designed specifically for using large datasets. One of the major databases providing a BioMart database implementation is the Ensembl (Hubbard *et al.*, 2005; Kasprzyk *et al.*, 2004) and Wormbase (Rogers 2008). Besides genome annotation databases several other biological databases are served through the BioMart system. The Bioconductor package biomaRt provides an Application Programming Interface (API) to BioMart web services, the programmable construction and subsequent analysis of large and complex queries to BioMart services from R (Durinck 2005). It allows the seamless embedding of identifier matching and annotation in statistical data analysis procedures. The biomaRt package is an add-on package for R that provides query ability to BioMart databases.

Microarray-based expression profiles have become a standard methodology in high-throughput analysis. Hierarchical clustering is the most widely used method for the analysis of patterns in gene expression. Hierarchical clustering produces a representation of data with the shape of a binary tree, in which the most similar pattern is clustered in a hierarchy of nested subsets. These techniques have already been applied to the study of gene expression patterns (Eisen *et al.*, 1998; Wen *et al.*, 1998). The normalization method solely relies on the experimental design of the microarray experiment (Kerr *et al.*, 2001).

The clustering step can be performed using point-based model-based or feature based grouping of data (Androulakis *et al.*, 2007) depending on the hypothesis adopted by the practitioner. All gene clustering algorithms is to discover the underlying gene pathways representing biological processes. Genes that are lying in the same pathway are often activated or depressed simultaneously or sequentially upon receiving stimuli. The biological signal is typically transmitted through intermediate gene interactions due to physical or chemical activities. The simultaneous or sequential activation or depression is delineated by the underlying network connection patterns (Ritesh *et al.*, 2010). The present study is to normalize the microarray data for the cluster analysis as well as to identify and short list the immune system pathway. This leads to predict the structure for docking studies.

#### Materials and Methods:

##### *Array express*

Array Express is a public database for high-throughput functional genomics data. The ArrayExpress Data Warehouse stores gene-indexed expression profiles and related measurements from a curated subset of experiments in the repository. Currently 1,00,000 hybridizations comprising

2,30,000 arrays are available from ArrayExpress. Retrieving publicly available data for analysis is a repetitive and error prone task in which the automation is desirable. Bioconductor is useful because it contains many widely used tools for data analysis to make a connection with public databases. The ArrayExpress query package was developed to load Affy datasets into Bioconductor.

The following functions have been used in ArrayExpress package

```
queryAE      : Query the database
ArrayExpress :
  Build object from raw data
getAE        :
  Download MAGE-TAB files
magetab2bioc :
  Convert MAGE-TAB files (local or from the
  database) into an object
getcolproc   :
  Extracts the column names from processed
  MAGE-TAB
procset      :
  Converts local MAGE-TAB files into an
  ExpressionSet
```

##### *Querying the Database*

The queryAE function was used to query the ArrayExpress repository. The current release of R was downloaded. The needed packages were installed from Bioconductor by starting R, which gives the following instructions to the R interpreter:

```
>source(http://bioconductor.org/biocLite.R)
>biocLite()
>library(ArrayExpress)
>influenza = queryAE(keywords =
  influenza+treatment, species =
  homo+sapiens)
>dim(influenza)
>influenza[1:7]
```

The output is a data frame with identifiers and two columns which enables the availability of raw and processed data. The

raw and processed data are available for the same dataset a unique identifier is given for both. There is no way to distinguish between the processed and raw dataset just from the identifier. The column 'Raw' from the output of ArrayExpress leads to know the raw data available for dataset.

#### *Getting Raw Data - Affymetrix*

```
> MEXP-1274 = ArrayExpress(E-MEXP-1274)
> MEXP-1274
> colnames(pData(MEXP-1274))
> pData(MEXP-1274)[1:31:8]
> head(fData(MEXP-1274))
> experimentData(MEXP-1274)
```

Transcription profiling of human bronchial epithelial cells stably expressing a dominant negative form of TLR3 after infection with influenza A virus not infected. To explore the host responses associated with TLR3 signaling IAV infection used the human bronchial epithelial cell line stably expressing a dominant negative form of TLR3 (pZERO-hTLR3) and examined global impact of TLR3 through an analysis of gene expression changes compared to control cells at 24 hours after infection.

#### *Normalizing Microarray Data*

The purpose of normalization is to identify and remove the effects of systematic variation other than differential expression in the measured fluorescence intensities. It is necessary to normalize the fluorescence intensities before any analysis which involves comparing expression levels within or between slides.

Preprocessing the Affymetrix expression arrays involve three steps: background adjustment, normalization, and summarization. Self-contained routines were used for background correction and normalization; it takes an Affy Batch as input and returns a processed Affy Batch. Routines in summarization produced exprSet objects containing expression summary values. This effect can be

illustrated by studying boxplots of the raw influenza A (E-MEXP-1274) data, in the present work to examine probe intensity behavior for a number of different arrays the box plot methods were used which is shown in the Figure 1(b). Boxplot provides a useful graphical view for identifying differences in the level of raw probe-intensities between arrays. A difference between arrays in shape or center of distribution often highlights the need for normalization. The boxplot command is in the made4 package and so the figure is produced by using the R commands.

An object (eset) has been created and defining it as the outcome of running rma on the previously-named data. rma does three things: convolution background correction, quantile normalization and summarization based on multi-array model fit robustly using the median polish algorithm. The normalized data stored in eset were exported to a text file with the name of Normalized.txt.

