Analysis of gene interaction networks in neuronal aging and HIV infection

DOCTOR OF PHILOSOPHY

By

Pankaj Singh Dholaniya

Department of Biotechnology and Bioinformatics
School of Life Sciences
University of Hyderabad
PO Central University, Gachibowli
Hyderabad – 500 046, Telangana, India

September 2016

Analysis of gene interaction networks in neuronal aging and HIV infection

A thesis submitted to University of Hyderabad
for the award of a Ph.D. Degree in Biotechnology and Bioinformatics

By

Pankaj Singh Dholaniya

Department of Biotechnology and Bioinformatics
School of Life Sciences
University of Hyderabad
PO Central University, Gachibowli
Hyderabad – 500 046, Telangana, India

Enrollment No: – 11LTPH09
September 2016

UNIVERSITY OF HYDERABAD
(A central university established in 1974 by an act of Parliament)
Department of Biotechnology and Bioinformatics
School of Life Sciences
University of Hyderabad, Hyderabad 500 046, India

CERTIFICATE

This is to certify that thesis entitled "**Analysis of gene interaction networks in neuronal aging and HIV infection**" is a record of *bona fide* work done by **Mr. Pankaj Singh Dholaniya**, a research scholar for Ph.D. program in Department of Biotechnology and Bioinformatics, University of Hyderabad under my guidance and supervision.

The thesis has not been submitted previously in part or full to this or any other university or institution for the award of any degree or diploma. I recommend his thesis for submission towards the partial fulfillment of Doctor of Philosophy degree in Biotechnology.

Supervisor	**Head** Department of Biotechnology and Bioinformatics	**Dean** School of Life Sciences

UNIVERSITY OF HYDERABAD
(A central university established in 1974 by an act of Parliament)
Department of Biotechnology and Bioinformatics
School of Life Sciences
University of Hyderabad, Hyderabad 500 046, India

DECLARATION

I, Pankaj Singh Dholaniya, hereby declare that the work presented in this thesis, entitled as "**Analysis of gene interaction networks in neuronal aging and HIV infection**" has been carried out by me under the supervision of **Prof. Anand K. Kondapi**, Department of Biotechnology and Bioinformatics. To the best of my knowledge this work has not been submitted for the award of any degree or diploma at any other university or institution. I hereby agree that my thesis can be deposited in Shodganga/INFLIBNET. A report on plagiarism statistics from the University Librarian is enclosed.

Place: Hyderabad
Date:

Pankaj Singh Dholaniya
11LTPH09

Dedicated to My Parents

Acknowledgement

"Ph.D. is a training of being an independent researcher."

This was the thing I hear from my supervisor during my first conservation with him. First and foremost, I express my heartfelt gratitude and utmost respect to my supervisor, **Prof Anand K. Kondapi**, who has supported me throughout my work with his patience and guidance. His critical inputs and fruitful discussions have always been a great support throughout my research work.

I offer my gratitude to my doctoral committee members Dr J.S.S. Prakash and Dr. Vaibhav Vindal for their constant support. I am thankful to present and former Dean, School of Life sciences, for giving me the opportunity to use necessary facilities to carry out my work. I am also equally thankful to present and former Head, Department of Biotechnology and Bioinformatics for their support.

I would also like to thank Prof. Bapi Raju (former faculty in DCIS, UoH, now in IIIT hyderabad) for his valuable suggestions. I thank Prof. Arndt Telshow, University of Münster, Germany, and Prof. Appa Rao for giving me an opportunity to join student exchange program of UoH and Münster University. This program proved to be very much useful for me to learn new things releted to my research work.

I thank my lab mates Dr. Preeti, Dr. Anil, Dr. Satish, Dr. Bhaskar, Dr. Kishore, Dr. Balakrishna, Dr. Upendhar, Dr. Sarada, Dr. Farhan, Dr. Venkanna, Kurumurthy, Godan, Harikiran, Sonali, Srujana (Late), Jagadeesh, Kiran, Akhila, Suresh, Srujana, Kushagra, Ajay, and Kriti for their co-operation, support and cheerful nature all through my research. I thank Mr. Soumitra Ghosh from School of Computer and information Sciences for helping me during my project. I thank Mr. Srinivas Murthy, Mr. Sreenu, Mr. Bhanu, Mr. Chandra for their cooperation in lab. I thank Mr. Rajesh, from Bioinformatics Infrastructure Facility, for his support in server related issues. I thank to Dr. Arun, Paparao, Sanmukha, Kiranmayi and Pavani for having a wonderful company during my stay in Germany.

I feel blessed to have a wonderful friend circle with Angamba, Prateek, Nidhi, Tanvi, and Kalyani who made each moment joyful with lots of fun and late night parties. The circle extends with kind and caring friends Deepak, Babita, Prashant and Sanjay for being there with me in every part of time. Special thanks to Prashant for serving wonderful tasty food several times and also for his gentle support in my laboratory work. I am also thankful to the political group

of the campus to provide me an opportunity to contest the students' union election. It was a nice experience to work with them.

I thank all the faculty members of School of Life Sciences for their cooperation and extended support during my work. I am highly thankful to DBT, Government of India for providing the funds and fellowship to carry out my research work successfully.

I wish to express my deepest gratitude to my parents, brothers and sister.

Lastly, I offer my regards to all of those who supported me in any aspect during my work and all those who are connected to me during my stay at University of Hyderabad.

- Pankaj Singh Dholaniya

Abbreviations

AIC	Akaike information criterion
AIDS	Aquired Immunodeficiency Syndrome
ARM	Association Rule Mining
CGN	Cerebellar Granule Neurons
DA	Dopaminergic
DEq	Differential Equations
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribosenucleic Acid
EBI	European Bioinformatics Institute
FBS	Fetal Bovine Serum
HIV	Human Immunodeficiency Virus
IAEC	Institutional Animal Ethical Committee
JAK	Janus Kinase
MEM	Minimum Essential Media
NCBI	National Center for Biotechnology Information
PCR	Polymerase Chain Reaction
PD	Parkinson's Disease
PDL	Poly-D- Lysine
RNA	Ribosenucliec Acid
RT-PCR	Real Time Polymerase Chain Reaction
SNPc	Substantia Nigra Pars compacta
SOCS	Suppressor of Cytokine Signaling
STAT	Signal Transducer and Activator of Transcription
Top2b	Topoisomerase 2 beta
VM	Ventral Mesencephalon

Chapter 1

Introduction

1.1 Background:

Ever since the human race has evolved and started living a social life, it has always been challenging to live a healthy and disease-free life. Reaching back to Ayurveda and ancient Egyptian medicine, we have continuously improved our understanding of the human body and the way it functions. In order to keep ourselves healthy, we are exponentially increasing our knowledge about living biological system and its interaction with the surroundings. It includes the knowledge about the various natural and synthetic products used for the purpose of improving human health and curing various diseases. The desire of living a healthy and disease free life has slowly allowed us to move into a well-established field of science called as Life Science. Life science is a study of the biological system and its phenomenon. The key concept is that the more we understand the biological processes, the better we can combat the diseases. Understanding any biological process at cellular and molecular level is a fundamental step in life science.

The invention of the microscope was a major breakthrough in science when scientist started observing things a microscopic level. It was the 17th century when Robert Hooke, by making use of a light microscope, proposed his cell theory and described it as the basic unit of life. The scientists now started to observe the biological processes at the cellular level and were more interested in studying the basic facts about cell behavior such as cell division and sexual reproduction. In the 19th century, Gregor Mendel came up with a completely new concept by linking the phenotypic traits of living organism with the "factors" (now widely known as *genes*) located inside the cells. His findings provided a new direction for studying life as he proposed the basic principle of genetics. Later in the 20th century in 1953 James D. Watson and Francis H. Crick proposed the molecular structure of DNA (Watson & Crick, 1953b, a), which proved to be another milestone in biological science. This outstanding discovery accelerated the research in the field of molecular and cellular biology. *Deoxyribose Nucleic Acid* (DNA) is a molecule that carries genetic information inside the nucleus of the cell. DNA is responsible for growth, development, reproduction and function in all living organisms. DNA is made up of two strands coiled on each other, and each strand consists of four nucleotide base namely as Adenine (A),

Cytosine (C), Guanine (G) and Thymine (T). The backbone of the strand is made up of the five carbon phosphate-sugar deoxyribose. The two strands of DNA link to each other via base pairing in which A pairs with T with a double hydrogen bond and G pairs to C with triple hydrogen bonds. Three consecutive nucleotide bases serve as one codon, which codes for one amino acid in proteins. Proteins are macromolecules consisting of one or more long chains (known as polypeptide chain) of amino acid residues. One protein may have more than one polypeptide chains. Proteins perform a vast variety of functions in an organism, including metabolic reactions, transport molecules and response to any kind of external or internal stimuli. The structure of a protein is determined by its sequence of amino acid present in the polypeptide chain, and thus each protein differs from other protein by their arrangement of amino acids in a polypeptide chain. This sequence of the amino acid is determined by a specific region of DNA which is called as *a gene*. The synthesis of protein from DNA is termed as *the central dogma* of molecular biology, and it consists of two major steps. In first ribose nucleic acid (RNA) is synthesized using one strand of DNA as a template. This step is called *Transcription*. RNAs are single-stranded molecules also composed of four nucleotides. In place of Thymine (T), RNA contains Uridine (U). These RNA synthesized using DNA as templates are called messenger RNAs (mRNA) and are used as a template for Protein synthesis, the process known as *Translation*. Each gene in a particular cell type expresses a certain amount of mRNA, which may also increase/decrease with the time or under specific conditions. The protein expressed from one gene may affect the expression of other genes that means proteins also regulate the expression of other proteins when needed.

Therefore studying the gene expression profile of a cell under a given condition reveals a lot about cellular behavior. Studying gene expression pattern of a specific cell type in particular condition can help us to identify the gene interaction networks which can be considerably beneficial in the identification of key regulators for any biological process. Using high-throughput experiments such as Microarray we can measure the expression value of all the genes present in a genome and compare the differential expression pattern under two different conditions.

1.2 Introduction to Microarray Technology:

Functional genomics aims the analysis of a genome-wide dataset of information resulting from various biological experiments, for example, measuring the gene expression level of all the genes present in the genome under various biological conditions. DNA Microarray is one such technology which allows us to measure the genome-wide expression under a given biological condition in a single experiment. A DNA microarray also known as DNA chip is prepared by adhering DNA molecules on a solid surface, typically a glass slide, in an orderly manner. These DNA molecules are fixed in the form of spots. Each spot contains millions of copies of identical DNA that uniquely correspond to a particular gene (Schena *et al.*, 1995).

The advantage of using microarray technology is that it simplifies the process of comparative analysis of differential gene expression under different biological conditions. The process involves the extraction of mRNA from the cells. These mRNA were then used to make cDNA and then are spotted on the microarray plate, each spot representing one gene. These cDNA are then labeled with different color dyes such as green cy3 and red cy5. The microarray can be done in two ways, using single dye and two dyes hybridization. In single dye microarray, the cDNA from only one type of cells is hybridized onto the microarray chip while in two dye the cDNA of two different samples are labeled with two different color dyes and used for hybridization. Depending on the color intensity at each spot the expression of a particular gene is determined as upregulated or downregulated (figure 1).

On the basis of biological conditions microarray can be divided into two different categories: i) Time ordered microarray ii) Perturbation microarray. In time ordered microarray experiments the cells are grown for a certain time period, and the samples are collected at specific time intervals. While a perturbation microarray experiment compares the gene expression patterns of two or more cell cultures grown under different conditions or given some specific treatments. Time ordered microarray allows us to measure the expression patterns of genes in a particular cell type over a certain period of time and is useful in studying the time-dependent changes in the expression of genes for the corresponding biological process.

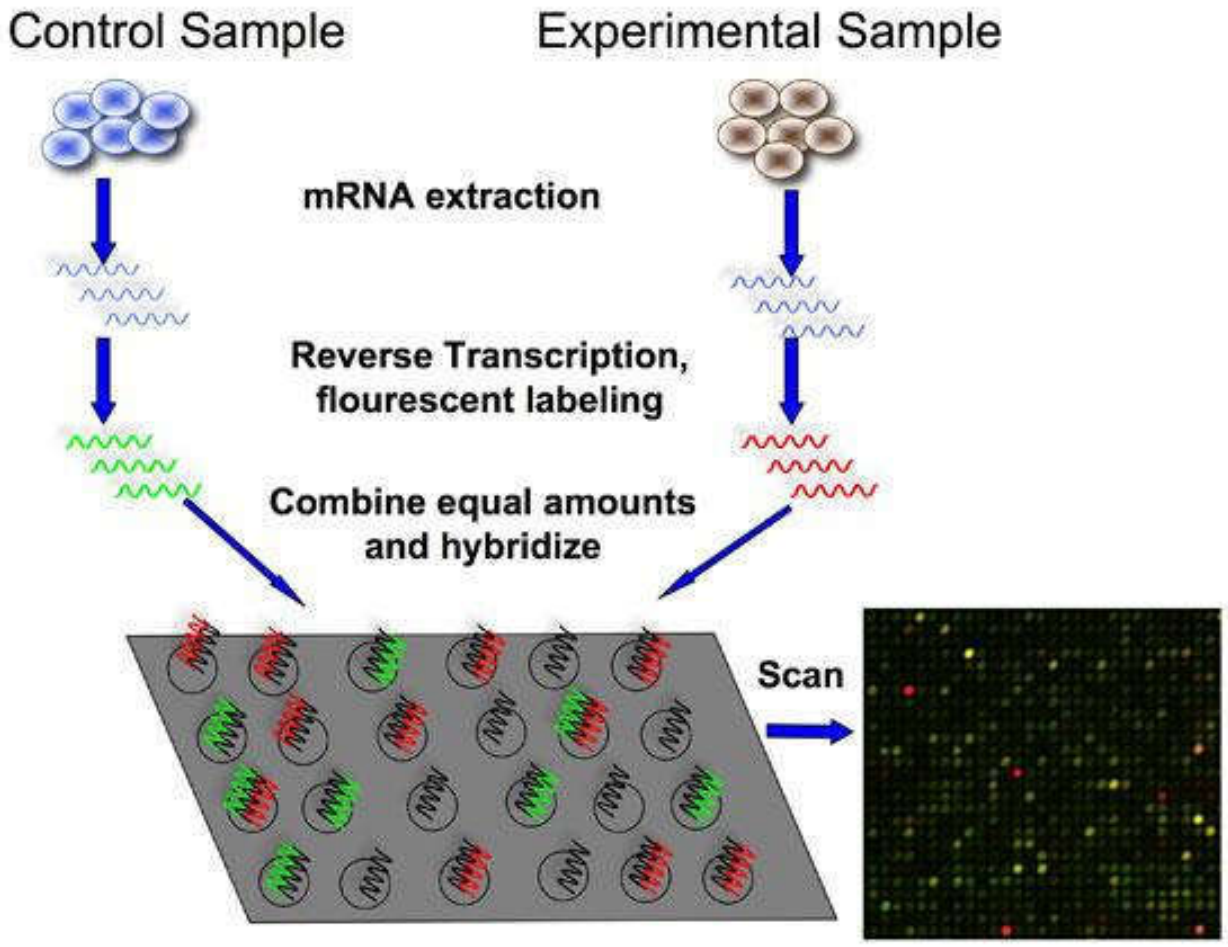

Figure 1.1: *Step involved in microarray technology showing the comparison of gene expression from two different types of cell conditions. (Source: http://bitesizebio.com/articles/introduction-to-dna-microarrays/).*

1.3 Gene network construction from microarray data:

Microarray generates a large set of data helping us to estimate functional association among genes in understanding cellular dynamics. The question is how this data can be employed to infer the functional associations among genes? Thus inference of gene interaction network from microarray data has become an essential step. The expression

data obtained from microarray experiment also called features can be used to discover the relationship between two genes. These features are used to design an approach or build a model to predict the relationship between the genes. Various approaches, such as Boolean networks, Bayesian networks, Differential equation models, Machine learning methods, have been described to infer gene interactions from expression data (de Hoon *et al.*, 2003, Eisen *et al.*, 1998, Friedman *et al.*, 2000, Imoto *et al.*, 2002, Soinov *et al.*, 2003).

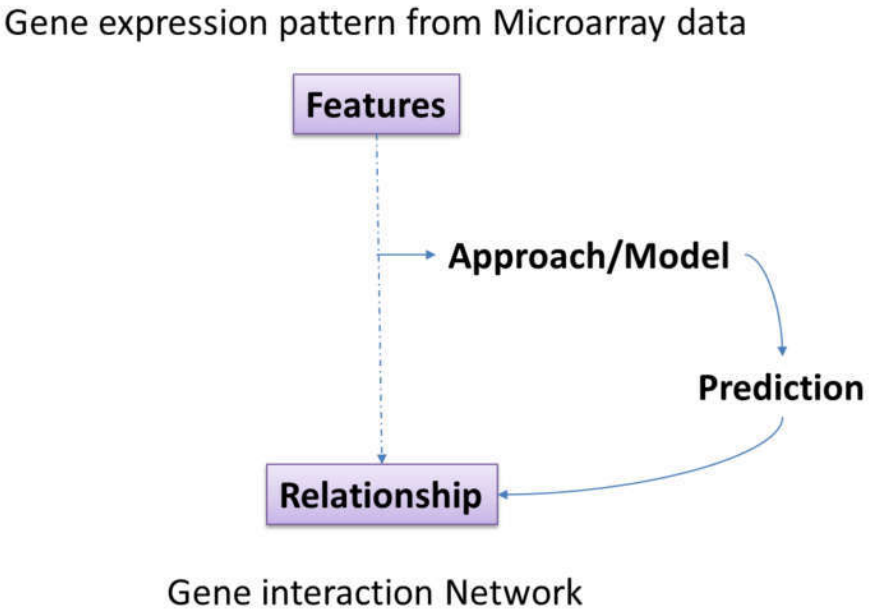

Figure 1.2: *Extraction of gene interaction from expression data. The relationship between any two genes depends on their expression patterns. Based on these expression patterns an approach or a model need to be developed which can be used to for the prediction of the relationship among genes.*

The simplest one is cluster analysis in which genes are grouped together based on the similarity between their gene expression profiles (Eisen *et al.*, 1998). These methods use either static or time series microarray data. The mathematical methods help in generating a dynamic model for gene regulatory networks (Hecker *et al.*, 2009). Recently, methods such as compressive sensing (Chang *et al.*, 2014), evolutionary algorithms (Thomas & Jin, 2014) and literature-based knowledge (Chen *et al.*, 2014) are being used for reconstruction of gene regulatory networks. Compressive sensing method takes advantage of the network's

sparseness; it assumes the biological networks are sparse networks which mean most of the genes have interaction with a small number of genes (Burda *et al.*, 2011). Evolutionary algorithms employ optimization algorithm based on Darwin's theory of evolution and make use of mechanisms involved in biological evolution such as reproduction, mutation, recombination, and selection (Khalid Raza, 2012). One benefit of the reconstruction methods involving evolutionary algorithms is due to the identification of both connectivity and parameters simultaneously from a given experimental data (Thomas & Jin, 2014). Chen et al. (Chen *et al.*, 2014) reported literature-based knowledge to infer gene regulatory network.

In the present study, we have implemented machine learning method to fetch the relationship among genes. We have used supervised learning method for time series data and unsupervised learning method for perturbation data.

1.3.1 Supervised learning method:

Supervised learning is also called classification or inductive learning. This is similar to human learning from past experiences to gain new knowledge in order to improve our ability to perform real-world tasks. This is a decision tree based classification method (Soinov et al., 2003). The key problem to address in finding the gene network is that the expression of which gene has an effect on the expression of other genes. The supervised learning algorithms learn from the expression data and identify patterns in the data. These patterns are then used to predict the expression or change in expression of a given gene from the change in expression of other genes. The predicted results are presented in the form of classifiers i.e. decision trees and decision rules. In a study by Soinov et al., (2003) used supervised learning to reconstruct the gene network from the microarray. The expression of a gene is defined by a finite number of states, the gene either expressed more than the average or expressed less than the average. The classifier is constructed to discriminate between these two states of the gene, which are labeled as upregulated and downregulated respectively. The state of one or more genes can change the state of another gene which can be described by a decision tree. The main advantage of using decision tree based classification is that they are easy to interpret.

1.3.2 Unsupervised learning method:

As described previously supervised learning method learns from the training data to predict the relationship among test data. Where, training data can also be called as labeled data because the predicting state is known for us. Unlike supervised learning method, unsupervised learning methods are used to describe the unlabeled data. Two major examples are clustering and associative learning. Clustering is a very useful technique which helps in the identification of different patterns lie in the dataset. Clustering has widely used to describe the gene expression data (Ben-Dor *et al.*, 1999, Qin, 2006). Clustering based on co-expression of genes provide a hint of dependencies of expression of genes on each other. For example, if two genes show the same pattern of expression across a number of samples it suggests that there are high chances that either of the genes regulates the expression of another gene, or the expression of both genes is regulated by some common factor. The conditions are also true if the two genes show exactly opposite pattern of expression i.e. high negative correlation.

Another popular example of unsupervised learning is associative learning or association rule mining. It was first introduced by Rakesh Agrawal (Agrawal *et al.*, 1993). Association rule mining (ARM) is the method for discovering the relationship among large dataset. Some of the algorithm used for generating rules are, Apriori, Eclat, and FP-Growth (Han *et al.*, 2004, Han *et al.*, 2000, Zaki, 2000, Jiawei Han, 2011).

These algorithms require an input dataset consisting of transactions and a record of items in a particular transaction find all frequent itemsets. The next step is to generate strong association rules from these frequent itemsets. Frequent itemsets are those item sets which occur at least as frequently as a predetermined minimum support count. Malik *et. al.* (Mallik *et al.*, 2015) introduced a Rank-Based Weighted Association Rule Mining algorithm and implemented on gene expression data. Usually, the apriori algorithm produces plenty of association rules, in order to reduce the number of rules Malik et. al. assigned a weight to each rule and ranked them to identify the high weighted rules, ultimately reducing the total number. These algorithms are widely acceptable for transactional data and market-basket analysis (Giudici, 2005). We have considered pathways as transactions and genes as items

which form the initial dataset. The association rules obtained from this algorithm elaborates the gene relationship based purely on pathway information and is used to form the gene network.

1.4 Problems and challenges:

The input for any computational model for gene network construction is the expression data which can be provided in the form of matrix e.g. rows represent the genes and columns represent the samples. In microarray experiments, we can gather expression data of a large number of genes (usually in thousands), but the number of experiments or samples are always limited. Thus the number of genes is always much larger than the number of samples. This kind of dimensionality is a common problem, when dealing with the analysis and organizing the data in high dimensional spaces and is popularly known as *"curse of dimensionality";* the term coined by Richard E. Bellman (Bellman, 1966). Regarding microarrays, this means that either we need a large number of observations i.e. more number of samples in order to obtain a good estimate of a model explaining the relationship between the genes or we need to reduce the dimension of the data by reducing the number of genes. One way to reduce the number of genes is by clustering the genes showing same patterns of expression across different samples into one cluster. Clustering can be effective and gives the relationship among the gene clusters (de Hoon *et al.*, 2003). But in such cases, the identity of the gene is lost and instead of identifying the interactive behavior among genes we can only study the interaction of gene clusters. Therefore the major drawback of a high dimension data is that it restrict us to study the gene interactions at the individual level and if we consider each gene as a separate entity then the chances of getting irrelevant interaction increases due to spurious correlation. This can be explained by an example; let us say that there are two genes which are showing the same pattern of expression across the sample, any computation model will treat them in a similar way and most probably will identify a significant relationship between them. As we know that the microarray experiment provides the expression data large set of genes together, typically thousands of genes, there is a high probability that the two genes showing the same pattern of expression may be completely biologically independent. In other words, the change in

the expression of one gene may not effect in the expression of another gene; thus it may lead to an irrelevant interaction (figure 3). We have handled this issue by adding the biological knowledge to the gene network reconstruction methods which I will be explaining in the later chapter.

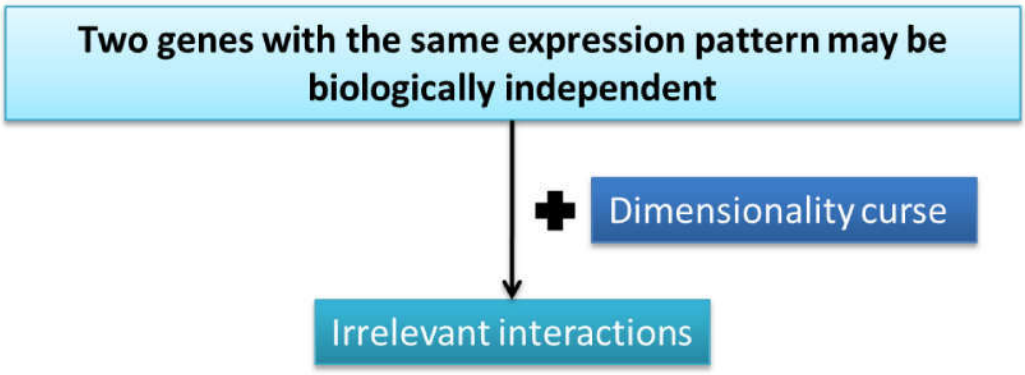

Figure 1.3: *the High dimension of data increases the chance of spurious correlations which led to the irrelevant interactions in the resulting gene interaction network.*

1.5 Work structure:

The objective of this work is to design an approach to minimize the search space from a large set of differentially expressed genes to a certain number of genes from which we can identify potential relevant targets.

Following are the steps we followed-

- Select DE genes
- Pathways categorization
- Reconstruction of gene networks
- Network analysis

We have worked on two kinds of microarray data – time ordered and perturbation. Under timer ordered data we have studied neuronal senescence and under perturbation microarray experiment we have studied HIV infection in the presence of curcumin and rotenone-induced Parkinson's disease.

The whole work was divided into three parts. The first part is the *analysis of time ordered microarray data* which include first objective (Chapter-2): "***Analysis of differentially expressed genes during neuronal senescence and identification of potential markers*.".** The second part includes *Top2b gene-associated regulatory model in neuronal activities and senescence, which is i*ncluded in the second objective (Chapter-3): "***Construction of hypothetical model showing the role of top2b in neuronal maturation and axon guidance*".** Third and last part of my work is to deal with perturbation microarray data, and it includes two short objectives (Chapter 4 and 5): "***Identification of crucial pathways involved in the anti-HIV activity of curcumin***" and "***Construction of gene interaction model in rotenone-induced PD in VM neurons***".

Chapter 2

Analysis of differentially expressed genes during neuronal senescence and identification of potential markers

2.1 Introduction

Senescence is a series of signaling events leading to gradual down-regulation of survival factors leading to deterioration of functional activities and recovery. Cellular senescence means gradual depletion in cellular function which ultimately leads to cell death. Cellular senescence was first described by Hayflick in 1965 as a process that limits the growth of cultured cells (Hayflick, 1965). CGNs are usually preferred to use to study the neuronal apoptosis or senescence/aging-related mechanism (Contestabile, 2002). To identify the genetic basis of neuronal aging has always been a challenge. Neuronal Aging is not caused by a single factor; there are several factors that affect neuronal aging such as i) Altered protein and molecular interactions. ii) Reactive oxygen species. iii) Signal transduction pathways. iv) Stress resistance mechanism and intracellular calcium homeostasis (Bishop *et al.*, 2010, Jellinger & Attems, 2013, Mattson & Magnus, 2006). Several studies in our lab have already demonstrated a crucial role of top2b in neuronal senescence (Bhanu *et al.*, 2010). To identify the potential genetic markers of neuronal senescence we have studied the neuronal aging model in Wistar rat. Indeed, Rats are the ideal animals to study neurocognitive aging (Gallagher *et al.*, 2011), and CGNs are most preferred model to study senescence *in vitro* (Bhanu *et al.*, 2010). We have identified the differentially expressed gene by performing microarray on the samples obtained from *in vitro* culture of cerebellar granule neurons (CGNs). Microarray provides a large set of differentially expressed genes. Several mathematical/computational models are available to reconstruct the gene networks from microarray data (as mentioned in the previous chapter), we have studied two models; first is Differential equation model explained by Chen *et. al* (1999) and the second is Supervised learning method as explained by Soinov *et. al* (2003).

Differential equation model is specifically used for the time series microarray data in which the expression pattern of gene can be described by a system of linear differential equations. Because the method is very sensitive to the dimension of the data we have first build the model on yeast cell cycle data and validated the accuracy of the model by predicting the expression of genes at given time points. In another method, we used supervised learning to reconstruct the gene network from the microarray data. This method predicts the

expression of genes based on expression of other genes. The prediction results are presented in the form of classifiers called as decision trees and decision rules.

2.2 Materials and Methods

We have performed microarray and then we studied two models for reconstruction of gene networks *viz.* Differential equation models and supervised learning method. The appropriate model was selected to reconstruct gene network, and the probable markers were identified and validated through experiments.

2.2.1 Reagents used:

Poly-D-lysine, DAPI (Sigma Chemical Co., MO, USA), Earle's balanced salt solution (EBSS), Bradford reagent, Trypsin, Glutamax (GIBCO, NY, USA), dNTPs (Fermentas, MD, USA), minimal essential medium (MEM), pure enzymes T4 DNA ligase (Invitrogen, NY, USA), T4 polynucleotide kinase (Bangalore Genei, India), penicillin streptomycin and fetal bovine serum (FBS)

2.2.2 Animals:

Wistar rats provided by National Institute of Nutrition, Hyderabad, India (NIN, Hyderabad) were maintained in accordance with the Institutional Animal Ethics Committee, in a pathogen-free environment with a 12-h light and darkness cycle. Sufficient Food and water were provided.

2.2.3 Cell culture: CGNs were isolated from 6-8 days old Wistar rat pups and culture in MEM media supplemented with 10% FBS and other essential components. Cells were plated in 60mm dishes and were grown at 37^{o}C under 5% CO_2. Cells were divided into four different sets and grown for four weeks. Each set of cells is harvested after an interval of 7 days. All four samples were collected and sent for the microarray to Genotypic Technology Bengaluru Pvt. Ltd. The steps involved in the microarray sample preparation are depicted in figure 2.1.

2.2.4 Microarray: Fold change was calculated by considering first week sample as control and other as treated. The log fold change values were calculated, and two models were selected for the study. The microarray data was submitted to ArrayExpress (EBI) and is freely accessible with accession ID **E-MTAB-4552**.

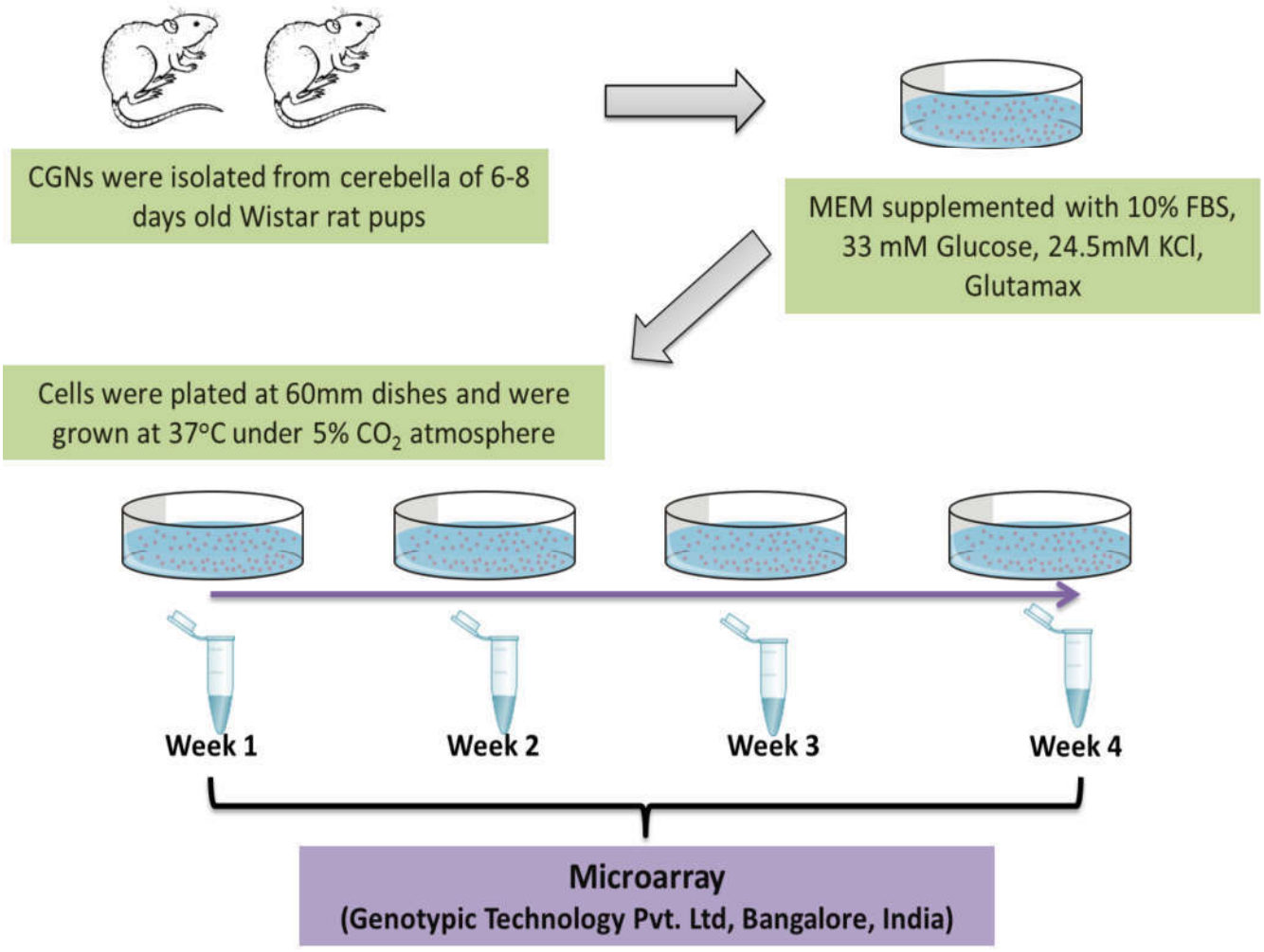

Figure 2.1: *Work plan for the preparation of samples for microarray*

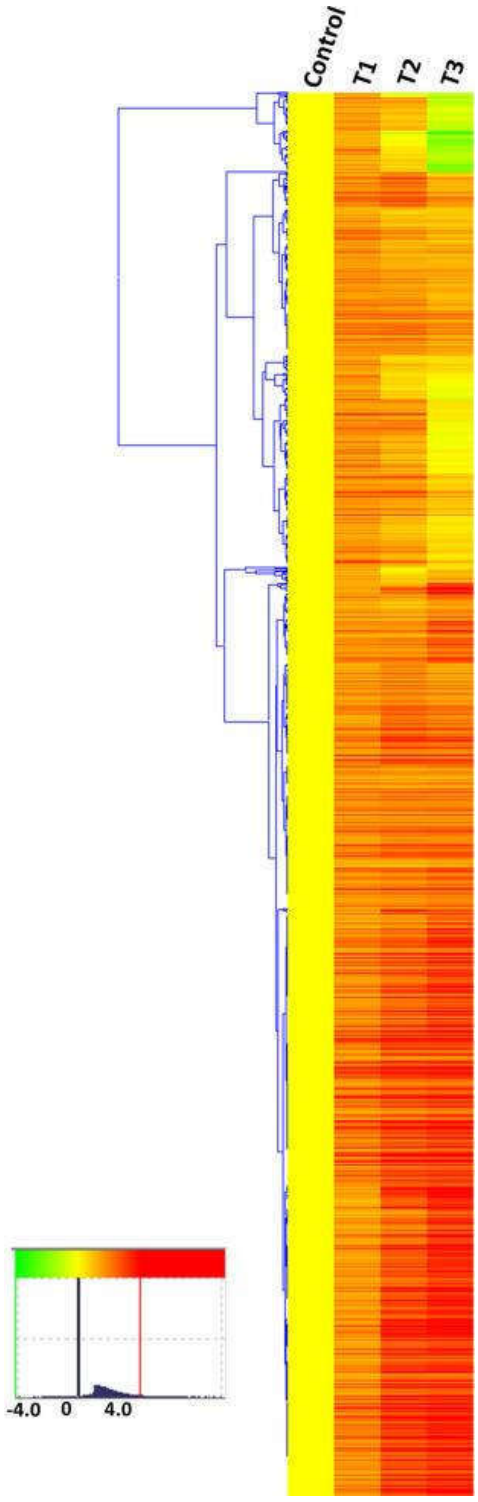

Figure 2.2: *Heat map showing the expression of differentially up-regulated genes. A heat map was generated using hierarchical clustering Explorer version 3.0. Fold change used for up-regulation was >0.8 and down-regulation was <−0.8 during the period of first to second week. The first week sample was taken as control. Second, third and fourth week samples were*

considered as treated denoted by T1, T2 and T3, respectively. In the figure, normalized expression signals are represented on a log scale for which green color corresponds to lower levels of expression and red color corresponds to higher levels of expression.

2.2.5 Data analysis and gene selection

Data analysis was performed using GeneSpring GX version 11.5 (Agilent Technologies) and Microsoft Excel. Genes with more than 0.8 and less than -0.8 fold change were considered upregulated and downregulated respectively. Geomean fold of greater than 1 is taken into account for replicated samples and P value <0.05 was considered statistically significant. There were total of 2773 differentially up-regulated genes. A heatmap showing the expression pattern of these genes is presented in Fig. 2.2. Differentially regulated genes were classified based on functional category and pathways using GeneSpring GX Software and Biointerpreter (Genotypic Technology Pvt. Ltd., Bangalore). For network analysis we have shortlisted the genes showing the upregulation more than 1.5 fold during first week. This search left us with a total of 157 genes falling under 108 pathways.

2.2.6 Network construction

2.2.6.1 Differential equation model

Differential equation model (DEq) was first described by Chen *et. al (Chen et al., 1999, de Hoon et al., 2003)* to reconstruct gene network from microarray data. The change in expression over time *dt* can be explained by the following equation.

$$\frac{\mathrm{d}}{\mathrm{d}t}\underline{x}(t) = \underline{\underline{\Lambda}} \cdot \underline{x}(t) \qquad \text{Eq –(1)}$$

Where matrix lambda is a weight matrix showing the effect of each gene on every other gene. A matrix element Λ_{ij} represents the effect of gene j on gene i $[\Lambda_{ij}]^{-1}$ being the reaction time.

Where vector x(t) represents the expression level of a gene at certain time t. Matrix $\underline{\underline{\Lambda}}$ is constant with units of [second]$^{-1}$. A matrix element $\underline{\underline{\Lambda}}_{ij}$ represents the effect of gene j on gene i $[\underline{\underline{\Lambda}}_{ij}]^{-1}$ being the reaction time.

As from the microarray data, we get the expression of genes at discrete time periods; we can replace the differential equation (eq.1) to by a difference equation.

$$\frac{\Delta \underline{x}}{\Delta t} = \underline{\underline{\Lambda}} \cdot \underline{x} \qquad \text{Eq -(2)}$$

or

$$\underline{x}\,(t + \Delta t) - \underline{x}\,(t) = \Delta t \cdot \underline{\underline{\Lambda}} \cdot \underline{x}\,(t) \qquad \text{Eq – (3)}$$

The matrix $\underline{\underline{\Lambda}}$ may contain positive or negative values and zeroes. A zero element represents no influence of a gene on other which means no edge between two genes. Since the gene networks are considered to be sparse, most of the value in the matrix $\underline{\underline{\Lambda}}$ should be equal to zero. Hoon (de Hoon *et al.*, 2003) used an error ε to be associated with the expression data which is considered to normally distributed independent of time with the standard deviation σ and calculated the maximum likelihood of variance (Eq- 4) and the weight matrix $\underline{\underline{\Lambda}}$ (eq- 5).

$$\hat{\sigma}^2 = \frac{1}{nm} \sum_{i=1}^{n} \Big[\left(\underline{x}_i^{\mathrm{T}} - \underline{x}_{i-1}^{\mathrm{T}}\right) \cdot \left(\underline{x}_i - \underline{x}_{i-1}\right) + \left(t_i - t_{i-1}\right)^2 \underline{x}_{i-1}^{\mathrm{T}} \cdot \underline{\underline{\Lambda}}^{\mathrm{T}} - 2 \left(\underline{x}_i^{\mathrm{T}} - \left(t_i - t_{i-1}\right) \underline{x}_{i-1}^{\mathrm{T}}\right) \cdot \underline{\underline{\Lambda}} \cdot \underline{x}_{i-1} \Big]$$

Eq – (4)

and

$$\hat{\underline{\underline{\Lambda}}} = \underline{\underline{B}} \cdot \underline{\underline{A}}^{-1} \qquad \text{Eq – (5)}$$

Where,

$$\underline{\underline{A}} \equiv \sum_{i=1}^{n} \left[(t_i - t_{i-1})^2 \cdot \underline{x}_{i-1} \cdot \underline{x}_{i-1}^{\mathrm{T}} \right]$$

$$\underline{\underline{B}} \equiv \sum_{i=1}^{n} \left[(t_i - t_{i-1}) \cdot (\underline{x}_i - \underline{x}_{i-1}) \cdot \underline{x}_{i-1}^{\mathrm{T}} \right]$$

We have written a script in R package to calculate the σ^2 and $\underline{\Lambda}$ from the expression (Progam 1 and Program 2). The weight matrix calculated using above method may contain all the non-zero elements. To make gene network sparse we must have zero elements in the gene connectivity matrix. Therefore Hoon used **Akaike's Information Criterion** (AIC) (Akaike, 1974) to determine which element of the matrix should be kept nonzero. Akaike's Information Criterion is an approach used for the statistical model selection. The model with the lowest AIC is considered to be most favorable.

$$\mathbf{AIC} = -2 \left[\begin{matrix} \mathbf{log}\ \boldsymbol{likelihood\ of\ the} \\ \boldsymbol{estimated\ model} \end{matrix} \right] + 2\ [\boldsymbol{number\ of\ estimated\ parameters}]$$

Eq – (6)

Program 1.

- **R program to calculate weight matrix from microarray data (p1.R)**

```
maindataset <- read.table(file = "data_genes_15.txt")
A=0
for (i in 1:24){
     y1 <- matrix(c(maindataset[,i]), ncol=1, byrow=T)
     z1 = t(y1)
     p1 = 10*10*(y1%*%z1)
```

```
     A = A+p1
}
AI <- solve(A)
B=0;
for (i in 2:24){
     j=i-1
     x2 <- matrix(c(maindataset[,i]), ncol=1, byrow=T)
     y2 <- matrix(c(maindataset[,j]), ncol=1, byrow=T)
          dx <- (x2-y2)
          z2 <- t(y2)
          p2 <- 10*(dx%*%z2)
          B = B+p2
     }
     W <- B%*%AI
write.table(W,"wtMat_data_15.txt",quote=FALSE, sep="\t",
row.names=FALSE)
```

Program 2.