Open R.

```
> setwd(path of cel file directory )
> source(http://bioconductor.org/biocLite.R
)
> biocLite()
> library(affy)
> Data<-ReadAffy()
> eset<-rma(Data)
> write.exprs(eset      file=Normalized.txt
sep=\\t)
```

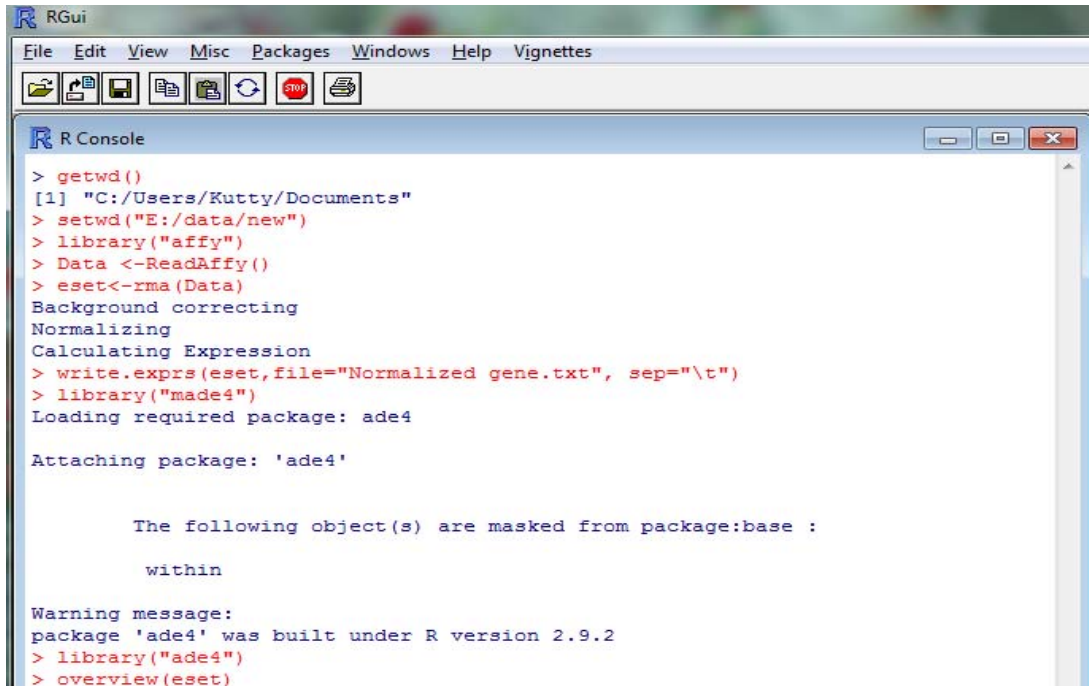
Install and load the made4 package for plot analysis

```
> biocLite(made4)
> library(made4)
> overview(eset)
```

Quick overview of the data boxplot has been obtained by using the above command.

#### *Clustering the normalized data*

MeV has integrated an R (CRAN) hook; therefore the R functions and libraries can be called within the java instance using



```

RGui
File Edit View Misc Packages Windows Help Vignettes

R Console
> getwd()
[1] "C:/Users/Kutty/Documents"
> setwd("E:/data/new")
> library("affy")
> Data <-ReadAffy()
> eset<-rma(Data)
Background correcting
Normalizing
Calculating Expression
> write.exprs(eset,file="Normalized gene.txt", sep="\t")
> library("ade4")
Loading required package: ade4

Attaching package: 'ade4'

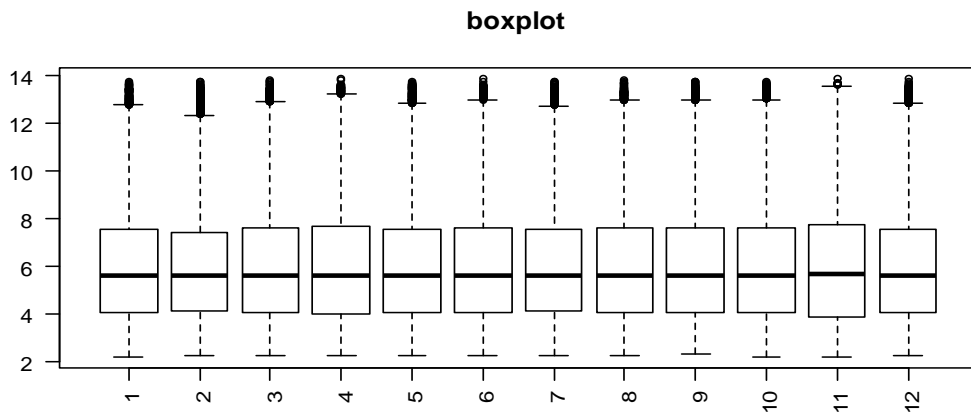
The following object(s) are masked from package:base :

    within

Warning message:
package 'ade4' was built under R version 2.9.2
> library("ade4")
> overview(eset)

```

**Figure 1(a): Normalizing the micro array data**



**Figure 1(b): Boxplot**

shared libraries. A variety of normalization algorithms and clustering analyses allow the user flexibility in creating meaningful views of the expression data. MeV tool was used for cluster analysis and the normalization was predicted using the following steps:

The normalized data was loaded from the local disc using the File Loader option. The drop-down menu was clicked to load .txt format files. At the top left the Clustering drop down menu was clicked and first

option Hierarchical Clustering was selected. The button OK was clicked at the bottom right. To view the clusters, HCL on the navigation tree on the left was double clicked. The cluster was stored by clicking on the groups. By clicking one particular line that particular group of gene was highlighted. The option Store Cluster was chosen by Right clicking on the line.

#### *Gene to pathway*

An important area in microarray data analysis is to infer signaling pathways. The

signaling pathway is a sequence of gene interactions leading to a specific biological end point function. The R package `gene2pathway` implements a classification model in which a specific gene of interest can predict the mapping to a KEGG pathway based on its domain signature. Usage of the R package `gene2pathway` mainly involves two functions: `gene2pathway` and `gene2pathway.signaltrans`. `Gene2pathway` predicts the KEGG pathway membership for a given list of genes. The mapping of genes to InterPro domains can be done automatically via Ensembl if Entrez gene IDs are passed. Alternatively the user can provide its own mapping in form of a list. In this case arbitrary gene identifiers can be used. `Gene2pathway` package used to identify the signaling pathway for the genes using commands.

Open R

```
>source(http://bioconductor.org/biocLite.R)
>biocLite(gene2pathway)
>library(gene2pathway)
> library(kernlab)
> library(KEGG.db)
> library(AnnotationDbi)
> library(Biobase)
>gene2pathway(Entrez Gene ID
KEGG.package=TRUE)
```

#### *Identifying the gene interaction for the immune system genes using HPRD*

The current release of HPRD has annotated a total of 2750 protein sequences. This number includes proteins encoded by 1484 genes that represent all genes with allelic variants which are annotated as linked to a human disease in OMIM database. The data base serves as a nice protein-centric complement to the OMIM database that provides extensive annotations about genes and their variations associated with human diseases.

The interactions (Receptor and Ligand) of immune system gene were taken from Human Protein Reference Database. It contains annotations pertaining to human proteins based on experimental evidence from the literature.

#### *Homology modelling*

CPH Model 3.0 Server is a novel method for fold recognition/homology modeling in which the larger sequence database is iteratively searched to construct a sequence profile until a template is found in a database of proteins with known structure. The receptor and ligand sequence were taken from HPRD. The sequence was load into the CPH Model 3.0 Server. The output of the modeled structure was imported as PDB file.

#### *Energy minimization using Deepviewer*

The structure of Human receptors and ligands were modeled using the CPH 3.0 server. All the receptors and the ligands were minimized using Swiss PDB Deep Viewer. These minimized structures are saved in PDB format. Geometry-based molecular docking algorithm was used for the studies.

#### *Docking using PatchDock*

The Receptor and the ligand were imported into PatchDock web server. The docking form was submitted and its result was send to user email.

Many types of docking has been done such as Human Receptor with Human Ligand, Human Receptor with Viral Ligand (Hemagglutinin), Human Receptor with Viral Ligand (Neuraminidase), Human Receptor - Viral Ligand (Neuraminidase) complex with Human Ligand, and Human Receptor - Viral Ligand (Hemagglutinin) complex with Human Ligand.

### **Results and Discussion:**

#### *Normalizing the micro array data*

The goal of normalization is to make the same distribution of probe intensities for array.

**Table 1: Gene Pathway results**

Pathways	No of genes
Cellular Process	77
Environment Information Processing	73
Human Disease	54
Signaling Transduction	52
Cell Growth and Death	38
Immune System	34
Signaling Molecule and Interaction	31
Apoptosis	31
Cancer	21
MAPK signaling pathway	21
Cytokine-cytokine receptor interaction	21
Toll-like receptor signaling pathway	21
Genetic information Processing	22
Neurodegenerative killer cell mediated cytotoxicity	15
Immune Disorders	15
Lipid Metabolism	13
Infectious Disease	11
Insulin signaling pathway	11

Normalization result was obtained using R Affy package. The normalized gene expression data was stored in Normalized gene.txt as shown in Figure 1(a). Similarly normalization of intensities is an important step aiming to allow data comparison by removing some systematic bias. Though it is not yet consensus about which normalization method is the most appropriate one (Shippy *et al.*, 2006), it has been noted that in general, the non-linear dependence on signal intensity performs better (Workman *et al.*, 2002).

#### Clustering the Normalized data

Rosa *et al.*, 2011 suggested some procedures for validating the clustering solutions for K-means cluster analysis. In this present work; the hierarchical clustering method was applied to produce

large amounts of data with difficult interpretation. The schematic representation is shown in figure 2(a&b). The genes which levels were up-regulated are shown as red color and the down-regulated genes are represented in green color. The expression graphs give a more detailed representation of the gene behavior. Each individual line represents one gene and the pink line in the middle shows the average expression of all the genes in each sample. Eisen *et al.*, 1998 reported some biological case studies by point-wise methods, using this method clustering gene expression data can be found in genome wide expression patterns. However, the clustering algorithm uses splines to cluster the continuous representation of time series expression data. In some cases, prior knowledge has been used to fit the models to the expression profiles (Bar-Joseph *et al.*, 2003)

#### Identifying the gene pathway

Ulitsky *et al.*, 2010 identified pathways in the interaction network that exhibits ample evidence in the disruption of transcription which is specific to diseased patients, but without requiring that any gene is significantly expressed differentially across all the cases. In this present investigation, Pathways of clustered genes were identified and its result is shown in Table: 1. many of the genes are found in biological processes like Cellular Process cell growth and death, Human Disease, Immune System, Signaling Transduction, Signaling Molecule and Interaction metabolism. This result is given in Figure 3.

#### Identify the gene interaction for the immune system genes

The Human Protein Reference Database (<http://www.hprd.org>) is a web-based resource based on open source technologies for protein information about several aspects of human proteins including

**Table 2: Gene Interaction**

Gene	Human Receptor	Name of the receptor
FAS	FAS	FAS receptor
CFB	Complement receptor 1-complex	Complex
C1R	CD93	Complement component 1q receptor
C1S	CD93	Complement component 1q receptor
C4A	C3AR1	Complement component 3a receptor 1
CSF1	CSF1R	Colony stimulating factor 1 receptor
HLA-F	LILRB1	Leukocyte immunoglobulin like receptor subfamily B member 1
IFNB1	IFNAR1	Interferon alpha receptor
IL1B	IL1R1	Interleukin 1 receptor type I
IL6	IL6R	Interleukin 6 receptor alpha
CXCL10	CXCR3	CXC chemokine receptor type 3
IRF7	TICAM2	Toll like receptor adaptor molecule 2
PAK1	ESR1	Estrogen receptor alpha
	TGFBR1	TGF beta receptor type I
	ACVR1	Activin A receptor type I
PIK3CB	AGTR2	Angiotensin II receptor type 2
PPP3CA	GABRA1	GABA-A receptor alpha 1
	GABRG2	GABA-A receptor gamma 2
	GABRB2	Gamma aminobutyric acid receptor beta 2
CCL5	CC chemokine receptor 5	CC chemokine receptor 5
CXCL11	CXC chemokine receptor 3	CXC chemokine receptor 3
RIPK1	TNFRSF1A	Tumor necrosis factor receptor 1
TNFSF10	TRAIL receptor3	TRAIL receptor3
TNFRSF10B	Death receptor 5	Death receptor 5
MAPK9	Grb2	Growth factor receptor-bound protein 2
RELA	Glucocorticoid receptor	Glucocorticoid receptor
MAP3K14	TRAF2	TNF receptor-associated factor 2
IKBKE	TNFRSF1A	Tumor necrosis factor receptor 1
STAT1	EGFR	Epidermal Growth Factor Receptor
PRKCA	EGFR	Epidermal Growth Factor Receptor