- **R program to find the minimum AIC (p2.R)**

```
maindataset <- read.table(file = "data_genes_15.txt")
wtMat <- read.table(file = "wtMat_data_15.txt", header=TRUE)
m = 15
n = 24
tm = 10
AICMat <- matrix(0:0, nrow = m, ncol = m)
MNew <- matrix(0:0, nrow = m, ncol = m)
     for(steps in 1:(m*m)){
          for(rows in 1:m){
               for(cols in 1:m){
                    M <- matrix(0:0, nrow = m, ncol = m)
                    M <- M + MNew
                    M[rows,cols] = 1
                    wtNew = wtMat*M

                    # Calculation for SIGMA SQUARE
                    wtNew <- as.matrix(wtNew)
                    wtNew_t = t(wtNew)
                    A = 0
```

```
                    B = 0
                    C = 0
                    S = 0
                    for(p in 2:24){
                         p2 =p-1
                         xi <- matrix(c(maindataset[,p]),
ncol=1, byrow=T)
                         xi_1 <- matrix(c(maindataset[,p2]),
ncol=1, byrow=T)
                         xi_t <- t(xi)
                         xi_1_t <- t(xi_1)
                         #temp = ((xi_t - xi_1_t)%*%(xi -
xi_1))+ (tm * tm * xi_1_t %*% wtNew_t %*% wtNew %*% xi_1) - (2 *
(xi_t - (tm*xi_1_t)) %*% wtNew %*% xi_1)

                         A = (xi_t - xi_1_t)%*%(xi - xi_1)
                         B = tm * tm * (xi_1_t %*% wtNew_t %*%
wtNew %*% xi_1)
                         C = 2 * (xi_t - (tm*xi_1_t)) %*% wtNew
%*% xi_1

                         temp = (A + B - C)
                         S = S+temp
                    }
                    sigma = S/(n*m)
                    #print(S)

                    # Calculation for AIC
                    sum_of_mask = sum(M)
                    AIC = ((n * m * log(2*pi*sigma)) + (n * m) +
2 * (1 + sum_of_mask))
                    AICMat[rows,cols] = AIC
               }
          }
          minCords <- which(AICMat == min(AICMat), arr.ind=TRUE)
          minRow = minCords[1,1]
          minCol = minCords[1,2]
          MNew[minRow,minCol] = 1
     }
filename <- sprintf("resMat_data_15.txt")
write.table(MNew, filename, quote=FALSE, sep="\t",
row.names=FALSE)
```

A binary Mask matrix M is used to set the element of $\underline{\underline{\Lambda}}$ equals to zero. This calculated maximum likelihood of the $\underline{\underline{\Lambda}}$ was multiplied with mask matrix M and the resultant matrix was used to calculate AIC.

$$\hat{\underline{\underline{\Lambda}}}' = \underline{\underline{M}} \circ \hat{\underline{\underline{\Lambda}}}, \qquad \text{Eq – (7)}$$

There may be many possible ways to calculate the optimum mask matrix M which gives the minimum AIC. For n number of genes, total possible binary M matrices are 2^{n*n}. So even if a small number of genes such as five the possible number of solutions goes very high which is 2^{25}. Hoon started from the random initial mask matrix and calculated AIC. The experiment was repeated 1000 times. This method is good enough for the small number of genes. But if we have more genes for example in our case we started with 25 genes and the total possible solution are 2^{625}, starting from random matrix will be difficult and we may not achieve the minimum.

We have used a slightly different approach to find the mask matrix which gives the minimum AIC. We started from the matrix with all elements equals to zero. Now keeping element one by one nonzero starting from position (1,1) to (n,n) AIC was calculated. The position which gives the minimum AIC was fixed to be nonzero and the step repeated n^2 times (Program 2).

Dataset:

We have calculated Λ matrix from the expression data and identified the best model using Akaike's Information Criterion. The model was first developed on yeast cell cycle gene expression data, downloaded from http://genome-www.stanford.edu/cellcycle (Spellman *et al.*, 1998). We have taken the cdc15 dataset; which contains 24 time points. The experiment was performed on 15 genes (Table 2.1).

	cdc15_10	cdc15_30	cdc15_50	cdc15_70	cdc15_80	cdc15_90	cdc15_100	cdc15_110	cdc15_120	cdc15_130	cdc15_140	cdc15_150
CTS1	-2.61	1.04	0.11	-1.94	-1.69	-2.48	-2.13	-2.94	-1.77	1.22	2.44	2.16
YRO2	0.31	-1.48	-3.16	-1.6	-0.2	0.1	0.98	1.36	2.14	1.02	0.76	-0.46
RME1	-0.32	1.63	0.68	-1.25	-1.6	-1.74	-1.79	-1.44	-0.24	1.64	1.35	1.63
PMS1	0.27	1.25	0.84	-0.26	-0.22	-0.44	-0.26	-0.17	0.27	0.81	0.64	0.68
CWP1	-1.85	-2.36	0.08	0.57	1.36	0.59	0.22	-0.46	-0.26	-0.59	0.48	0.83
PSA1	-2.35	-0.38	0.91	0.56	0.83	-0.04	-0.42	-1.54	-1.42	-0.65	0.4	1.2
AGA1	-0.78	-0.45	-1.15	-1.1	-1.32	-1.03	-0.62	-0.43	-0.1	0.62	0.4	-0.13
CDC5	0.03	-1.7	-1.2	0.98	0.44	0.64	0.17	1.54	0.05	-0.03	-1.3	-1.25
CLN2	-1.26	1.6	1.54	0.31	-0.14	-0.88	-1.7	-1.88	-1.7	0.61	1.54	1.72
CLB6	-1.25	1.93	0.76	-1.06	-0.45	-1.59	0.09	-1.63	-0.34	0.49	1.6	1.25
ASH1	1.48	1.87	-1.31	-1.73	-2.23	-1.78	-1.16	0.54	1.34	1.82	0.92	0.31
HO	-0.28	0.22	-0.12	-0.12	-0.11	0.17	-0.46	0.35	-0.25	-0.09	0.55	0.62
SWI4	0.25	0.45	0.16	0.08	-0.42	0.03	0.01	0.01	-0.04	0.59	0.13	0.22
CDC8	-0.51	-0.2	-0.03	0.08	0.2	0.12	0.06	0.07	0.04	0.02	-0.01	0.11
TIR1	-0.3	-0.13	-0.2	1.41	0.38	0.07	0.21	0.09	-0.12	-0.08	-0.52	0.65

	cdc15_160	cdc15_170	cdc15_180	cdc15_190	cdc15_200	cdc15_210	cdc15_220	cdc15_230	cdc15_240	cdc15_250	cdc15_270	cdc15_290
CTS1	2.32	1.08	0.91	-0.54	-0.41	-1.35	-0.85	-0.73	1.8	1.8	2.32	2.24
YRO2	-0.73	-1.67	-1.39	-0.52	0.75	0.61	1.51	0.93	1.44	0.19	-0.14	-0.75
RME1	0.89	0.68	-0.6	-0.74	-1.36	-0.97	-0.81	0.21	0.11	1.62	1.33	1.12
PMS1	0.38	0.52	-0.01	-4.63	-0.19	-0.54	0.21	0.38	0.21	0.16	0.11	-0.01
CWP1	1.12	0.66	0.7	0.5	0.82	0.08	0.24	-1.02	-0.73	-0.63	-0.29	-0.07
PSA1	1.27	1.16	1.02	0.82	-0.18	0	0.04	-0.72	-1.31	-0.18	0.36	0.61
AGA1	-0.72	-0.84	-1.03	-1	-0.57	-0.35	0.48	0.7	1.16	2.45	2.83	3.01
CDC5	-1.7	0.03	-0.32	1.13	0.53	1.31	0.6	0.89	-0.46	0.2	-0.37	-0.23
CLN2	1.51	1.18	0.84	-0.26	-0.34	-1.03	-0.85	-0.85	-0.21	0.13	0.13	0
CLB6	0.86	0.7	0.28	-0.92	-0.25	-1.12	-0.01	0.55	0.88	-0.11	-0.3	-0.32
ASH1	-0.55	-0.53	0.58	-1.48	-1.44	-0.48	-0.46	1.31	1.48	1.23	0.48	-0.23
HO	0.58	-0.03	-0.44	-0.48	-0.32	-0.21	0.28	0.15	0.36	-0.57	0.1	-0.21
SWI4	-0.24	0.17	-0.49	-0.14	-0.59	-0.04	-0.47	-0.49	0.01	0.3	0.3	0.24
CDC8	-0.01	0.08	0.32	-0.28	-0.15	0	0.43	0.39	0.07	-0.27	-0.39	-0.15
TIR1	0.02	-0.02	0.14	-0.41	-0.22	-0.01	0.39	-0.72	0.33	-0.36	-0.22	-0.38

Table 2.1

2.2.6.2 Supervised learning method

In the present study, we have implemented the decision tree based supervised learning method (Soinov *et al.*, 2003) which is also known as classification or inductive learning. This is comparable to human learning as humans learn from their past experiences to gain new knowledge in order to improve the ability to perform real-world tasks. This is a decision tree based classification method (Soinov *et al.,* 2003). We have used the C4.5 algorithm to generate the decision tree and then the decision rules were extracted to construct the gene network. (List of decision rule is given in Appendix I). Differentially expressed genes were selected and pathways classification was done. In order to reduce the dimension of the data, we have grouped the genes based on their biological function

and then applied supervised learning method to construct gene interaction network for each group. All these sub-networks were then merged into a single large network based on multipath genes. The modular analysis of the resulting network was performed using Cytoscape plugin called Moduland (Shannon *et al.*, 2003, Szalay-Beko *et al.*, 2012). The detailed list of gene in each module and their corresponding pathways is provided in Appendix II.

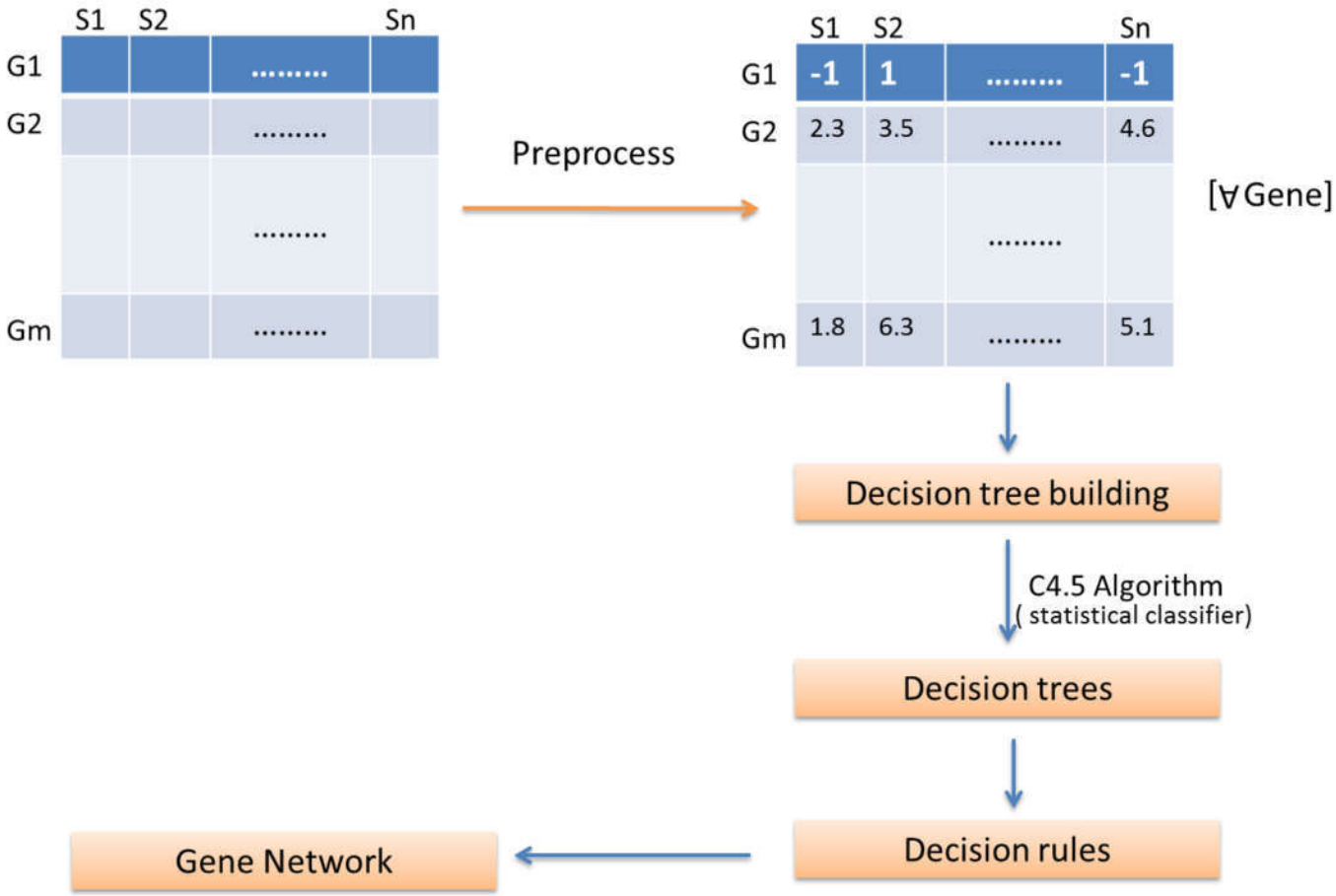

***Figure 2.3**: Steps involved in the extraction of decision rule from expression data using the C4.5 algorithm.*

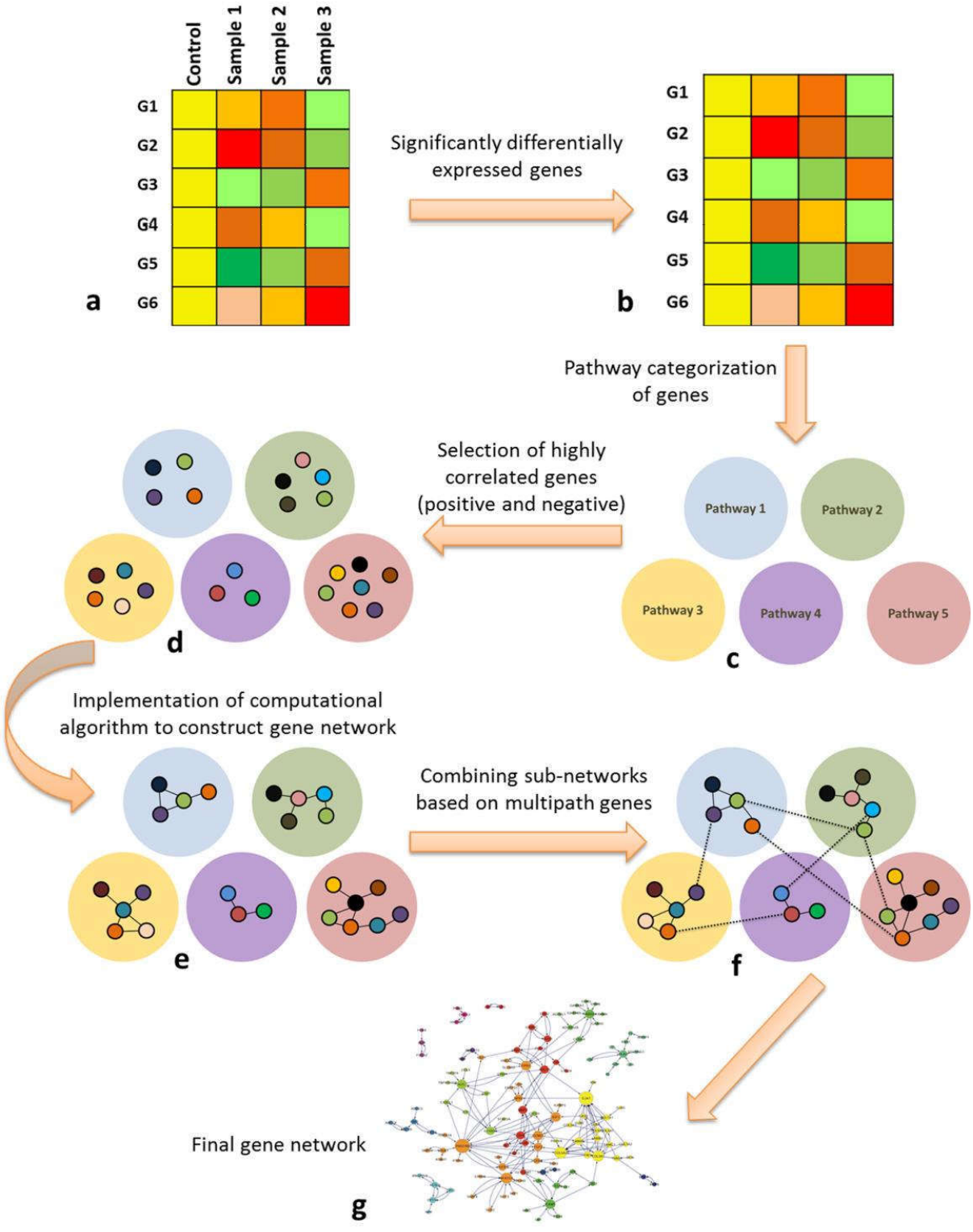

Figure 2.4*: Schematic representation of the flow of methodology. (a) Microarray data set. (b) Selection of significantly differentially expressed genes. (c) Genes are classified into pathways.*

(d) Genes within each pathway are shortlisted based on the correlation coefficient among other genes of the same pathway. Gene pairs having high positive and negative correlation coefficient are chosen. (e) The supervised learning method is used to extract the relationship among genes with each pathway. This step gives a sub-network for each pathway. (f) All the sub-networks are merged using multipath genes. (g) Resulting gene interaction map is obtained. The classifier is constructed to discriminate

2.2.6 Real-Time PCR:

The microarray data was validated by quantitative real-time PCR (qPCR). 10µL of cDNA synthesis was carried out using 1µg of total RNA, random hexamers and Super ScriptTM First-Strand Synthesis System (Invitrogen, NY, USA) using gene specific primers as in Table 2.1 (Eurofins Genomics, Bangalore, India) and Power SYBR Green PCR Master Mix (Applied Biosystem, CA, USA). The absence of genomic DNA in total RNA was ensured by DNaseI treatment (Fermentas, GmbH, Germany). 18s rRNA was used as an internal control. To ensure no contamination of PCR reagents with the cDNA, no template controls (NTC) were set up. PCR analysis was performed with ABI Prism H7500 fast thermal cycler (Applied Biosystem, CA, USA). Each sample was run in triplicate in a final volume of 10µL containing 0.2µL of 1st strand cDNA template, 20 pmol of each primer, and 5µL of Power SYBR Green PCR master mix. Fluorescence resulting from DNA amplification was analyzed. Relative fold change was assayed using $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen, 2001).

S. No	Gene	Forward primer (5' →3')	Reverse primer (5' →3')	Amplicon size
1	UNG	CGC ACA CCA AGC CAA TTC CCA TAA	AGG TTC TGA TTC AGC CAC GAC ACA	81
2	SMUG	TTT CCA GAG CCT GTG GGT GTC ATC TA	TTG GCA GTA GCG AGT CAC GTA GTT	84
3	APEX	GTT CTT CCT CAC CAA TGC CAT AAG AG	AGG CTT GGA TTG GGT AAA GGA AGA AGC	215
4	POL B	CAC AGC TCA ATG GCA CCT AAC	AGT GAC CAG ACG CTG TGA TG	220
5	POL M	AGA GGT GAC ACA TGT GGT GAT GGA	TCC TGC TGC CAT GCT CTC TGT AAA	142

6	POL L	AGG GTT CCT CAC AGA TGA CTT GGT	AAA CTC ACT GTA GGG CAC CAC GAT	136
7	LIG 1	AGA CAG CAG AGG CCA GAA AGA TGT	TCT GGG AGA CTT TCC AAG CCA TGT	130
8	LIG 3	TAC TGG AGG CAG CAA TGG TGA GAA	TTC TTG GCA CTG GCA GAG GAC TTA	102
9	TOP2B	GAC AGA GGA AGG TAG TAG AGC CTG	CGT TTT CTT CGG TTT CTT GCT GGC	168
10	A2M	TCA CTC ATC CTG TTG TCC GCA ATG CCC TCT	ACC AGC AAG GGC AAA TGC ATA GGC CAA CA	128
11	GNA14	ACC AAA GCA AGA TGT CAA AGC TGC CAG GGA	TCC TTG ACA GCA GCA AAC ACG AAG CGG AT	147
12	GRIA1	ATG CCA ACC AGT TTG AGG GCA ATG ACC GCT	TTC TCC CAC CAT GCC ATT CCA AGC CTT TGT	167
13	MASP	TGC CGA GTG GAA TGC AGT GGC AAT CTC TT	AGG GCA CCT CGG GAT GGT CTT CAA TGT CAA	184
14	NPY	TGC TCG TGT GTT TGG GCA TTC TGG CTG A	ATC AGT GTC TCA GGG CTG GAT CTC TTG CCA	169
15	SLIT2	AGA ACG GCA CCA GCT TCC ATG GCT GTA T	TGG GCA CAC ACT TTC TTG TGG CAT GGT TCA	133
16	GLB	TCA AGG ATG GGC AGC CAT TCC GCT ACA T	TGG ATT GCA TCC AGC CCA GCC ATC TTC AT	118
17	18S	GCT ACC ACA TCC AAG GAA GGC AGC	CGG CTG CTG GCA CCA GAC TTG	200

Table 2.1: *Primer list for Real-time PCR*

2.2.7 Gene silencing studies: The network obtained from method mentioned above was analyzed and from the further literature survey, six genes (A2m, Gna14, Gria1, Masp1, Npy, and Slit2) were selected for further analysis. The expression patterns of these genes were validated by real-time PCR. Also, the gene silencing study was done with these genes using siRNA, and the viability of neurons was verified.