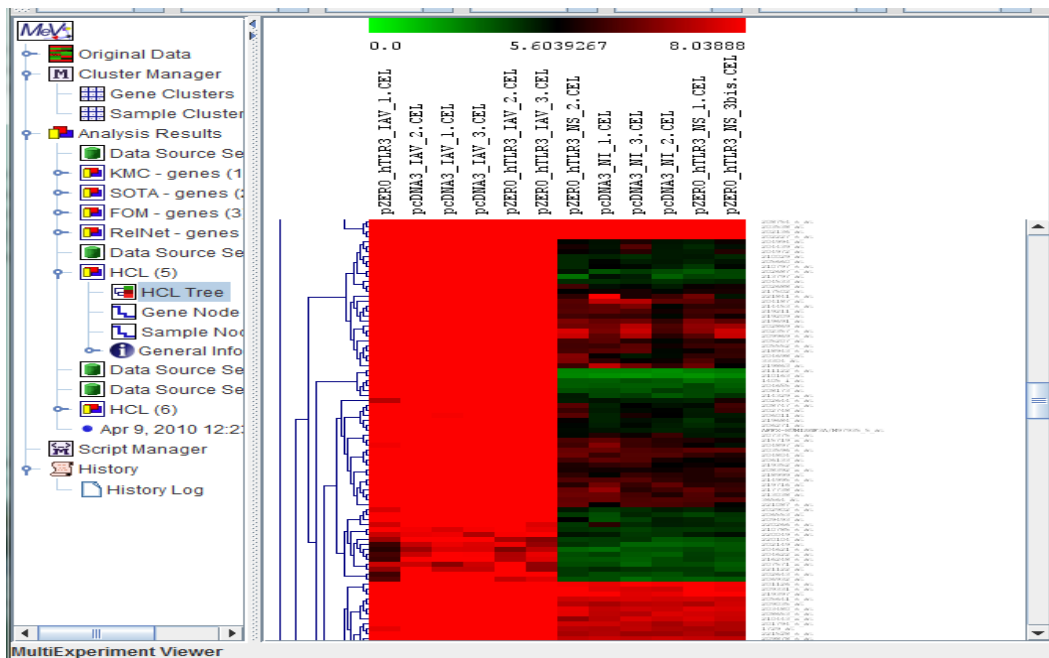
protein-protein interactions, post-translational modifications, enzyme-substrate relationships and disease associations (Peri *et al.*, 2004). The immune system gene interactions were identified and its result is shown in Table 2. Many of the immune system genes (RASGRP3, TAP2, AKT3, LGMN and IFI30) had no interaction. Among 34 genes, only 9 genes had expression values.

#### Docking

Patch Dock results were viewed using Swiss PDB Deep Viewer. It is found from the PATCH DOCK result that first result has a very high value of area, atomic contact energy, score and transformation. The energy values were obtained for each gene which is tabulated in table 3. The expressions genes are shown from figures 4-12.

**Table 3: Docking results**

Gene ID	Gene Name	Name of the Receptor	Human (Receptor-Ligand) Energy value	Human Rec-viral ligand (Neuraminidase) complex value	Human Rec-viral ligand (hemagglutinin) complex value
6352	CCL5	CC chemokine receptor 5	-370.1	-295.21	-437.59
6373	CXCL11	CXC chemokine receptor 3	-370.3	-432.82	-188.61
5970	RELA	Glucocorticoid receptor	223.22	412.88	280.71
5578	PRKCA	Epidermal Growth Factor Receptor	365.55	-121.35	250.46
8743	TNFSF10	TRAIL receptor3	21.9	179.94	88.43
8795	TNFRSF10B	Death receptor 5	376.87	218.64	179.55
9020	MAP3K14	TNF receptor-associated factor 2	13.62	311.93	432.68
9641	IKBKE	Tumor necrosis factor receptor 1	384.9	246.87	143.5
355	FAS	FAS receptor	398.61	289.97	276.44