We have used double strand siRNA oligos for transient downregulation of indicated genes in rat CGNs. Lipofectamine 2000 (Invitrogen) was used for transfecting the double-strand siRNA oligos. Cultures, one day after plating, were used for transfection as standardized in our lab (Mandraju et al. 2008) (Table 3). Double-strand siRNA oligos were synthesized as described earlier (Donze and Picard 2002). For this, desalted DNA oligonucleotides were obtained from Sigma (India). The oligonucleotide-directed production of small RNA transcripts with T7 RNA polymerase has been implemented as described (Milligan and Uhlenbeck 1989). Sequences of SiRNA are given in Table 2.2.

For each transcription reaction, 1nM of each oligonucleotide was annealed in 50μL of TE buffer (10mM Tris–HCl, pH 8.0, and 1mM EDTA) by heating at 95 °C; after 2 min, the heating block was switched off and allowed to cool down slowly to obtain double-stranded DNA. Transcription was performed in 50μL of transcription mix: T7 transcription buffer (40mM Tris–HCl, pH 7.9, 6mM MgCl2, 10mM dithiothreitol (DTT), 10mM NaCl, and 2mM spermidine), 1 mM nucleotide triphosphates (NTPs), 0.1 U yeast pyrophosphatase (Sigma), 40 U RNaseOUT (Life Technologies), and 100 U T7 RNA polymerase (Invitrogen) containing 200pM of the dsDNA as template. After incubation at 37 °C for 2hr, 1 U RNase-free DNase (Genetix) was added at 37 °C for 15 min. Sense and antisense 21-nt RNAs generated in separate reactions were annealed by mixing both crude transcription reactions, heating at 95 °C for 5 min followed by 1hr at 37 °C to obtain 'T7 RNA polymerase synthesized small interfering double-stranded RNA' (T7 siRNA). The mixture (100μL) was then adjusted to 0.2M sodium acetate, pH 2, and precipitated with 2.5 volume of ethanol. After centrifugation, the pellet was washed once with 70 % ethanol, dried, and resuspended in 50μL of water. In the case of gene-specific knockdown studies, at three weeks after culturing, the neurons were treated with gene specific SiRNA (0.1–1 μM) using Lipofectamine-2000 (Invitrogen, cat. no. 11668) as transfection agent and also with 0.5μM of non-silencing siRNA (scrambled) as control. The effect of Topo IIβ or senescence-associated gene-specific siRNA on the viability of granule neurons in culture was determined using cell viability assay described above (Mosmann 1983). Cultured CGNs in poly-l-lysine coated 24 well plates after treatment at 3 weeks were assessed for cell viability and gene expression at fourth, fifth, and sixth weeks in culture.

Name	siRNA sense template sequence	siRNA antisense template sequence
TopoIIß siRNA	5'-AAA GCT TAA CAA TCA AGC CCG CTA TAG TGA GTC GTA TTA -3'	5'-AAC GGG CTT GAT TGT TAA GCT CTA TAG TGA GTC GTA TTA -3'
TopoIIß scrambled siRNA	5'-ACA CTC GAT CAA TCC AAG GAA CTA TAG GGA ATC GAA ATA -3'	5'-CAC TGG ATT GAT CGA GAT GTT CTA TAG TGA GTC GTA TTA -3'
SLIT 2 siRNA	5'-ATT GAT TCC CCA ACA ATT CAG TAG TAG TGA GTT ATA -3'	5'-AAT TGT CTA CAT CCA GAA TAT ATA TGG AGA GGC ATA ATA -3'
SLIT 2 scrambled siRNA	5'-TAT GGT TCA ACA CCA ATT CCG CAG CCG TTA ATT AAA TAA -3'	5'-CAC TGG ATT GAT CGA GAT GTT CTA TAG TGA GTC GTA CTA -3'
MASP siRNA	5'-GAA AAT CTA ATA CCA ATC TTG AAG AAG CAA ACC CGT ATA -3'	5'-AAC TAG CCT GAC AGC GAT ATT CTA AAG CAA GAC ATA GTA -3'
MASP scrambled siRNA	5'-AAT TAT TCC CCA TCA ACT CTG AAG TTG ATA ATT TTA CCT -3'	5'-GAC TAG AAT GAC AGC TAT ACG CTA CCG CAA GAC CTA CCA -3'
GLB siRNA	5'-CAT TTT TCC TTA TCA ACT CGG CAG TAG ACA AAT CTA GCT -3'	5'-TAC GAC TCT GAC TGC GGT CAT CAA AAG CGG AAC GTA GTA -3'
GLB scrambled siRNA	5'-ATC AAG CTT TAC AGC AAT CCT TTA AAG TGA AAC ATA TTA -3'	5'-CTT AGT AGA CCA TCAATT ACG CAG TTG ATA ATT CCT TAT -3'
GRIA1 siRNA	5'-AAT TAT TCC CCA TCA ACT CTG AAG TTG ATA ATT TTA CCT -3'	5'-CAC TGG ATT GAT CGA GAT GTT CTA TAG TGA GTC GTA TTA -3'
GRIA1 scrambled siRNA	5'-TAC GTG GTT AAT CCT GAT GTT CTA TAG TGA GTC ATA ATG -3'	5'-TTC TAC CAT TTC ACC GGT ATT CTA AAG CAA GAC GTA GTC -3'
A2M siRNA	5'-CTC AAA CAT AAT AGC GAT AAT CCA AGG GAA GAC ATC TTA -3'	5'-AGA GGT TCA ACA GCA AGT CCG CTG CCG ATA ATA ACC TAA -3'

A2M scrambled siRNA	5'-TAT TCT TCC CTA TCA AAT CTG TGG ATG ATA ATT AGA GCT -3'	5'-GAA AAT CTA ATA CCA ATC TTG AAG AAG CAA ACC CGT ATA -3'
GNA14 siRNA	5'-AAC TAG CCT GAC AGC GAT ATT CTA AAG CAA GAC ATA GTA -3'	5'-TAC AAG CCA GTA ATA ATC CAG AAG TAG CAT TCC GGT GTT -3'
GNA14 scrambled siRNA	5'-GAT GGT TCA ACA CCA ATT CCG CAG CCG TTA ATT ATT TAA -3'	5'-TAT GGT TTC GCA TAA ACT GTG TAG TTG CTA ATT TTA CCT -3'
NPY siRNA	5'-CGA CGT TTA AAG CCA ATG ATC AAG TAG GGA ACA TGT CTA -3'	5'-AAC GGG CTT GAT TGT TAA GCT CTA TAG TGA GTC GTA TTA -3'
NPY scrambled siRNA	5'-AAT AAA TCC CCA TCA ACT CTG AAG TTG ATA ATT TTA GCT -3'	5'-AAT GTT TCA ACA CCA ATT CCG CTG CCG TTA ATT CTT TAA -3'
T7 promoter	5'-TAA TAC GAC TCA CTA TAG -3'	

Table 2.2: *List siRNAs*

2.2.8 Cell viability assay:

The viability of granule neurons in culture was estimated by the redox activity of mitochondria in reducing MTT (3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyl tetrazolium bromide) (Sigma) in viable cells. Cells were cultured in poly-lysine coated 24 well plates (1×105 cells per well). After the cultures had reached indicated age, 500µL of 0.5 mg/mL of MTT in fresh medium was added to each well, and the cells were incubated at 37 °C for 4hr. The plates were then centrifuged at 1500 rpm for 20 min at room temperature, and the medium was carefully removed. Dimethyl sulfoxide (DMSO) (500µL) was then added to each well to dissolve the formazan crystals. The DMSO-dissolved formazan crystals were read immediately at 540 nm with DMSO as blank on a spectrophotometer (JASCO, Japan).

2.3 Results

The network obtained from DE model was shown in figure 2.5. To test the resulting model we have predicted the expression value of genes at particular time points given the expression value of previous time points. The predicted results were compared with the actual value, and the r-square values were calculated (figure 2.6). Most of the r-square values are close to 1, which confirms the accuracy of the model.

Program 3.

- **To predict the expression value from previous known time points (p3.R)**

```
maindataset <- read.table(file = "prdiction_set-test.txt")
wtMat <- read.table(file = "wtMat_data_15.txt", header=TRUE)
binMat <- read.table(file = "resMat_data_15.txt", header=TRUE)
regMat = wtMat * binMat

maindataset <- as.matrix(maindataset)
wtMat <- as.matrix(wtMat)
regMat <- as.matrix(regMat)
difMat <- matrix(0:0, nrow = 15, ncol = 23)
predMat <- matrix(0:0, nrow = 15, ncol = 23)
for(x in 1:23){
     Y = 10* wtMat %*%  maindataset[,x]
     X = Y + maindataset[,x]
     predMat[,x] <- X
}

filename <- sprintf("predicted_set-test_15.txt")
write.table(predMat, filename, quote=FALSE, sep="\t",
row.names=FALSE)
```

The major challenge in constructing the gene networks from microarray data is the dimensionality problem usually referred as dimensionality curse (Wang *et al.*, 2006, Wessels *et al.*, 2001), which means the number of genes is much more than the number of experiments. In the differential equation model if the number of genes becomes greater than the number of experiments then it is difficult to solve the Λ matrix (de Hoon *et al.*,

2003). To tackle this problem we can reduce the number of genes by replacing with the cluster of genes (de Hoon *et al.*, 2003, Eisen *et al.*, 1998).

Genes showing the same pattern of expression can be clustered into one cluster. But in the present data, there are only four time points, which give us three fold change values; therefore we were restricted to only 3 clusters. As the number is very less, the network construction using this type of data provides limited interactive counterparts leading loss of information.

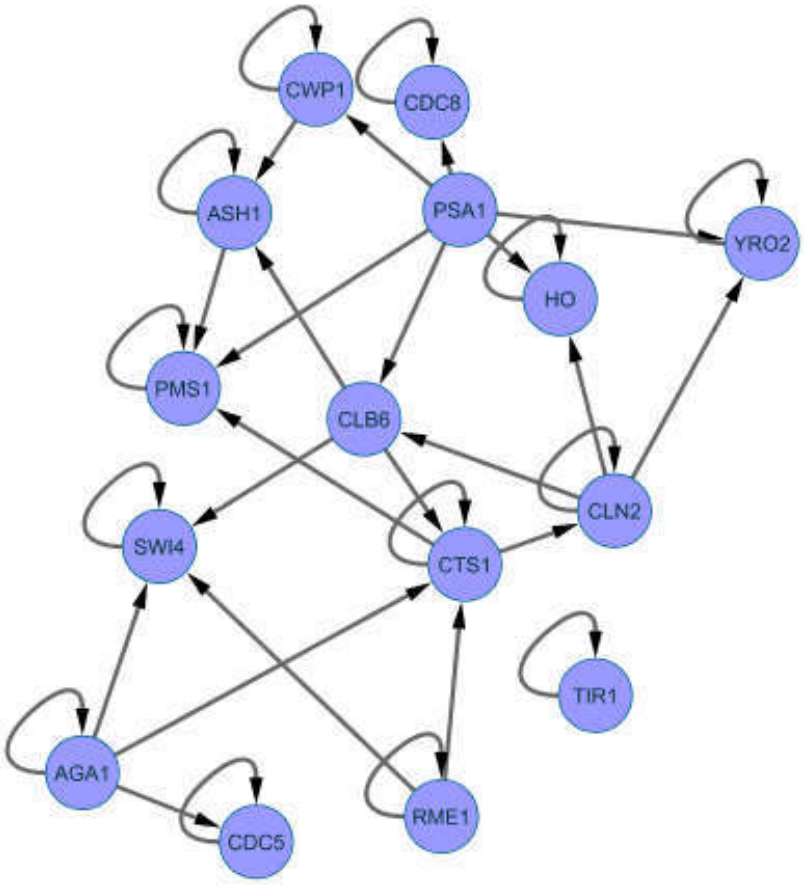

Figure 2.5: *Gene interaction network constructed using differential equation model. Edges with the arrow between the two genes represent the directional effect of on gene on other.*

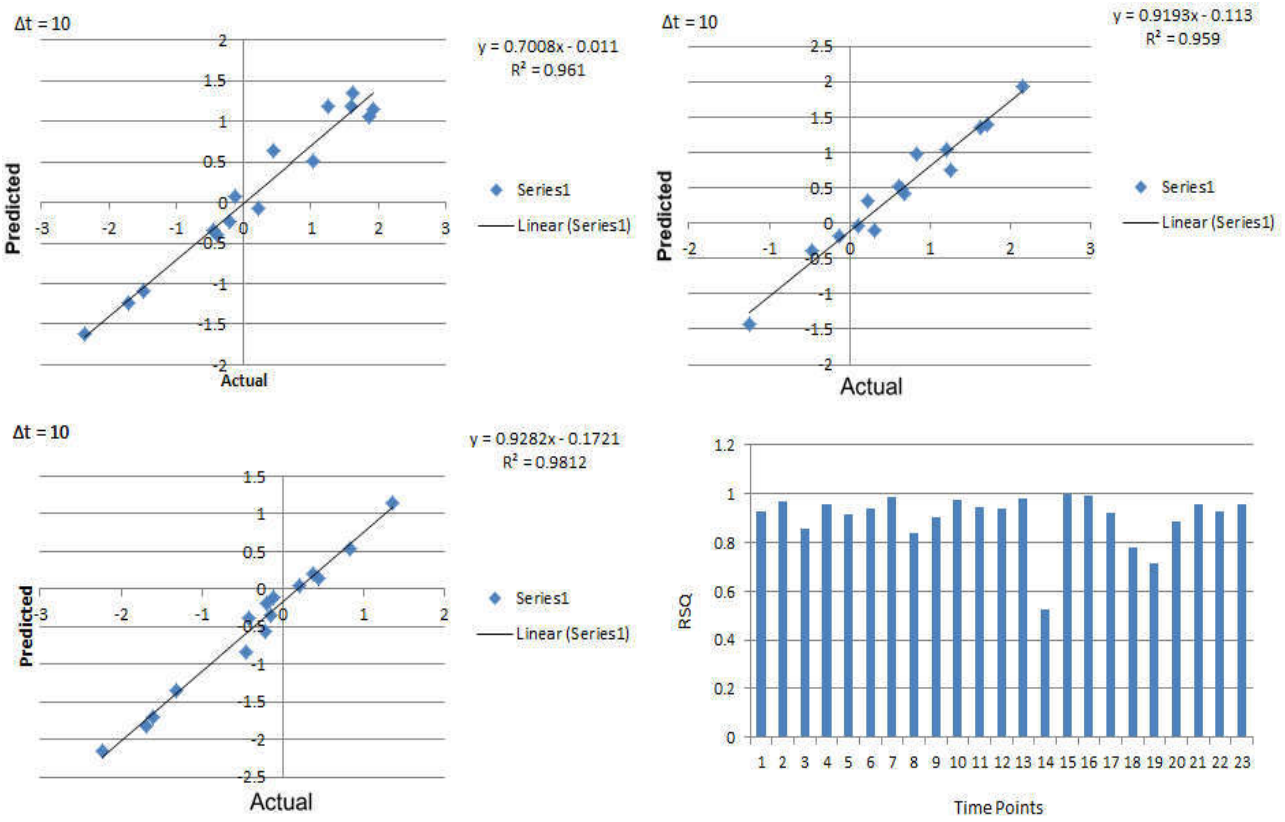

Figure 2.6: *R-square values between the predicted and actual values at different time points.*

The second method we have used to reconstruct the gene network from microarray expression data is supervised learning method. The network obtained from this method is shown in figure 2.7. The decision rules obtained from this method are listed in appendix 1. Modular analysis of the network using Moduland divided the network into 15 modules. The modular correlation table is given in Table 2.3.

Each module (M1 to M15) is represented by a gene which is the center node of the network *MMP2, PDGFRB, GJA1, KITL, GABRA6, ICAM1, PLAU, ACSL6, TAP2, MYH14, BMP4, NMNAT3, AGA, DHRS3, and LBP.* The Spearman's rank correlation values were calculated between the module pairs. Modules MMP2 and PDGFRB show the highest correlation (0.73). These two modules have a number of pathways which have significant roles in neuronal aging for example mTOR signaling pathway (Droge & Schipper, 2007, Laplante & Sabatini, 2012), p53 signaling pathway (Culmsee & Mattson, 2005), Axon guidance and Calcium signaling pathway (Nikoletopoulou & Tavernarakis, 2012). MMP2 and PDGFRB have also been

shown to play a role in the neuron degeneration and development (Gu *et al.*, 2002, Hoch & Soriano, 2003).

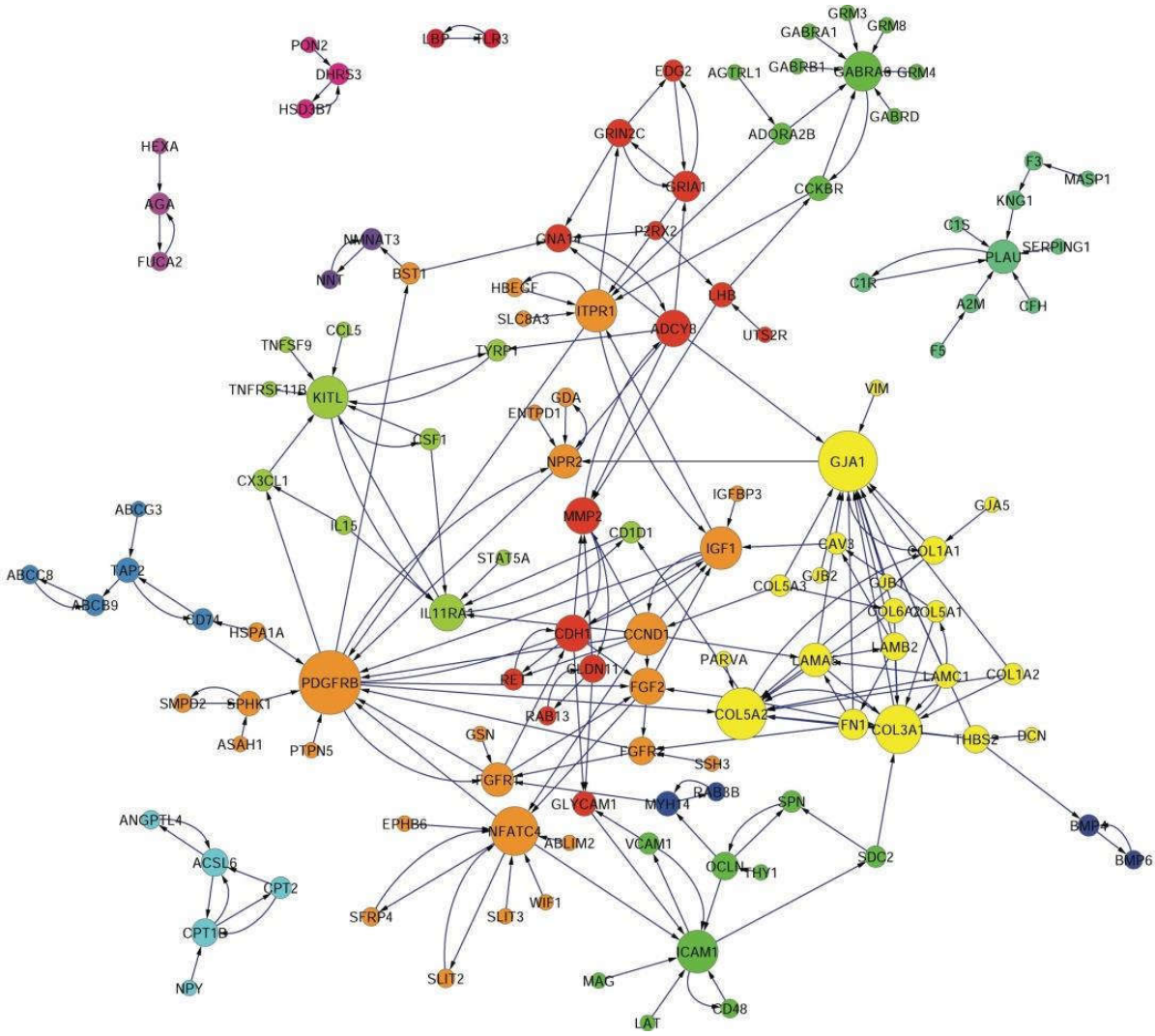

Figure 2.7: *The network structure of genes. The network is divided into 15 modules using moduland plugin of Cytoscape. Each module is shown in different colour. The size of the node represents the degree of the node. Higher the degree, bigger is the size of the node.*

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M14	M15
M1	1														
M2	0.73	1													
M3	0.15	0.19	1												
M4	0.09	0.05	0.09	1											
M5	-0.02	-0.03	-0.27	-0.14	1										
M6	0.23	0.19	-0.2	-0.16	-0.15	1									
M7	-0.29	-0.31	-0.21	-0.11	-0.11	-0.12	1								
M8	-0.2	-0.21	-0.15	-0.08	-0.07	-0.08	-0.06	1							
M9	-0.13	-0.1	-0.06	-0.08	-0.08	-0.09	-0.07	-0.05	1						
M10	0.13	0.07	-0.16	-0.08	-0.08	0.42	-0.07	-0.05	-0.05	1					
M11	-0.18	-0.19	0.17	-0.07	-0.07	-0.08	-0.05	-0.04	-0.04	-0.04	1				
M12	0.07	0.03	0.04	-0.06	-0.06	-0.06	-0.05	-0.03	-0.03	-0.03	-0.03	1			
M13	-0.15	-0.16	-0.11	-0.06	-0.06	-0.06	-0.05	-0.03	-0.03	-0.03	-0.03	-0.02	1		
M14	-0.15	-0.16	-0.11	-0.06	-0.06	-0.06	-0.05	-0.03	-0.03	-0.03	-0.03	-0.02	-0.02	1	
M15	-0.13	-0.13	-0.09	-0.05	-0.05	-0.05	-0.04	-0.03	-0.03	-0.03	-0.02	-0.02	-0.02	-0.02	1

Table 2.3: *Modular correlation table*

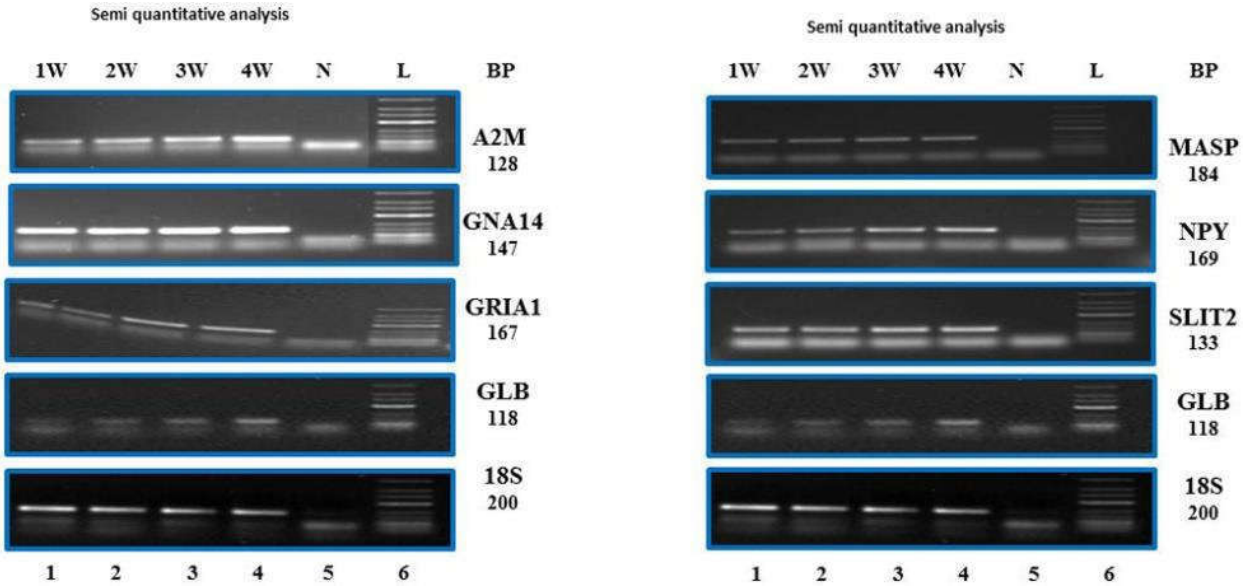

Figure 2.8: *Semi-quantitative PCR analysis. Panels A and B show semi-quantitative gene expression of A2M, GNA14, GRIA1, MASP1, NPY, SLIT2, and GLB1. All the genes showed significantly increased expression in 2W, 3W, and 4W over 1W. Lanes 1-4 correspond to 1W to 4W transcript expression; lane 5 is no template control (NTC), and lane 6 is 50 bp ladder. Semi-quantitative analysis was done to ensure single product amplification.*

With further correlational analysis in literatures, we have selected following six genes from the network for validation purpose

1. **A2m -** alpha-2-macroglobulin
2. **GNA14 -** guanine nucleotide binding protein, alpha 14
3. **Gria1 -** glutamate receptor, ionotropic, AMPA 1
4. **Masp1 -** mannan-binding lectin serine peptidase 1
5. **Npy -** neuropeptide Y
6. **Slit2 -** slit homolog 2

The expression patterns of these genes were confirmed using semi-quantitative and real-time PCR (Fig 2.8 and Fig 2.9).

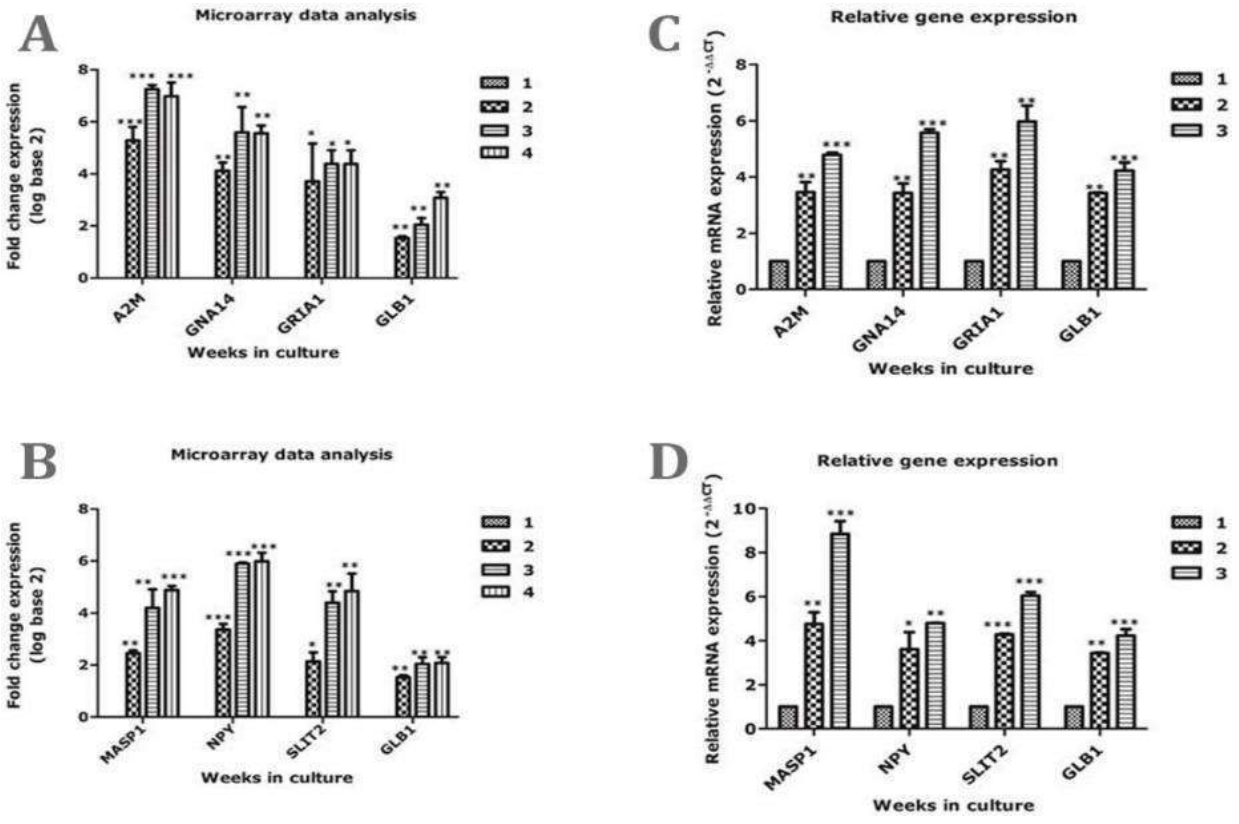

Figure 2.9: *Panels A and B show fold change expression by Microarray analysis. Panels C and D depict validation by q- RT-PCR of genes – A2M, GNA14, GRIA1, MASP1, NPY, SLIT2, and GLB1. q-RT-PCR analysis reveals a significant increase in relative expression of all genes through the weeks as per statistical analysis, One Way ANOVA (Tukey post hoc test) (*P< 0.05,*

***P< 0.01, ***P< 0.001). 1W healthy CGNs were taken as control. Values are represented as mean ± SD and n = 6. Comparative CT method was employed to evaluate the gene transcript pattern in ageing cultures.*

We have also observed the viability of neurons using MTT assay upon silencing these six genes. The result shows that out of six genes, only two genes Npy and Slit2 have shown a significant increase in neuronal viability (Fig 2.10). Thus suggesting that Npy and Slit2 may play a role in neuronal senescence and we propose Npy and Slit2 as potential markers for senescence in cultures CGN's.

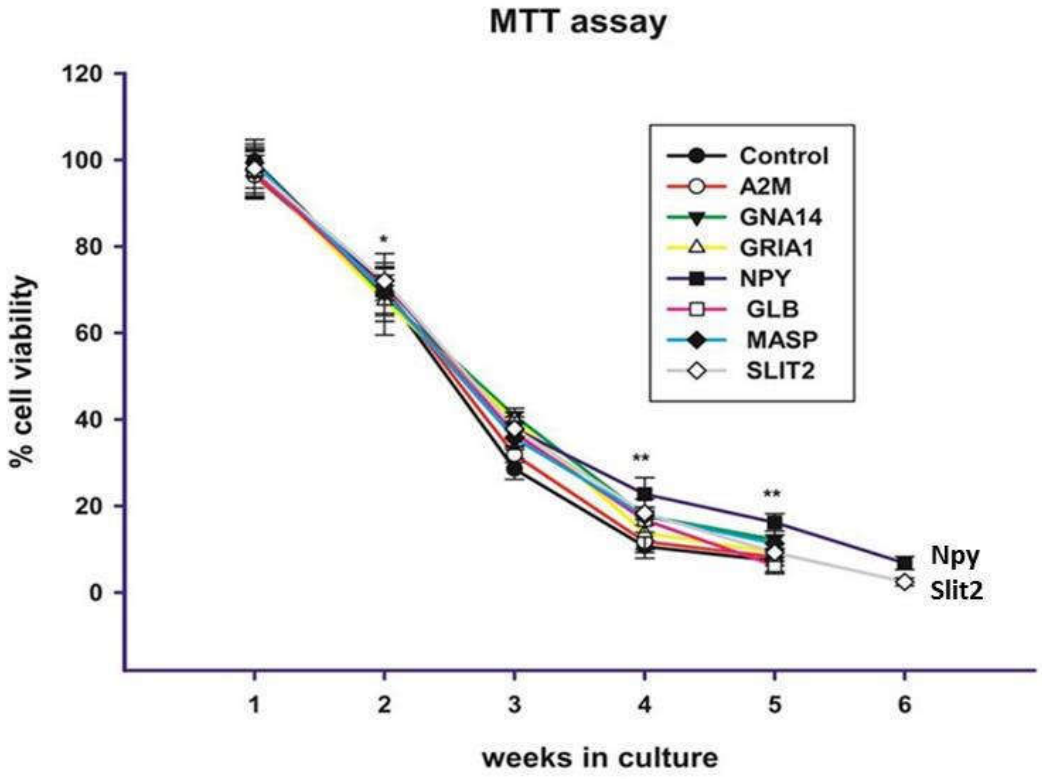

Figure 2.10: *Figure shows cytotoxicity detected by reduction of MTT in ageing CGNs. Results are expressed as mean ± SD (3 replicates in two independent experiments). Cell viability in the first week was considered as 100%. Values are presented as a percentage of activity in control cells. Cell viability is compared at every week based on MTT values presented. The results showed a marked increase in cell viability in NPY, SLIT2 downregulated cells compared to control when knock down was done at 3rd week in culture.*

2.4 Conclusion:

We studied two different methods to reconstruct gene networks from microarray data; 1. Differential equation model and 2. Supervised learning method. DE models are very sensitive to the dimension of the data and are very difficult to use for high dimensional data. Although we tested the DE model on yeast cell cycle data and evaluated the model accuracy by predicting the expression values at different time points. The model performed well on data with a large number of time points. Since we had only four time points, we preferred to use supervised learning method for gene network reconstruction. We used biological categorization of genes to overcome dimensionality issue in combination with machine learning methods. The modular analysis of the network and further literature survey allowed us to shortlist six genes from the network. Gene silencing studies of these six genes further validated their role in neuronal senescence. As a result, only two genes, Npy and Slit2, showed significant increase in the viability of the neurons thus we proposed Npy and Slit2 as potential markers for senescence in cultures CGN's

Chapter 3

Construction of hypothetical model showing role of top2b in neuron maturation and axon guidance

3.1 Introduction

Replication and transcription are the two important biological processes performed by DNA. DNA replication is a process by which DNA produces two identical copies from a one original DNA. DNA replication is the basis for biological inheritance. On the other hand DNA transcription is a biological process in which DNA is copied into RNA, which is followed by the translation process in which RNA is used for the synthesis of protein. Both of these events require the unwinding of DNA, as DNA inside the nucleus is present in coiled form of the double stranded helix. The strands of DNA are used as template to synthesize new strand of DNA (in replication) or RNA (in transcription), thus the unwinding of the two strands become a crucial step. During the unwinding of DNA few topological changes happens in the DNA structure for example during DNA replication DNA becomes overwound ahead of the replication fork. If such overwinding in the DNA is not resolved then the process would halt or would results in strand breakage for release of tortional stress. Topoisomerases are the enzymes that govern the overwinding or underwinding of DNA and regulate the topologic states of DNA during replication, transcription (Wang, 1996, Osheroff *et al.*, 1991). Topoisomerases performs these reactions by creating transient breaks in one or both strands of DNA (Champoux, 2001). *Depending upon the number of strands cut in one reaction, the topoisomerases can be classified into two classes; Type 1 and Type 2 topoisomerases (Wang, 1991). A* type I topoisomerase cuts one strand of a DNA double helix which allows the relaxation of the strands after which the cut strand is re-annealed. Type 2 DNA topoisomerases cut both strands DNA double helix. These enzymes catalyzes the transient breaking and rejoining of two strands of duplex DNA, which allows the strands to pass through one another, thus altering the topology of DNA. Both type 1 and type 2 topoisomerases are further subdivided into two categories, alpha and beta. All four types of DNA topoisomerases are explained in table 3.1.

Topoisomerase	IA	IB	IIA	IIB
Metal Dependence	Yes	No	Yes	Yes
ATP Dependence	No	No	Yes	Yes
Single- or Double-Stranded cleavage?	SS	SS	DS	DS
Cleavage Polarity	5'	3'	5'	5'
Change in L	±1	±N	±2	±2

Table 3.1:

Topoisomerase 2 alpha (Top2a) expressed abundantly in highly proliferating cells, while topoisomerase 2 beta (Top2b) is known to be expressed in both proliferating and fully differentiated cells and tissues (Capranico *et al.*, 1992). The beta isoform of type 2 topoisomerase is predominant in the brain (Kondapi *et al.*, 2004, Tsutsui *et al.*, 2001). Previous studies in our lab have demonstrated that top2b plays crucial role in neuronal growth and survival (Kondapi *et al.*, 2004, Gupta *et al.*, 2012). The top2b expression is inversely synchronizes with senescence and thus forming an important biomarker for aging (Kondapi *et al.*, 2004). In order to functionally classify and identifying the crucial top2b associated pathways in various biological processes, we have first developed a database of top2b-associated genes. By making use of available information present in the database, we have developed a hypothetical map showing the role of top2b in neuron maturation and axon guidance.

3.2 Methodology

3.2.1 Data collection:

An intensive literature survey has been done and all the genes associated with top2b were listed. The information about all these genes, such as gene id, description, chromosome etc., was collected from NCBI GenBank database. To make the database more

comprehensive we have also collected the data from NCBI Human-HIV interaction database. The HIV proteins interacting with these genes listed in previous step were also submitted to the database. All the top2b interacting genes were hierarchically classified using KEGG database. The literature information has also been collected from various sources.

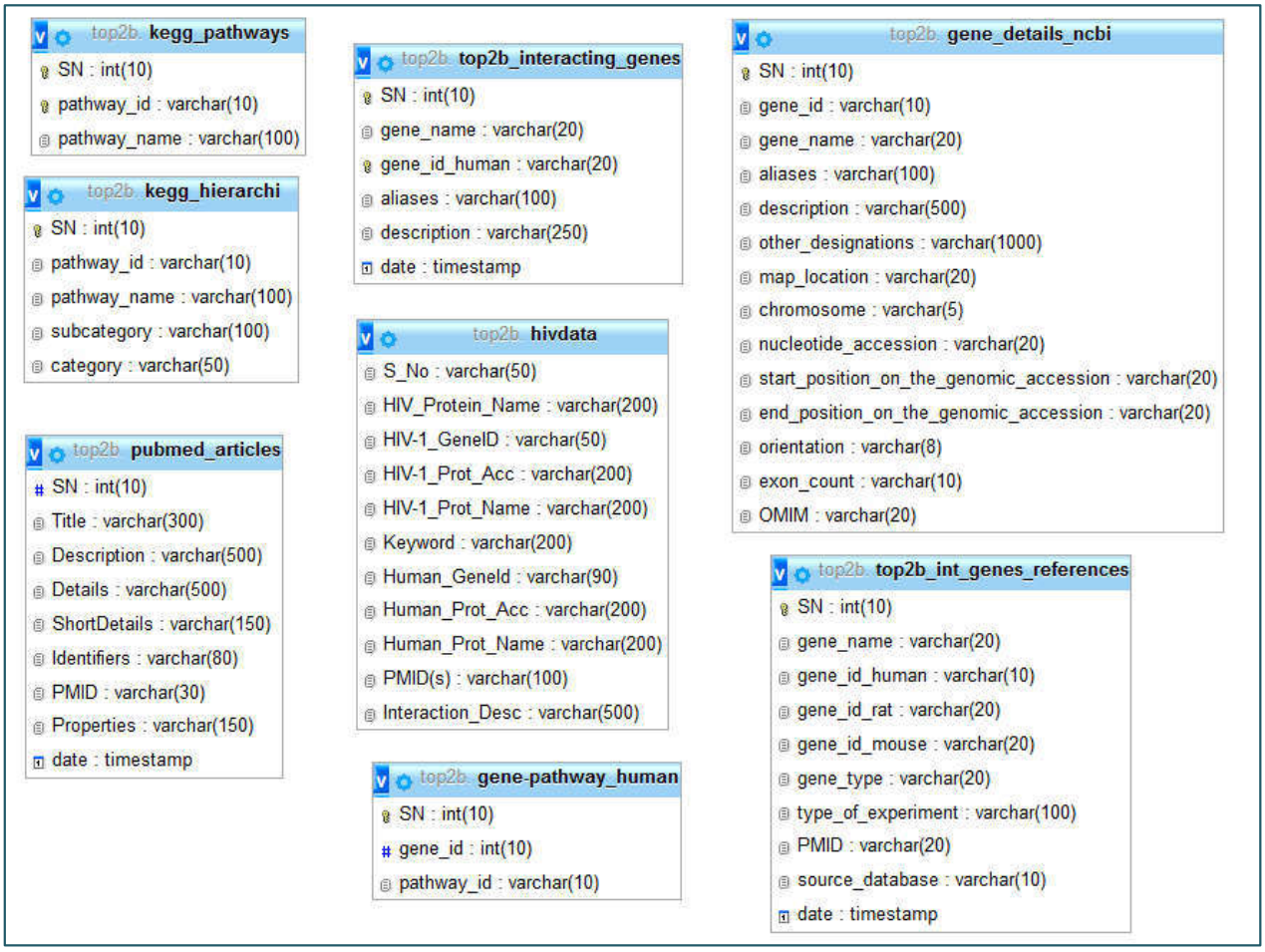

Figure 3.1: *Database tables*

3.2.2 Database development: The database was developed in MySql, and the web interface was developed using HTML, PHP, and JavaScript. Apache XAMPP open source server v3.2.1 was used to develop website and database. All the tables in the database are shown in figure 3.1. The database is accessible with the URL http://bif.uohyd.ac.in/top2b. The web interface provides a common search option for all types of search data. The search results were displayed in three separate tabs each belongs to genes, pathways, and HIV

interaction data respectively. The database also provides interactive pathways map which can be browsed to further subcategories. The link between two categories denotes that they have common top2b interacting genes.

3.2.3 Analysis of top2b associated genes involved in neuron maturation and axon guidance: Genes involved in axon guidance were listed, and a hypothetical model was developed which shows the relationship between different pathways related to axon guidance and neuron maturation. Top2b associated genes were mapped in the model displaying the effect of top2b in axon guidance and related pathways.

3.2.4 Validation: From the above model a hypothesis was developed which explains how inhibition of Npy and Slit2 results in the increased expression of Top2b. The validation was done by siRNA gene silencing. After siRNA inhibition of Npy and Slit2 a slight increase in top2b expression level was observed.

3.3 Results

A database of Top2b associated genes was created using available literature information about Top2b. The database can be accessed at http://bif.uohyd.ac.in/top2b (Figure 3.2). It allows users to browse interactive pathway interaction map of top2b associated pathways and is helpful in developing the Top2b centric models in various functions/disease pathways such as Cancer. Users can also view the interactions among HIV and Human proteins which are associated with Top2b. The database also allows users to upload the expression value of genes, which they can visualize on KEGG pathways. This database also features with new submission entry by any registered user. The basic features of the database are explained below.