**Figure 2(a) Clustered data**



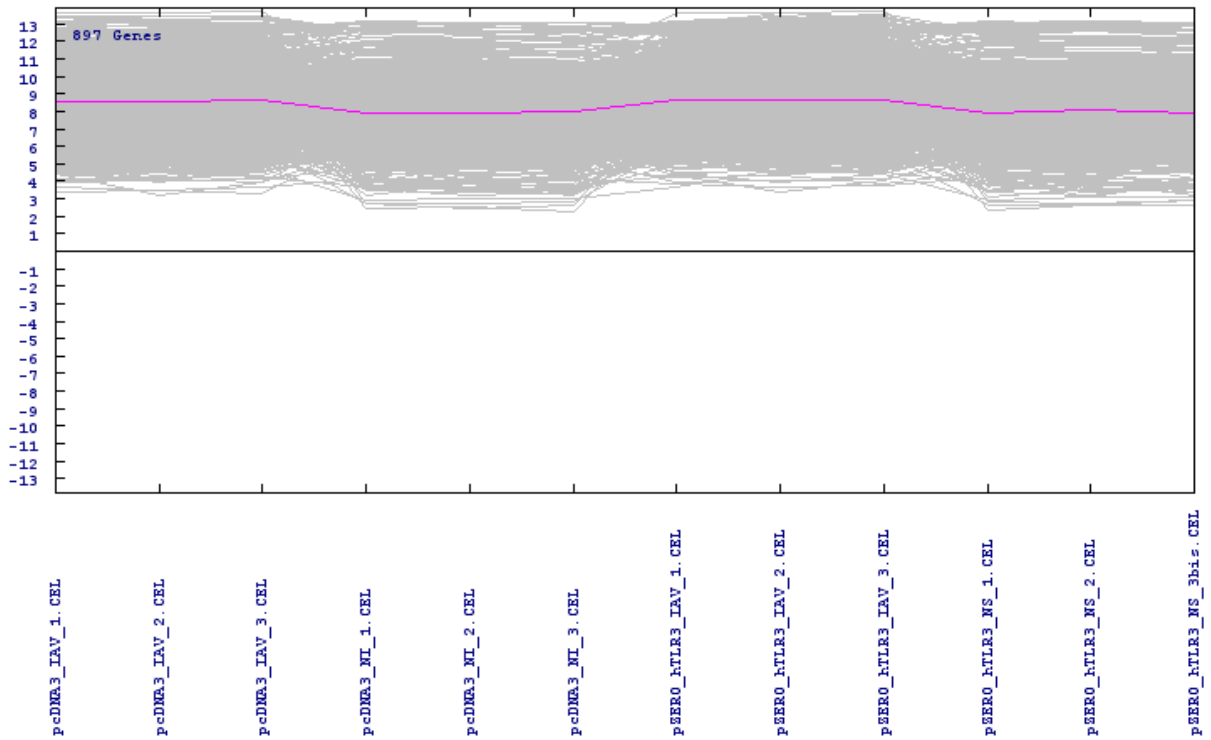


Figure 2 (b): Expression Graph

```

RGui - [R Console]
File Edit View Misc Packages Windows Help Vignettes

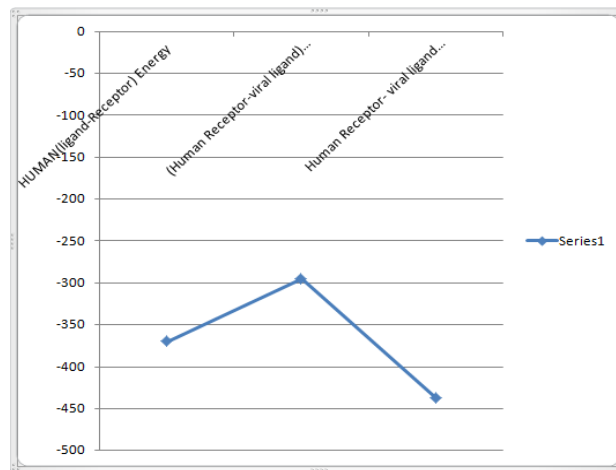
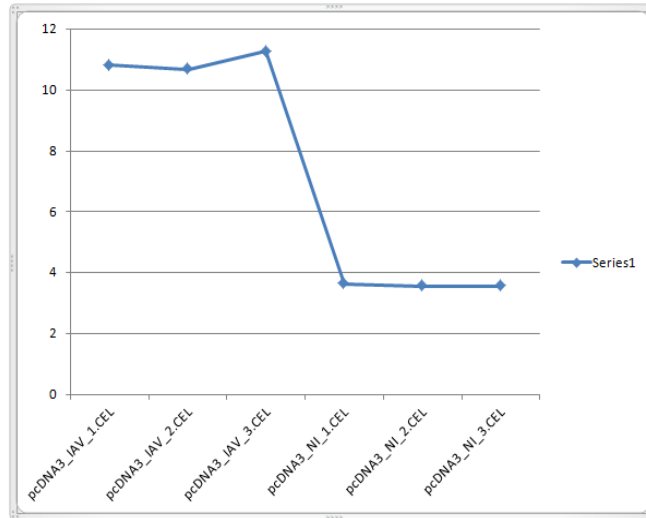
> library("Biobase")
> gene2pathway("3569", KEGG.package=TRUE)
Using KEGG information from KEGG.db package ...
Information from KEGG package available for 1 genes ...
finished
$gene2Path
$gene2Path$`3569`
 [1] "04060" "04620" "04630" "04640" "05060" "05332" "01130" "01133" "01140"
 [10] "01145" "01132" "01160" "01163" "01162"

$gene2Pathname
3569
04060 "Cytokine-cytokine receptor interaction"
04620 "Toll-like receptor signaling pathway"
04630 "Jak-STAT signaling pathway"
04640 "Hematopoietic cell lineage"
05060 "Prion disease"
05332 "Graft-versus-host disease"
01130 "Environmental Information Processing"
01133 "Signaling Molecules and Interaction"
01140 "Cellular Processes"
01145 "Immune System"
01132 "Signal Transduction"
01160 "Human Diseases"
01163 "Neurodegenerative Diseases"
01162 "Immune Disorders"
attr(,"names")
 [1] "3569" NA NA NA NA NA NA NA NA
 [11] NA NA NA NA

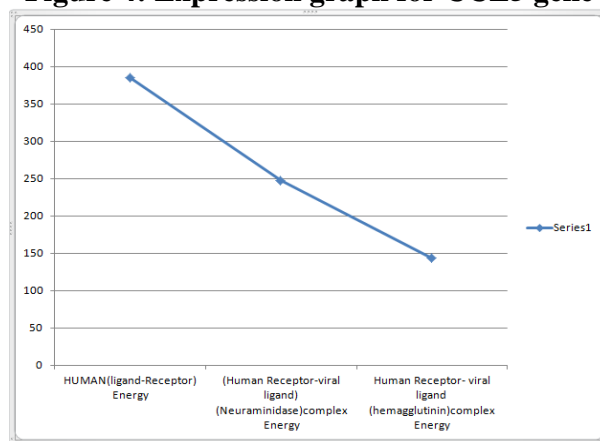
$byKEGG
3569
TRUE

$scores
$scores$`3569`
NULL
    
```