3.3.1 Main Search Engine:

The home page of the database displays a global search (Fig 3.2) option which accepts any keyword and displays results under three different categories (Fig 3.3a-c). The search engine is designed in such a way that it can retrieve relevant data from all three different fields in context to top2b. The “Genes” tab in the search result page lists all the genes

matching to the search keyword. These genes are either directly or indirectly linked to the Top2b based on available information present in the database. The information about the gene can be viewed by clicking on respective genes.

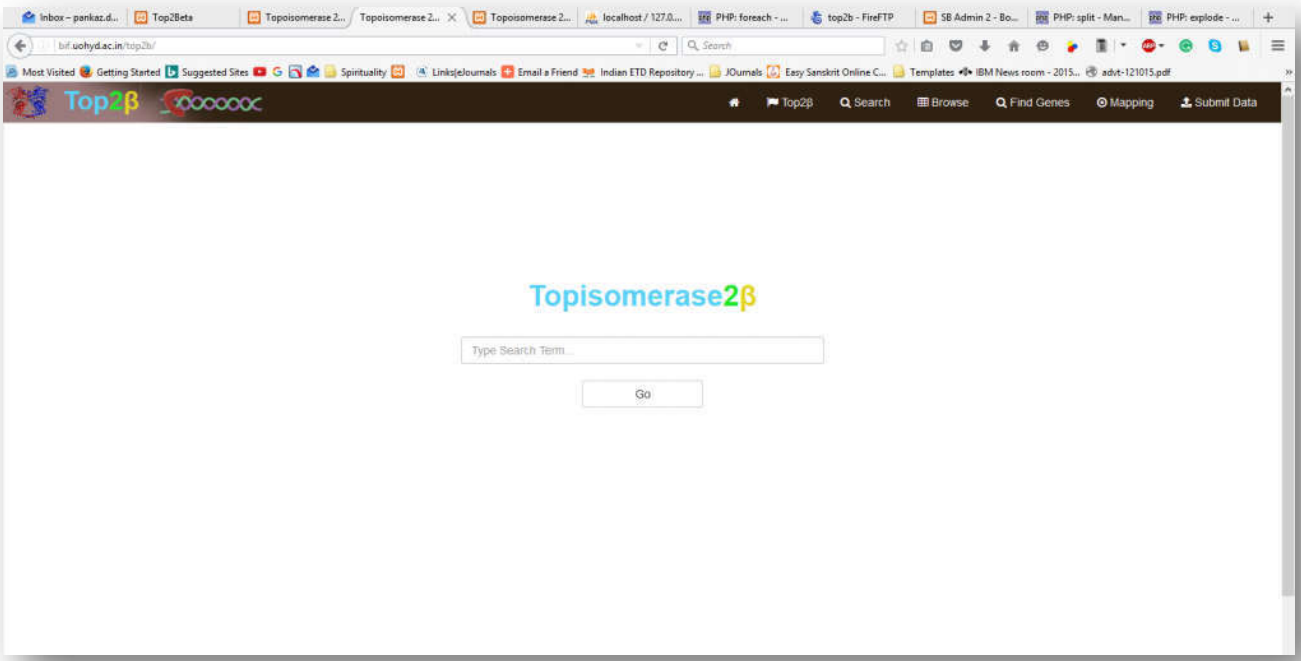

Figure 3.2: *Home page of the database*

3.3.2 Custom Search and Browsing Data

The search option provided on the top menu bar of the website allows users to search pathways and genes independently (Fig 3.4). An additional link "Browse" on the top menu bar list out all the genes in single page showing links to other information about genes such as Gene Descriptions, other common names, References (link to PubMed article which reports the relationship of corresponding with Top2b), and links to external database such as GeneCards and NCBI.

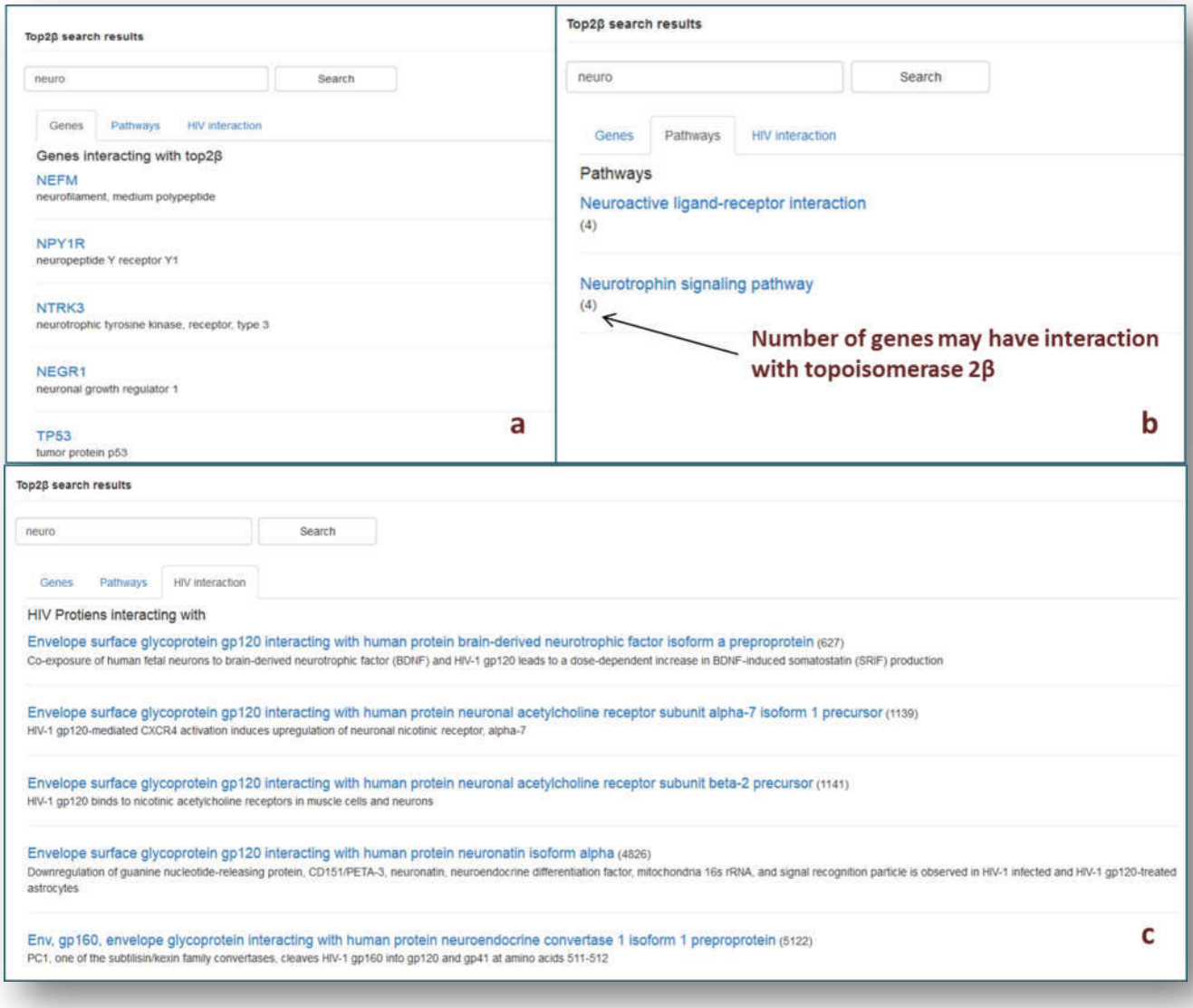

Figure 3.3: *The search result page for the main search engine. Three classes Genes, Pathways and HIV interaction data from the search results are displayed respectively in a, b, and c.*

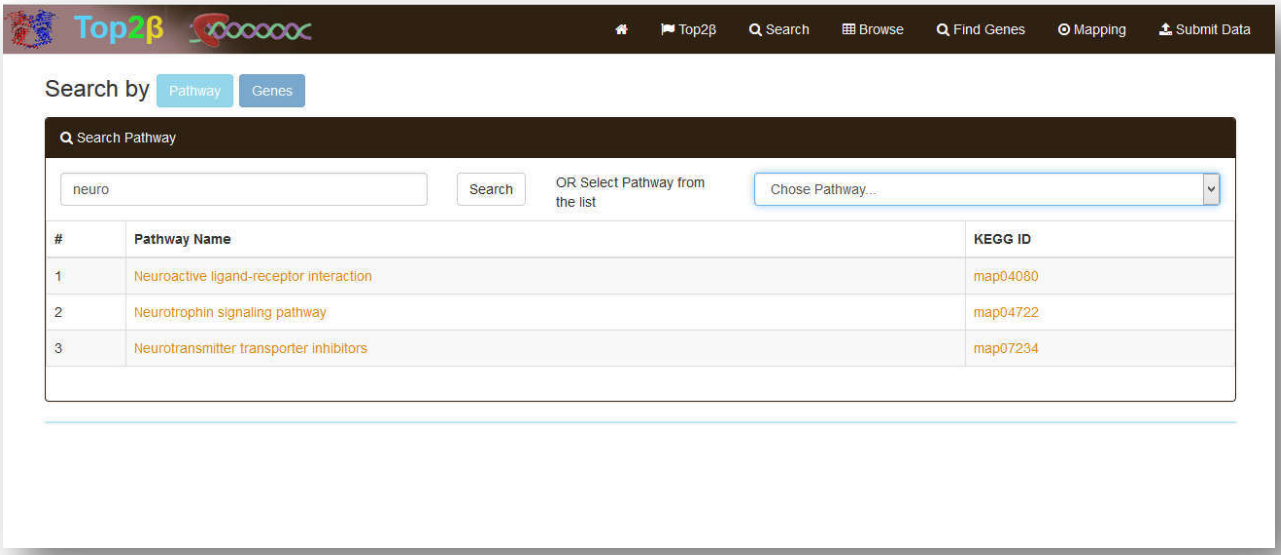

Figure 3.4: *Custom search for pathways and genes.*

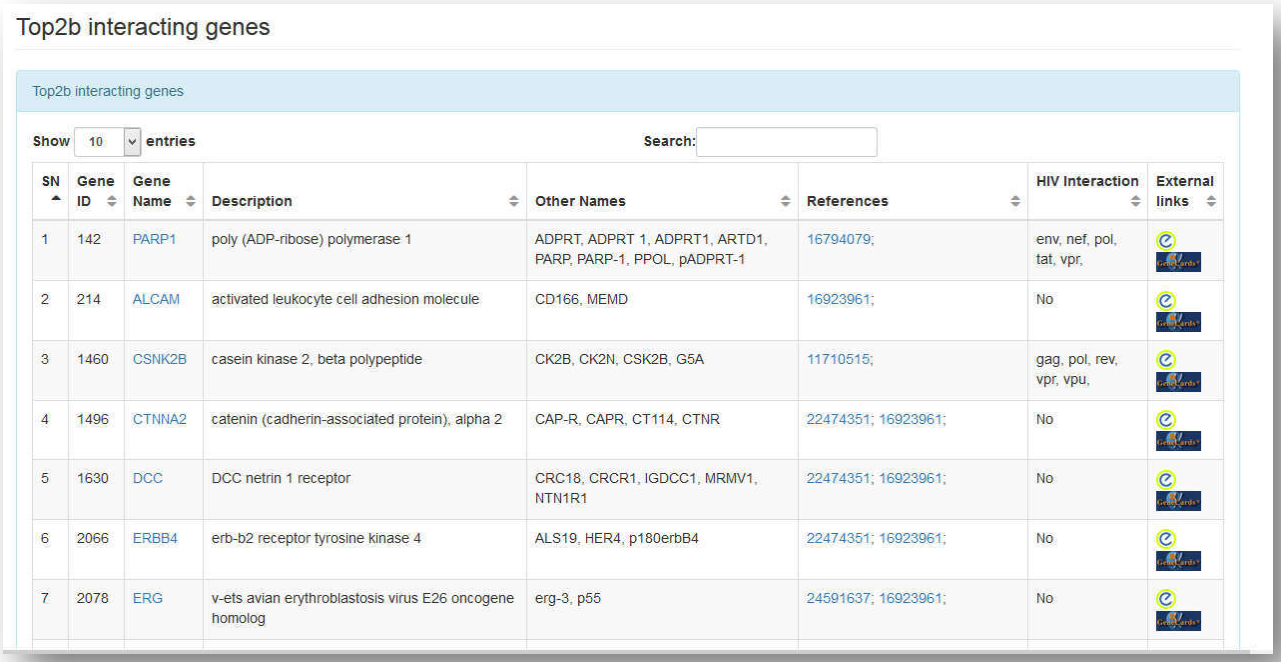

Figure 3.5: *Listing all the genes which are linked to Top2b identified in literature.*

3.3.3 Mapping expression data on KEGG pathways

We have integrated our tool to KEGG pathway interface. This feature allows the user to upload the gene along with their expression values which can be directly viewed on KEGG pathway map as a heat map (Fig 3.6 & 3.7).

3.3.4 New Data Submission

To make the database more live and updated we have also included a submission form in the database where the user can make new entries (Fig 3.8). All the new entries done by any user are first stored in the temporary database. These entries are then verified and curated by us and submitted finally to the main database making them accessible to other users.

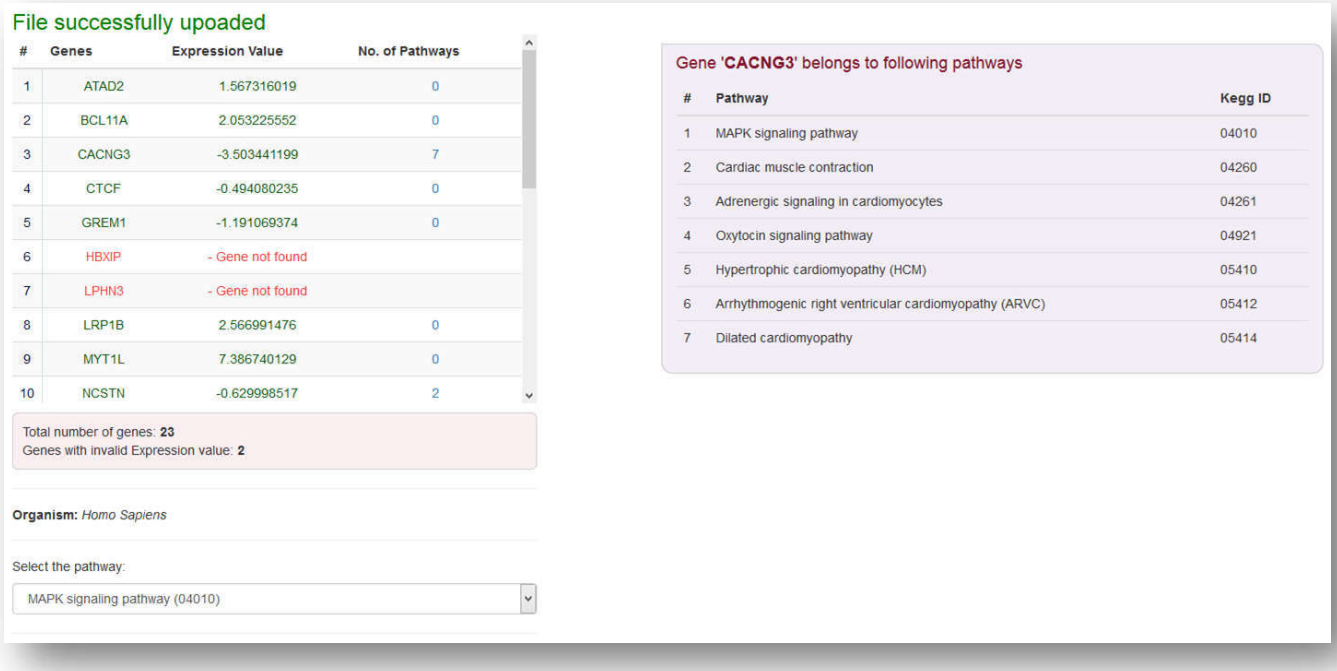

Figure 3.6: *Page is showing the genes and their expression value uploaded by the user. On mouse over on a particular gene name, the list of corresponding pathways is displayed on the right side of the page.*

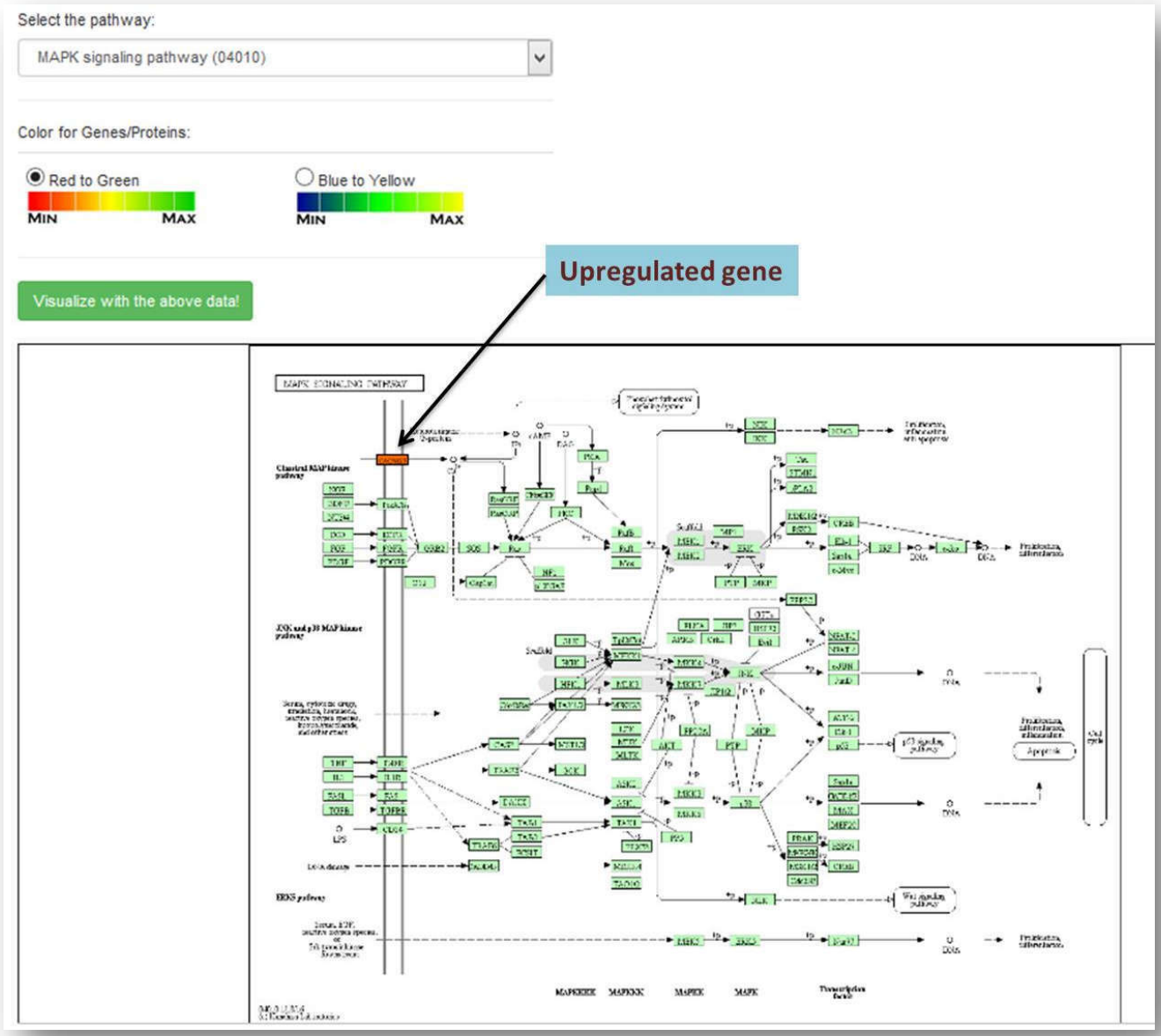

Figure 3.7: *Mapping of expression data of uploaded genes on the KEGG pathways. Two color gradient options are available, Red to Green and Blue to Yellow.*

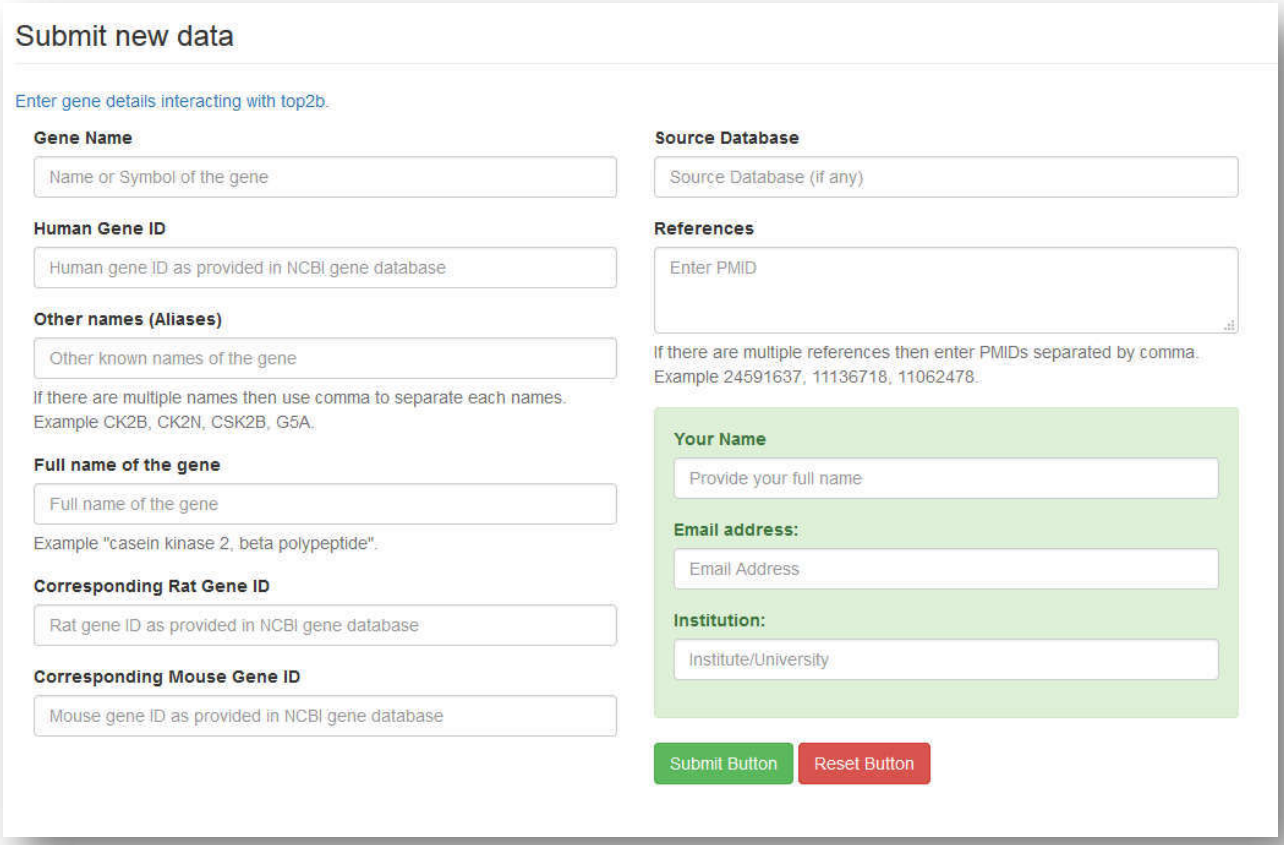

Figure 3.8: *Data Submission form allows any user to submit a new entry.*

Using the available information in the database, we have developed a hypothetical model explaining the role of Top2b in axon guidance and neuron maturation (Figure 3.9). The model explains the interference of Top2b in pathways related to axon guidance. Based on this model, a hypothesis was established, showing how Npy and Slit2 inhibition may increase the expression of Top2b (as explained in chapter 2). After the gene silencing study of Npy and Slit2, the expression of Top2b was found significantly increased in Npy and Slit2 silenced CGNs (Figure 10). This observation validates hypothetical model involving Slit2 and Npy in the regulation of cell survival and Topo2b.

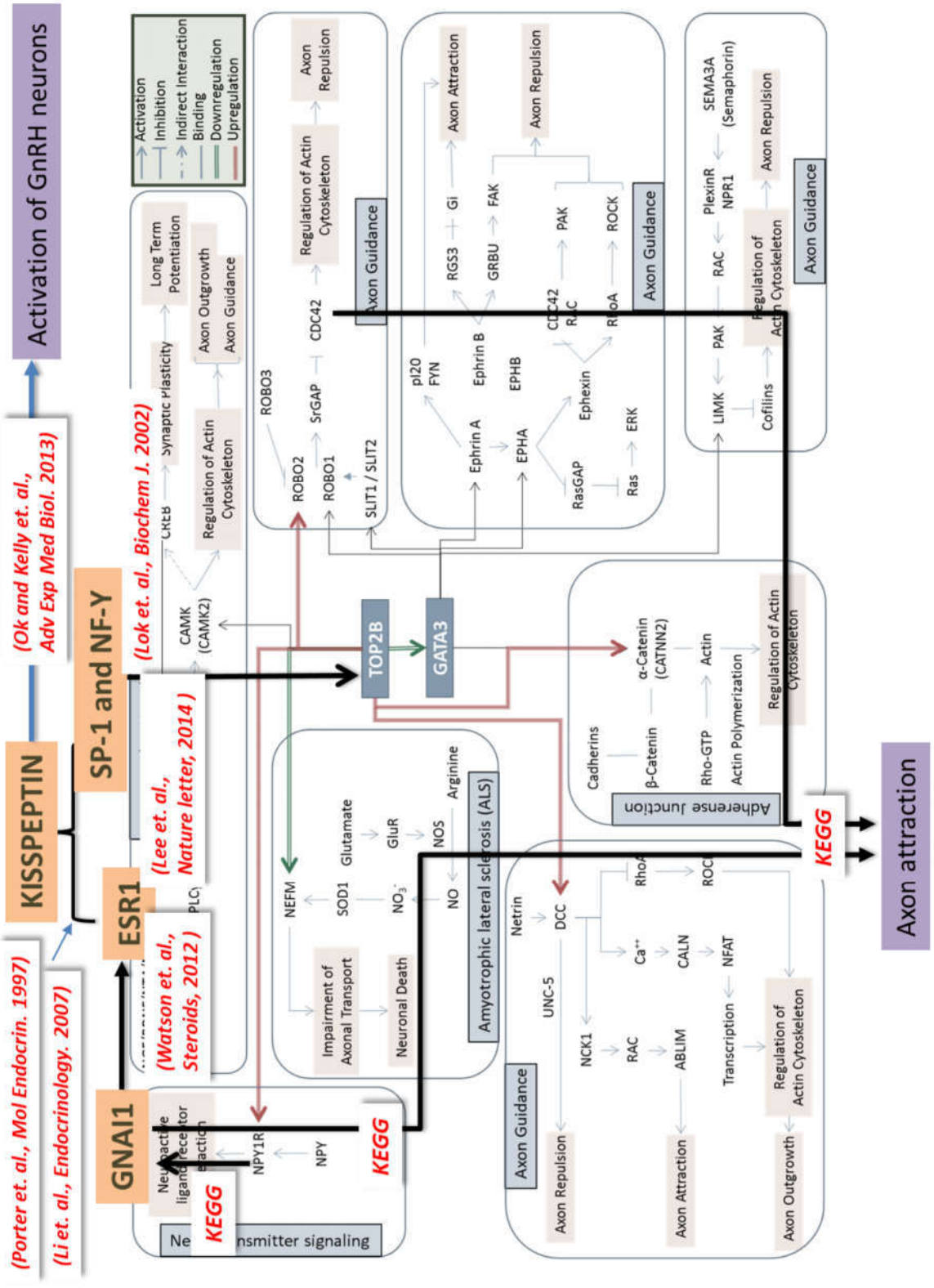

Figure 3.9: *A hypothetical model, showing the role of top2b in axon guidance pathway and the link between Npy, Slit2, and Top2b.*

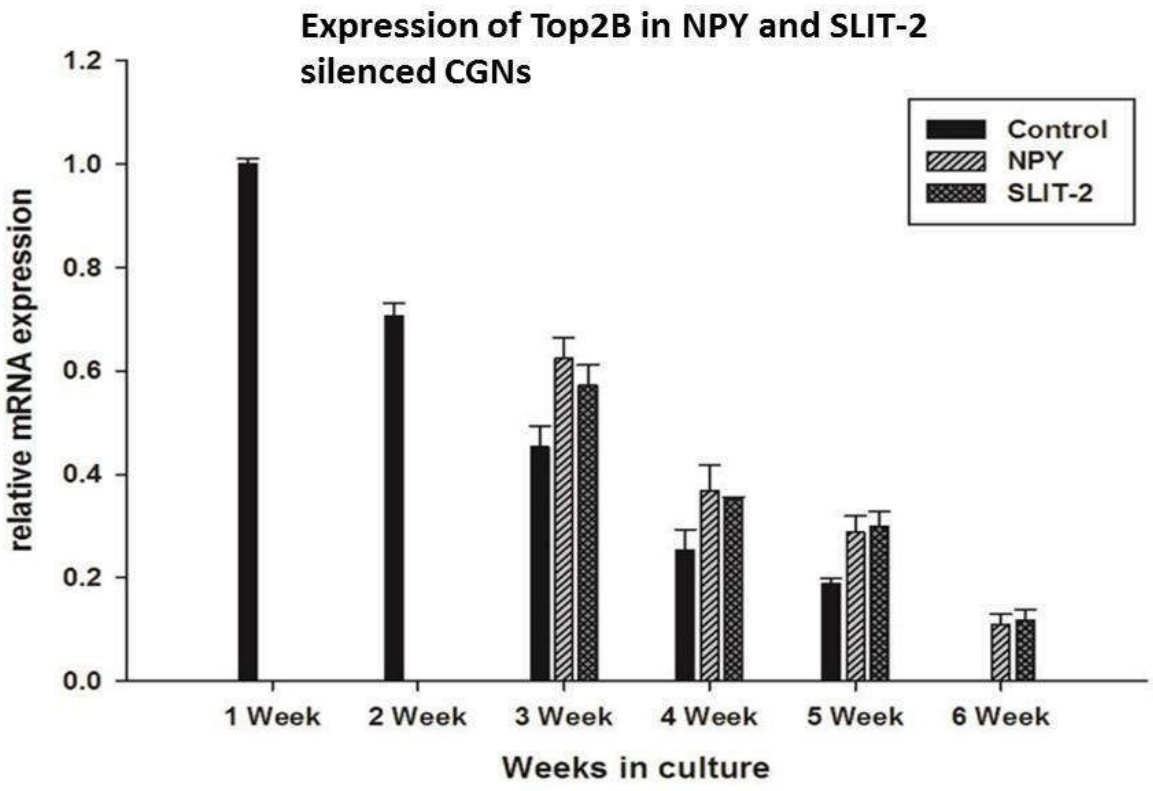

Figure 3.10: *Expression of Top2b in Npy and Slit2 silenced cells.*

3.4 Conclusion

We have listed all the genes which were showing interaction with Top2b from literature survey and created a database. Based on the available information a hypothetical model explaining the role of top2 beta in axon guidance and neuron maturation was developed. By analyzing the model we have established a hypothesis, showing how Npy and Slit2 inhibition may increase the expression of Top2b and Top2b expression was found significantly increased in Npy and Slit2 silenced CGNs. This increase in expression of Top2b was observed in the neurons when Npy and Slit2 genes were silenced. This database can also be used to develop such top2b centric models for any other particular disease/function.

Chapter 4

Identification of crucial pathways involved in anti-HIV activity of curcumin

4.1 Introduction

Curcumin is a major polyphenolic component of natural herb *Curcuma longa* (turmeric). It is known that curcumin exhibits various pharmacological properties including anti-inflammatory, antioxidant, anti-carcinogenic and anti-microbicidal effects (Aggarwal *et al.*, 2003, Jurenka, 2009, Kuo *et al.*, 1996, Moghadamtousi *et al.*, 2014, Si *et al.*, 2007). Curcumin modulates several molecular pathways, including infection, immune response and apoptosis (Bhattacharyya *et al.*, 2010, Ciftci *et al.*, 2010, De *et al.*, 2009, Martins *et al.*, 2009, Moghadamtousi *et al.*, 2014, Piwocka *et al.*, 1999, Rechtman *et al.*, 2010, Si *et al.*, 2007) by regulating the gene expression of target genes involved in these pathways, either by direct interaction or by modulating the signaling events involved in the transcriptional machinery (Gupta *et al.*, 2011, Misra *et al.*, 2011, Shehzad & Lee, 2013). Curcumin inhibits many viral infections, including HIV-1 through its specific interactions with viral proteins (Kutluay *et al.*, 2008, Mazumder *et al.*, 1995, Sui *et al.*, 1993), inhibition of viral gene expression (Taher *et al.*, 2003), modulating the recruitment of viral proteins to the gene promoters and also by inhibition of intracellular signaling pathways (Barthelemy *et al.*, 1998, Singh & Aggarwal, 1995, Tomita, Matsuda, *et al.*, 2006, Zhang *et al.*, 2011). Activation or repression of host signal transduction mechanisms by alterations in the expression of target genes involved in the pathway is a prerequisite for the establishment of viral infection and its persistence in the host. The sustainability of HIV-1 infection depends on the complex interplay between viral proteins and host cytokine signaling pathways. Signal transducer and activator of transcription (STAT) are the primary effectors of cytokine signaling in many haematopoetic cell lineages. HIV-1 is known to activate multiple STATs during acute and chronic infection, which further cascades in cytokine signaling and regulates host immune system for viral persistence and propagation. HIV-1 infection can activate STAT signaling through direct interactions of STAT with viral proteins, several HIV-1 proteins, including Tat and Nef, can directly affect JAK/STAT pathways or indirectly by paracrine stimulation of STATs through virus-host cell interactions in bystander cells (Bovolenta *et al.*, 1999, Federico *et al.*, 2001, Herbein *et al.*, 2010, Kohler *et al.*, 2003, Tomita, Kawakami, *et al.*, 2006). The immune suppression during the HIV-1 infection is majorly could be due to

the regulation of cytokine signaling pathways like STAT, in infected T cells and macrophages causing impairment of immune function.

Activation of JAK/STAT signaling during HIV-1 infection is acute phenomenon, which occurs immediately during early interaction with gp120 and $CD4^+$ cells, triggering multiple cytokine signaling cascades, which in turn leads to constitutive activation of the signaling pathway and modulation of host immune microenvironment (Bovolenta *et al.*, 1999, Liu *et al.*, 2012). The complex interplay between the proinflammatory and immunomodulatory cytokines during the acute phase of infection is crucial for the establishment of the viral infection and progression of the disease (Bahbouhi *et al.*, 2004).

The regulation of the cytokine signaling network is by Cytokine-inducible SH2-containing protein family of proteins (SOCS or SSI), which function as a negative feedback loop of cytokine signaling. The expression of Suppressors of Cytokine Signaling (SOCS) proteins is modulated by the activation state of JAK/STAT pathway, which in turn cascades cytokine signaling. Enhanced SOCS expression can inhibit the normal function of JAK/STAT-regulated pathways (Miller *et al.*, 2011). Effect of Curcumin on inhibition of JAK/STAT pathway was studied extensively in elucidating its anti-inflammatory properties (Moghadamtousi *et al.*, 2014). Curcumin inhibits phosphorylation of JAK1/2 and also by activation of SHP1 thus disrupting the inflammatory responses in activated brain microglia (Kim *et al.*, 2003). The study of the differential regulation of host gene expression during viral infection and curcumin treatment prompts the molecular pathways involved in the process. Pharmacological use of the soluble form of curcumin is limited due to low bioavailability. Thus curcumin loaded apotranferrin nanoparticles (nano-curcumin) would enhance curcumin bioavailability and anti-HIV-1 activity (Gandapu *et al.*, 2011).

To delineate the molecular pathways targeted by curcumin during HIV-1 acute infection, a genome-wide microarray analysis was performed during HIV-1 acute infection in the presence and absence of nano-curcumin. Differentially regulated genes were identified and categorized into pathways, and a gene interaction map was developed. The present study resulted in validation of JAK/STAT pathway as a potential novel target for nano-curcumin

for blocking HIV infection. This pathway can be further explored to identify the potential drug targets for HIV/AIDS.

4.2 Materials and Methods

The microarray experiment was performed to identify potential targets for the anti-HIV activity of nano-curcumin. This experiment resulted in a huge number of differentially expressed genes. The genes were then shortlisted based on their expression values and were subjected to pathway analysis. A gene interaction map was developed based on multipath genes and the genes sharing the same pathway were connected with an edge. We have established a scoring system for each gene based on two factors: the number of pathways in which a particular gene is involved and the connectivity of that gene in the network. The schematic representation of the complete workflow is shown in figure 4.1.

4.2.1 Cell culture and Microarray:

The cells were collected at 48 hours of post infection in presence or absence of nano-curcumin. CD4 positive T cell line, SupT1 and Subtype C virus, 93IN101 HIV type 1 were used in the study. Cells were cultured in RPMI with 10% FBS at 37^{o}C under 5% CO_2. The whole experimental module was divided into four different groups. SupT1 cells incubated in the absence (#1) and presence (#2) of nano-curcumin (50μM) for 48 hours. Sup T1 cells infected with HIV-1 in the absence (#3) and the presence of nano-curcumin (#4) for 48 hours. mRNA was isolated from the above samples and processed for microarray analysis at Genotypic Technology Pvt. Ltd, Bangalore, India.

4.2.2 Data analysis and selection of genes:

GeneSpring GX version 11.0 and Microsoft Excel were used for data analysis. For filtering upregulated genes fold change was considered more than one, and for down-regulated genes, it was considered less than minus one. P-value < 0.05 was considered statistically significant. To calculate the fold change in the expression of genes, the level of expression in the control sample (#1) was compared with that of #2, #3 and #4. Clustering of genes was carried out using Hierarchical Clustering Explorer version 3.0 (Figure 4.2). Fold

change in gene expression was derived from the ratio of the magnitude of gene expression in cells under treatment (#2, #3, #4) to that in normal cells (#1). Genes were shortlisted based on a fold change of ±2 in expression when infected with HIV-1 (#3:#1). Amongst these, genes that exhibit a fold change of ±1, when HIV-1 infection conducted in the presence of curcumin (#4:#1) were considered (Figure 4.1).

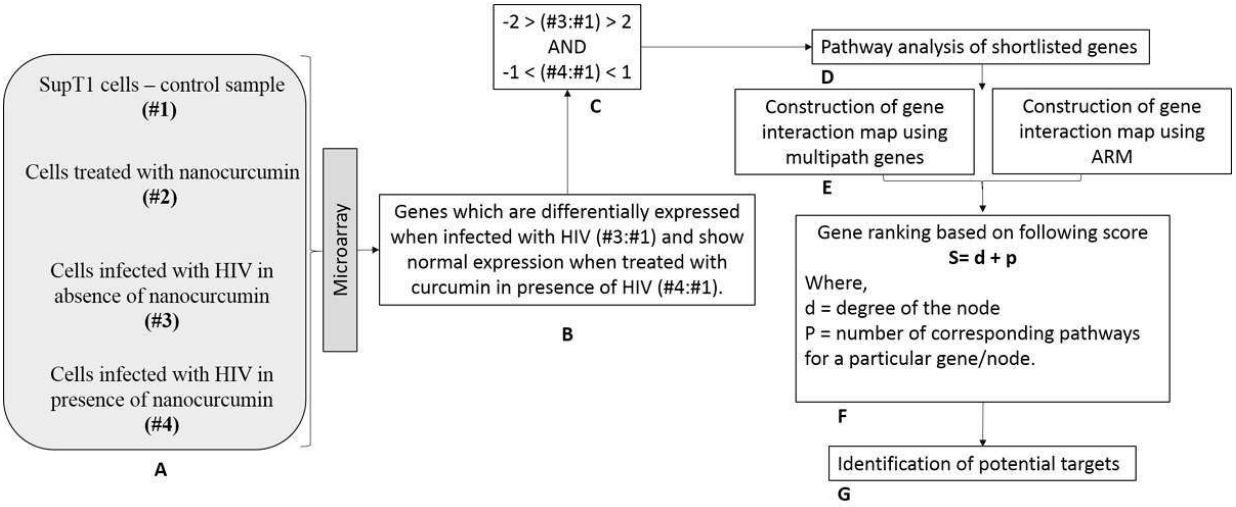

Figure 4.1 *Schematic representation of the methodology. A. samples prepared for microarray. B and C. differentially expressed genes selected whose expression is >2 and <-2 in the cells infected with HIV (#3:#1) and >-1 and <1 (#4:#1) in the cells infected with HIV in the presence of curcumin. D. Pathway analysis was performed with ToppGene server. E. The score assigned to each gene in each network based on their involvement in different pathways. G. Identification of the potential target by comparing the high scoring genes.*

4.2.3 Network construction:

Using the criteria mentioned above, 380 genes were identified which were subjected to pathway analysis. Pathway analysis was done by ToppGene server available at https://toppgene.cchmc.org. The list of Entrez gene IDs of 380 genes was uploaded to the web server, which in turn could return the pathway information of only 76 genes belonging to 90 different pathways with the p-value less than 0.05. List of pathways is given in the Appendix III.

Two methods were followed to construct the gene association networks, first using multipath genes and second using Apriori Algorithm of Association Rule Mining (ARM).

4.2.3.1 Construction of gene network using multipath genes

This is the simplest way to represent the probable biologically related genes in the form of network. Genes belonging to the same pathway are considered to be dependent on each other and hence an edge was placed between two nodes (genes). As there are several genes which are involved in various pathways, we get interconnections among genes. In other words, this network is merely a representation of the probable interactions of biologically related genes.