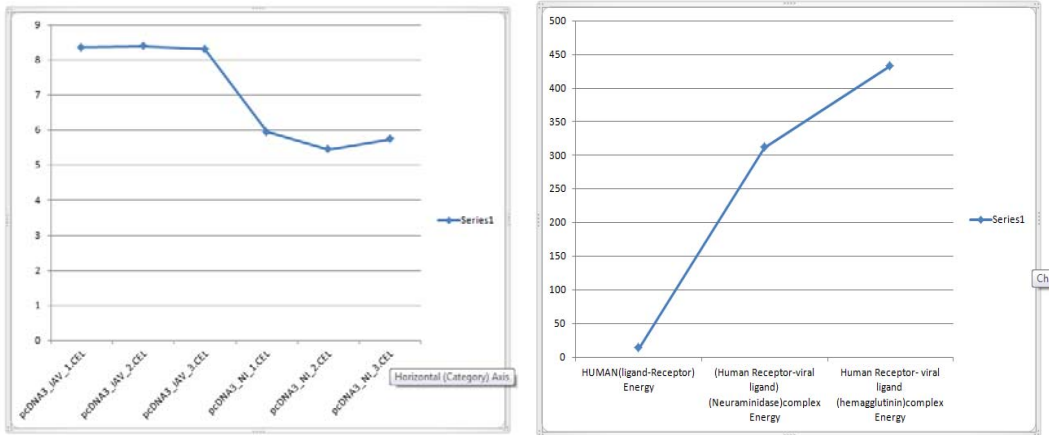
Figure 3: Pathway identification of genes



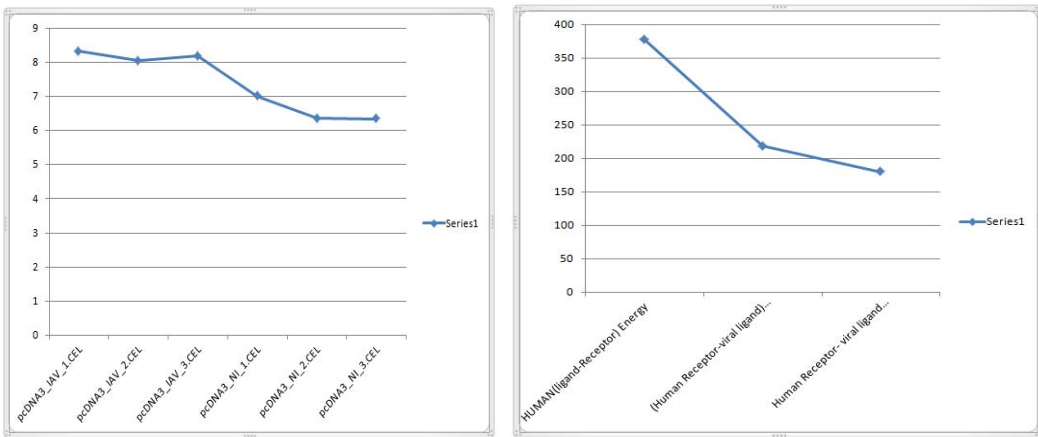
**Figure 4: Expression graph for CCL5 gene**



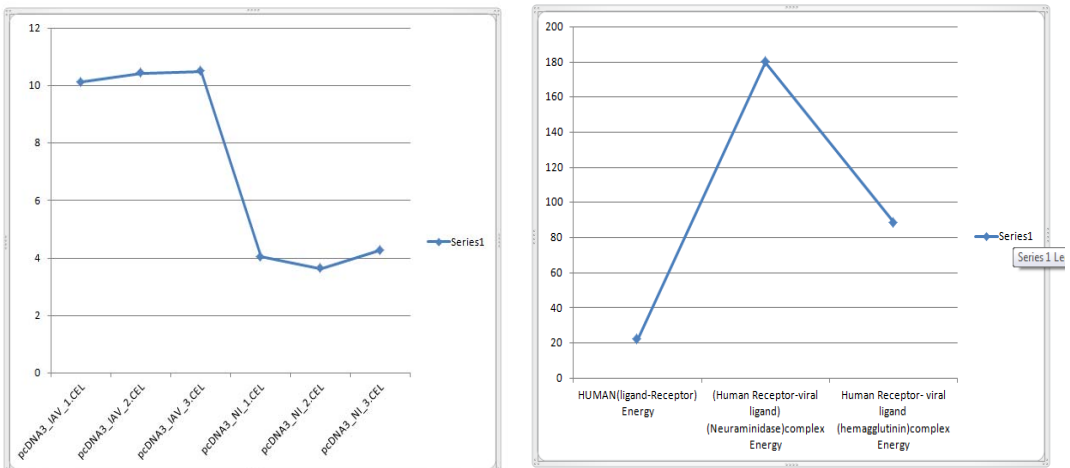
**Figure 5: Expression graph for IKBKE gene**



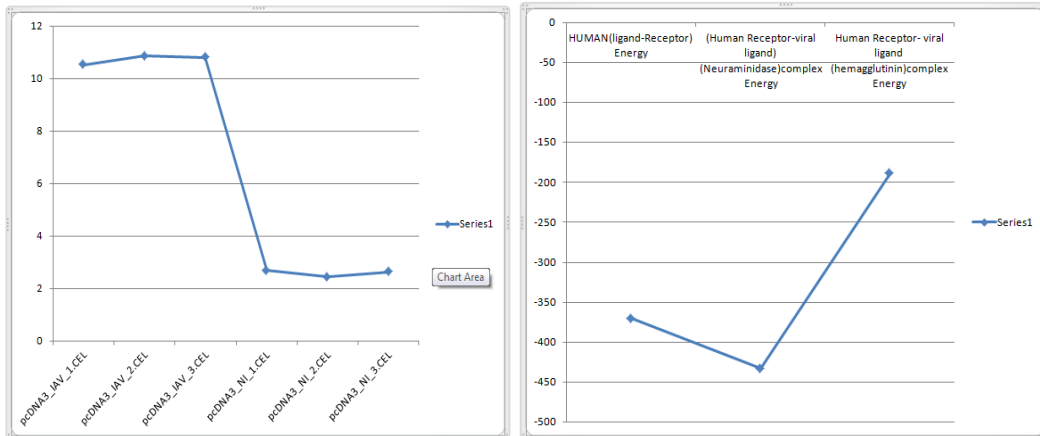
**Figure 6: Expression graph for MKP3K14 gene**



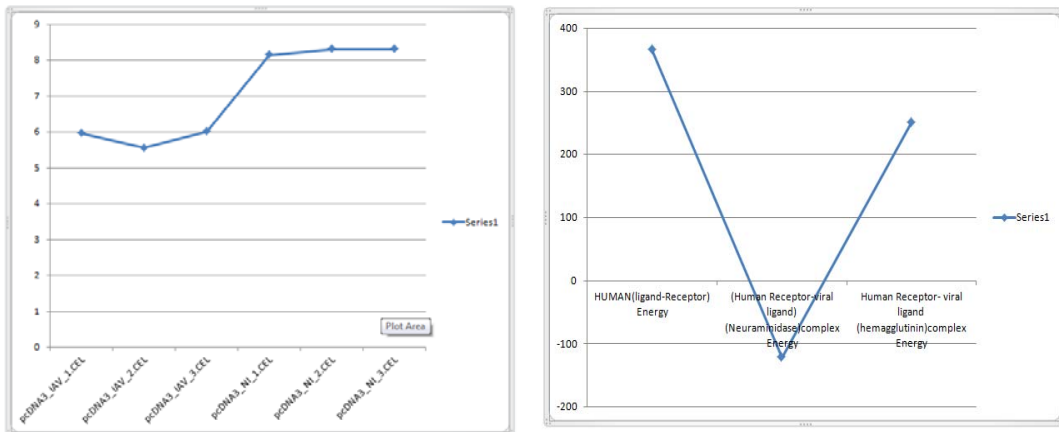
**Figure 7: Expression graph for TNFRSF10B gene**



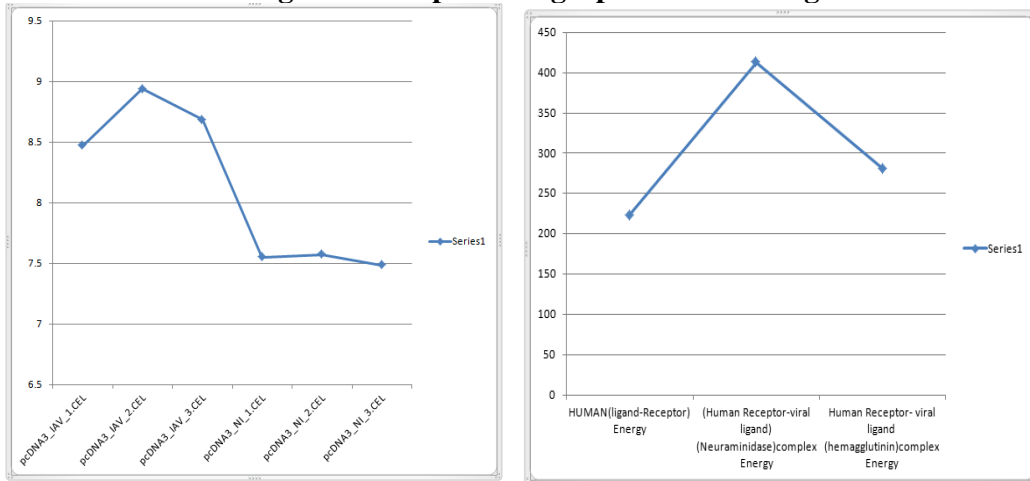
**Figure 8: Expression graph for TNFRSF10 gene**



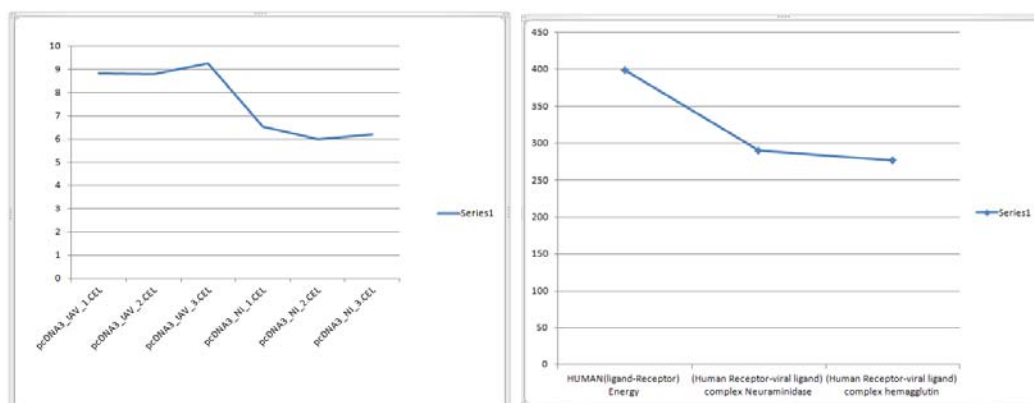
**Figure 9: Expression graph for CXCL11 gene**



**Figure 10: Expression graph for PRKCA gene**



**Figure 11: Expression graph for RELA gene**



**Figure 12: Expression graph for FAS gene**

### Conclusion:

The present works have represented a successful application of R package to analyse the microarray data. Based on the overall studies it can be concluded that a set of very important 9 genes plays a crucial role for the high expression of microarray data in Influenza A (H1N1). This indicates that these 9 genes have high competitive binding with the immune receptor except the one gene PKCA which shows decrease in expression. Hence it is considered as a novel approach for data mining to mine the useful data from the vast gene expression data.

### Acknowledgment:

I am extremely grateful to Dr. Vidhya Niranjana, for her stimulating guidance, valuable suggestion, critical and persistent encouragement.

### References:

- [1] Alan JH., Victoria G., Alan RD., and Yi PL., (2001). The evolution of human influenza viruses. *Phil. Trans. R. Soc. Lond* 356: 1861-1870
- [2] Androulakis IP., Yang E., and Almon RR., (2007). Analysis of Time-Series Gene Expression Data: Methods Challenges and Opportunities. *Annual Review of Biomedical Engineering*. 9: 205-228.
- [3] Bar-Joseph Z., Gerber G., Jaakkola T., Gifford D., and Simon I., (2003). Continuous representations of time series gene expression data. *J Comput Biology*. 3(4): 341-356.
- [4] Durinck S., (2005). BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics*. 21: 3439-3440.
- [5] Eisen MB., Spellman PT., Brown PO., and Botstein D., (1998). Cluster analysis and display of genome-wide expression patterns. *Proc. Natl Acad. Sci.* 95: 863-868.
- [6] Gentleman RC., Carey VJ., Bates DM., Ben B., Marcel D., Sandrine D., Byron E., Gautier L., Yongchao GE., Jeff G., Kurt H., Torsten H., Wolfgang H., Stefano I., Rafael I., Friedrich L., Anthony J., Gunther S., Gordon S., Luke T., Yang JY., and Zhang J., (2004). Bioconductor: open software development for computational biology and bioinformatics. *Genome Biology*. 5: 80-90.
- [7] Hubbard TJ., Andrews M., Caccamo G., Cameron Y., Chen M., Clamp L., Clarke G., Coates T., Cox F., Cunningham V., Curwen T., Cutts T., Down R., Durbin X., Fernandez-Suarez J., Gilbert M., Hammond J., Herrero H., Hotz K., Howe V., Iyer G., Proctor M., Rae D., Rios M., Schuster S., Searle J., Severin G., Slater D., Smedley J., Smith W., Spooner A., Stabenau J., Stalker R., Storey S., Trevanion J., Vogel S., White C., Woodwork T., and Birney E., (2005). Ensembl. *Oxford journals Nucleic Acids Research*. 33: 447-453.
- [8] Ihaka R., and Gentleman RC., (1996). R: a language for data analysis and graphics. *J. Comput. Graph. Stat* 5: 299-314.
- [9] Kasprzyk A., Damian K., Damian S., Darin L., William S., Craig M., Martin H., Philippe RS., Tony C., and Ewan B., (2004). Ensmart: a generic system for fast and flexible access to biological data. *Genome Res.* 14: 160-169.
- [10] Kerr MK., and Churchill GA., (2001). Statistical design and the analysis of gene expression microarray data. *Genet Res.* 77: 123-128.
- [11] Peri S., Navarro J.D., Kristiansen T.Z., Amanchy R., Surendranath V., Muthusamy B., Gandhi T.K., Chandrika K.N., Deshpande N.,

- Suresh S., Rashmi B.P., Shanker K., Padma N., Niranjana V., Harsha H.C., Talreja N., Vrushabendra B.M., Ramya M.A., Yatish A.J., Joy M., Shivashankar H.N., Kavitha M.P., Menezes M., Choudhury D.R., Ghosh N., Saravana R., Chandran S., Mohan S., jonnalagadda C.K., Kumar S.C., Deshpande K.S., Pandey A., (2004). Human protein reference database as a discovery for proteomics *Nucleic Acid Res.* D497-501
- [12] Ritesh K., Chang-Tsun L., and Buchanan-Wollaston T., (2010). A temporal precedence based clustering method for gene expression microarray data. *BMC bioinformatics.* 11: 68.
- [13] Rogers A., (2008). Wormbase. *Nucleic Acids Res.* 36: 612–617.
- [14] Rosa J.J.G., Marrufo M.V., Vaquero E., Duque P., Borges M., Gonzalez C.M.G., and Izquierdo G., (2011). Cluster analysis of behavioural and event-related potentials during a contingent negative variation paradigm in remitting-relapsing and benign forms of multiple sclerosis. *BMC Neurology* 11:64.
- [15] Shippy R., Smentek SF., Jensen RV., Jones WD., Wolber PK., Johnson CD., Pine PS., Boysen C., Guo X., Chudin E., Sun YA., Willey JC., Mieg JT., Mieg DT., Setterquist RA., Wilson M., Lucas AB., Novoradovskaya N., Papallo A., Turpaz Y., Baker SC., Warrington JA., Shi L., and Herman D., (2006). Using RNA sample titrations to assess microarray platform performance and normalization techniques. *Nat.Biotechnol.*, 24,1123–1131.
- [16] Steffen D., Paul TS., Ewan B., and Wolfgang H., (2009). Mapping identifiers for the integration of genomic datasets with the R / Bioconductor package biomaRt. *Nature journal.* 4 : 1184-1191.
- [17] Ulitsky I., Krishnamurthy A., Karp RM., Shamir R., (2010) DEGAS: De Novo Discovery of Dysregulated pathways in Human Disease. *PLoS ONE* 5(10)
- [18] Wen X., Fuhrman S., Michaels GS., Carr DB., Smith S., Barker JL., and Somogyi R., (1998). Large-scale temporal gene expression mapping of central nervous system development. *Proc.Natl Acad.Sci.* 95: 334-339.
- [19] Workman C., Jensen LJ., Jarmer H., Berka R., Gautier L., Nielser HB., Saxild HH., Nielsen C., Brunak S., and Knudsen S., (2002). A new non-linear normalization method for reducing variability in DNA microarray experiments. *Genome Biology* 3.