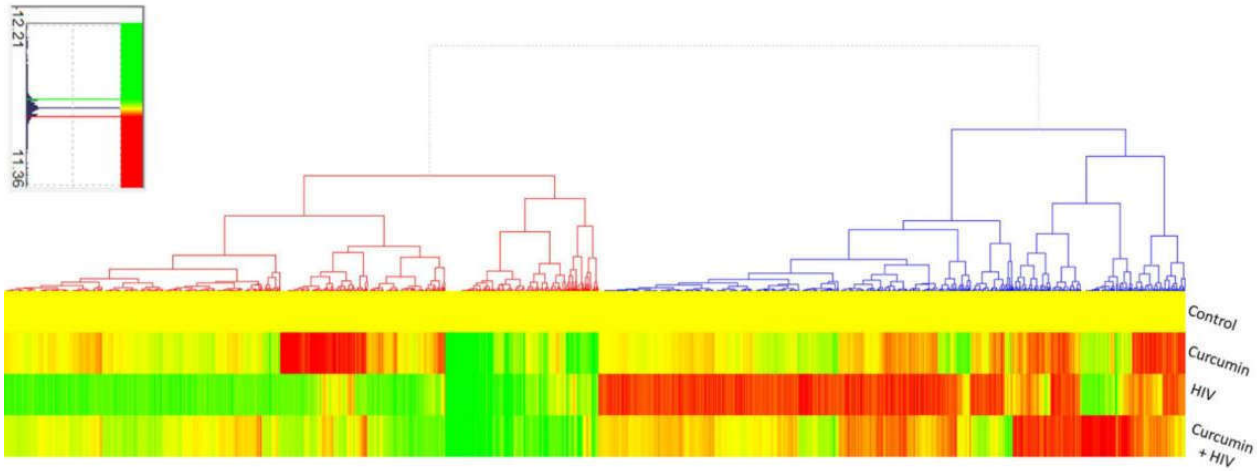

Fig. 4.2. *A heatmap was generated for significantly differentially expressed genes using Hierarchical Clustering Explorer version 3.0. Fold change used for upregulation was >1 and downregulation was <-1. In the figure, normalized expression signals are represented on a log scale for which green color correspond to lower levels of expression and red color correspond to higher levels of expression.*

4. 2.3.2. Gene interaction network using ARM (Apriori algorithm)

Weka data mining tool (version 3.6.13) (WEKA http://www.cs.waikato.ac.nz/~ml/weka) was used to perform ARM, i.e., Apriori algorithm. This requires an input dataset consisting

of transactions and a record of items in a particular transaction. It is a two-step process: First finds all frequent itemsets and then generate strong association rules from these frequent itemsets. Frequent itemsets are those item sets which occur at least as frequently as a predetermined minimum support count. We have considered pathways as transactions and genes as items which form the initial dataset. The association rules obtained from this algorithm elaborates the gene relationship based purely on pathway information and is used to form the gene network.

In the network, an Association rule, Gene_A -> Gene_B is represented as an edge from node Gene_A to Gene_B. We have observed that all the genes involved have bidirectional relations which is shown as a non-directional edge in the network.

The parameters used for this algorithm are as follows:
weka.associations.Apriori -N 1000 -T 0 -C 0.90 -D 0.05 -U 1.0 -M 0.99 -S -1.0 -c -1

- lowerBoundMinSupport: 0.90
- metricType: Confidence
- minMetric: 0.99
- upperBoundMinSupport: 1.0

4.2.3.3 Assigning score to the genes in the network

Genes in each network were ranked based on their interactions and the corresponding number of pathways. A specific score was assigned to each node in the network using following formula.

$$S = d + p \qquad \text{eq-4.1}$$

Where,
d = degree of the node
P = number of corresponding pathways for a particular gene/node.
This pathway-based scoring states that a particular differentially expressed gene affects a high number of genes belonging to several different pathways. Thus the gene itself suggests

of being a relevant candidate under a given microarray experiment. The individual score was calculated for each gene in both types of networks. Later the top scoring genes were compared to identify the potential candidate.

4.3 Results and Discussion

Networks obtained using above mentioned methods are shown in figure 4.3a,b and 4.4 respectively. Comparison of genes from both types of networks revealed several pathways which are reported in the literature for playing a crucial role in the anti-HIV activity of curcumin. Top 10 genes from both the networks based on the score are shown in figure 4.5. The high scoring genes from the first network are mostly the genes which are present in a high number of pathways such as PIK3R3 and PIK3R5. Other high scoring genes belong to Wnt signaling pathway, Jak/STAT signaling pathways and various other pathways. PIK3R3 and PIK3R5 (phosphoinositide-3-kinase, regulatory subunit 3 and 5) are genes up-regulated in HIV-1 infected cells and were down-regulated when HIV-1 infection is conducted in the presence of curcumin. These genes belong to a large number of pathways and hence do not give any clue to target a particular pathway. The high scoring genes from the network obtained using ARM method are mostly belong to related biological function such as regulation of cytoskeleton. The actin cytoskeleton plays a key role during the replication cycle of HIV-1 (Rocha-Perugini *et al.*, 2014). Interaction of viral proteins with the cell modulates the structure and function of the actin cytoskeleton to initiate, sustain and spread infections (Taylor *et al.*, 2011). The comparison of the networks obtained from two different methods shows that ARM-based method highlights more specific pathways while the first method only gives focus on the genes with a high number of pathways. Jak/STAT pathway obtained a high score in both the methods. Indeed, it is well known that HIV infection modulates JAK/STAT signaling pathway (Bovolenta *et al.*, 1999, Miller *et al.*, 2011), and also independently revealed in various studies that curcumin regulate genes in JAK/STAT signaling pathway (Kim *et al.*, 2003). Apparent expression of genes in JAK/STAT pathway observed in microarray was validated using real-time PCR analysis in HIV infection conducted for 4 hours in the presence and absence of curcumin. These results showed that STAT-5 were drastically upregulated during HIV-1 infection, whose

expression was significantly inhibited by treatment with curcumin suggesting that STAT-5 is one of the potential targets of curcumin in inhibiting HIV-1 replication (Fig. 4.6). To address whether the inhibitory pathways of JAK/STAT signaling are also differentially upregulated, modulating the JAK/STAT signaling during HIV-1 infection and curcumin treatment, we have selected some of the SOCS genes. The differential down-regulation of SOCS genes has been observed during HIV-1 infection. Further, upregulation of SOCS genes, especially SOCS-1 and SOCS-3 during curcumin treatment further inactivate JAK/STAT signaling pathway and associated cytokine response, which is crucial for the establishment of acute viral infection.

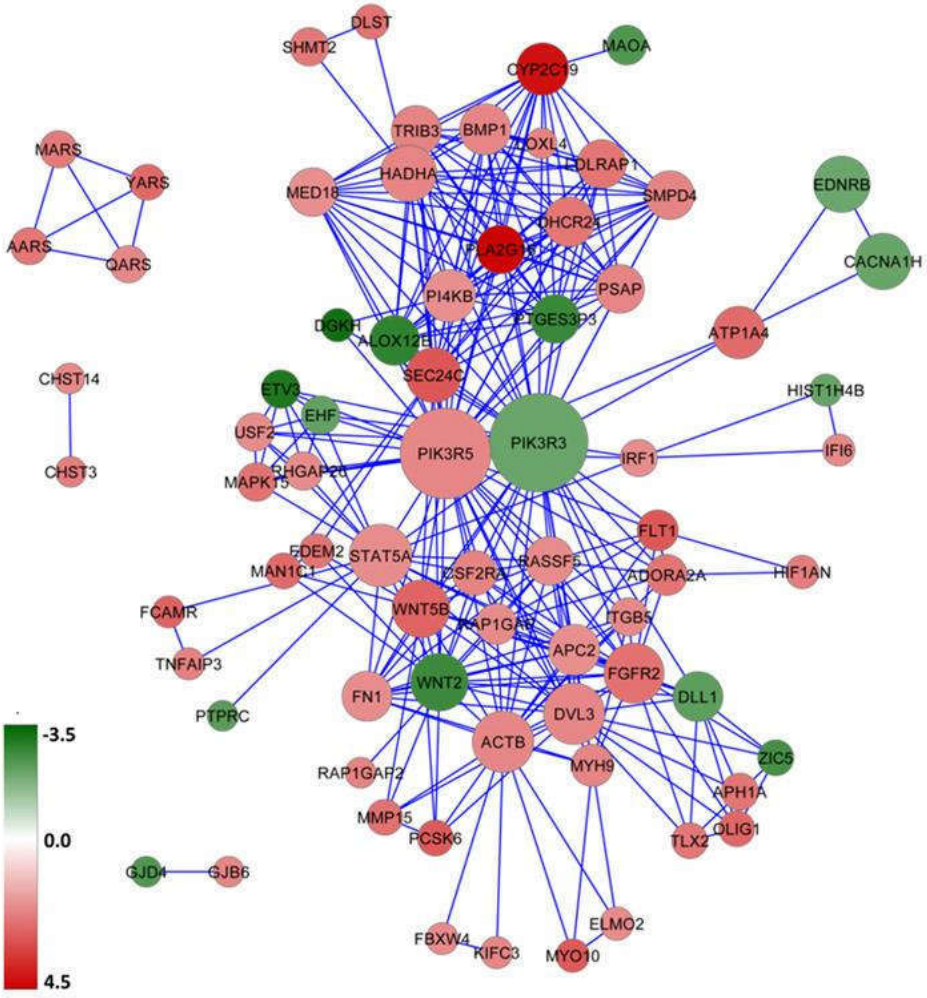

Figure 4.3a: *Gene network for differentially expressed genes using multipath genes when treated with HIV alone.*

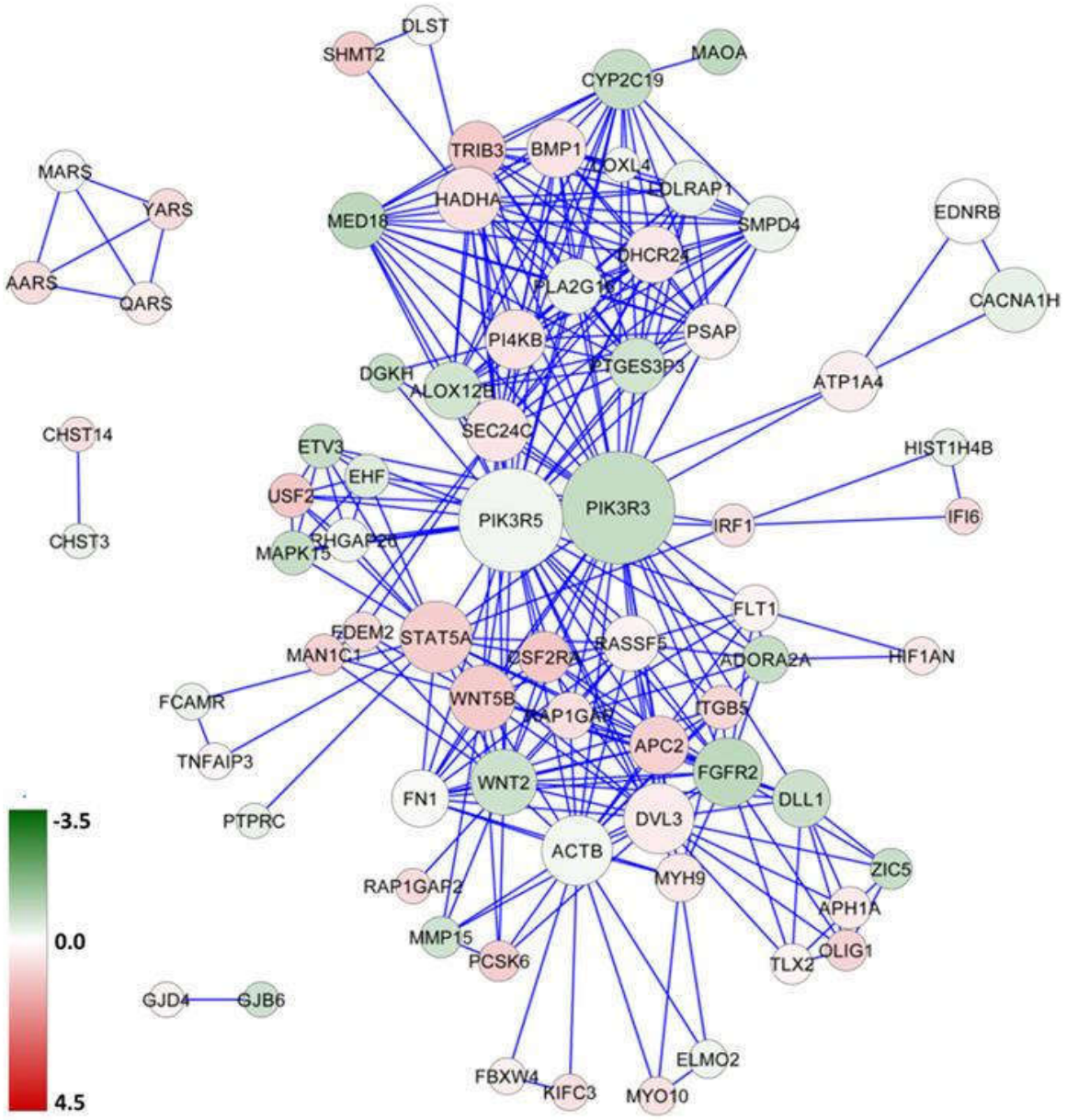

Figure 4.3b: *Gene network for differentially expressed genes using multipath genes when treated with curcumin in presence of HIV. The size of the node represents the corresponding score, and the color represents the fold change expression of the gene. Green color corresponds to the lower expression, and red color corresponds to higher expression of genes. Cytoscape v3.0.0 was used to visualize the gene network (Shannon et al., 2003).*

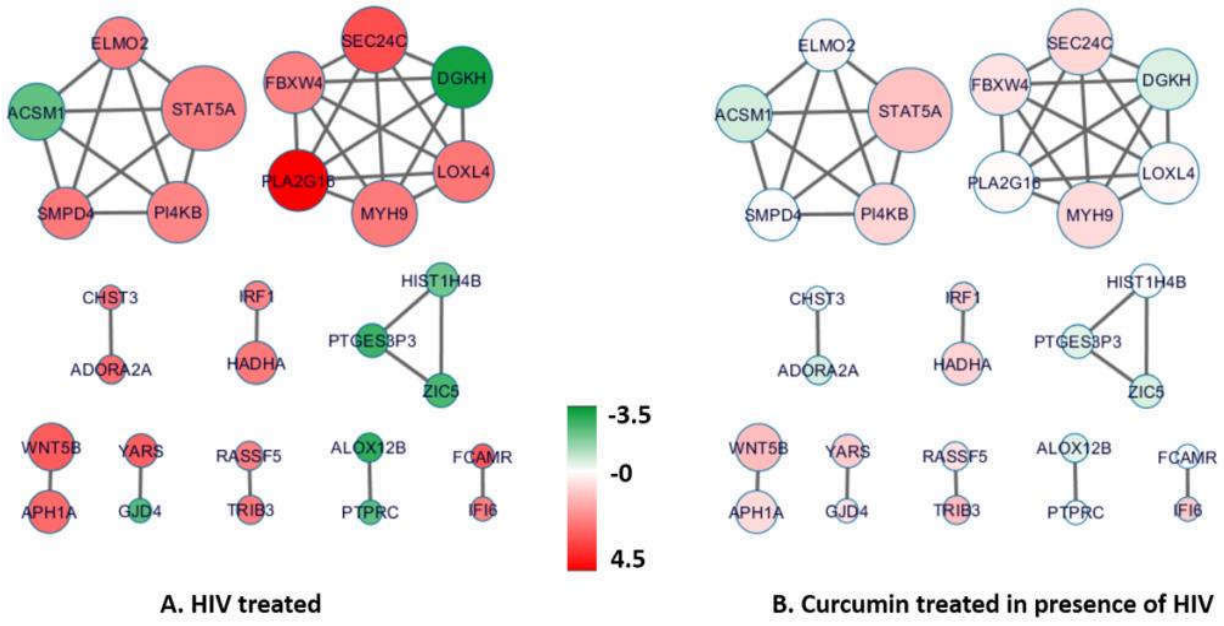

Figure 4.4: *Gene interaction map for differentially expressed genes obtained from ARM (Apriori algorithm). The size of the node represents the corresponding score, and the color represents the fold change expression of the gene. Green color corresponds to the lower expression, and red color corresponds to higher expression of genes.*

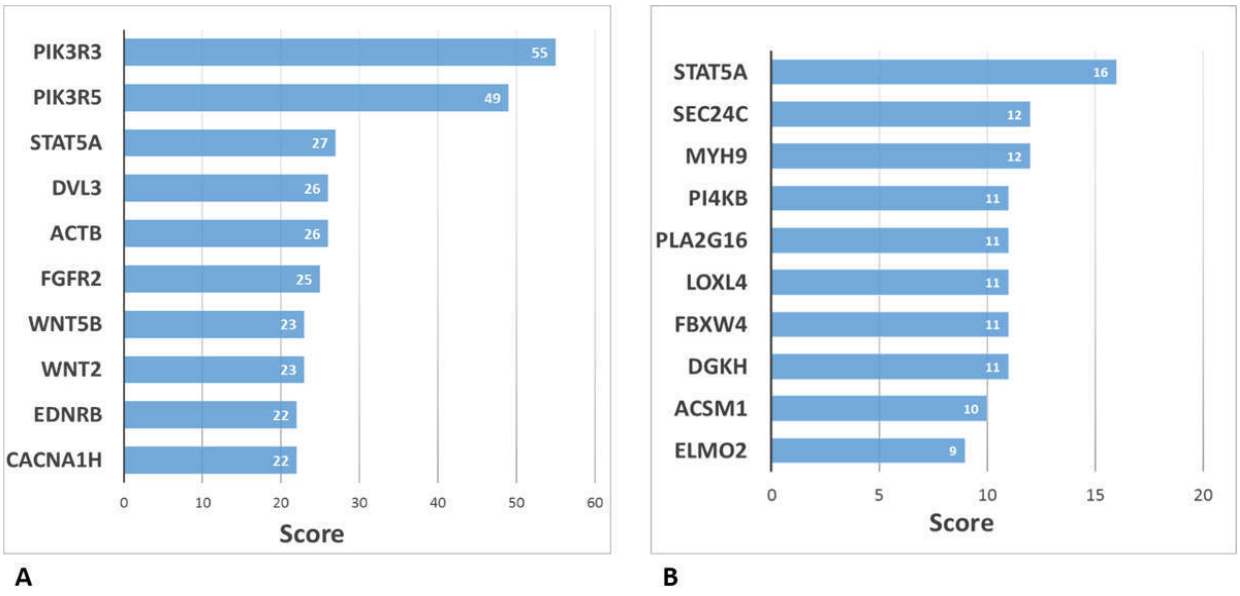

Figure 4.5: *10 highest ranking genes from the network based on their score. The score represents how actively a particular is involved in the biological conditions under which*

microarray experiment was performed. A. Genes from the network constructed using multipath genes. B. Genes from the network constructed using ARM.

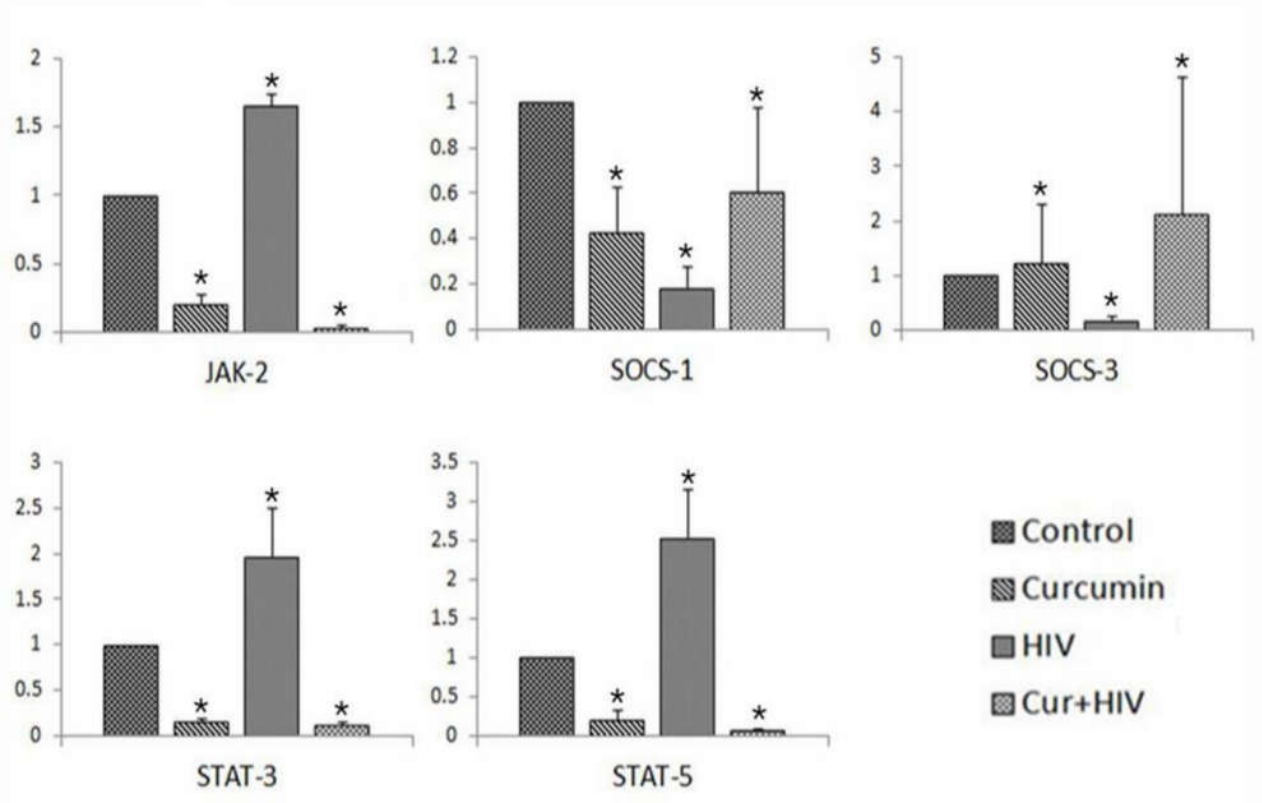

Figure 4.6: *Expression change of JAK/STAT genes confirmed by semiquantitative and real-time PCR. The figure shows the fold change of five genes achieved by real-time PCR. Relative fold change was assayed using the* $2^{-\Delta\Delta CT}$ *method. JAK-2, STAT-3, and STAT-5 genes were upregulated in HIV infection while SOCS-1 and SOCS-3 genes were downregulated.*

4.4 Conclusion:

In order to identify key genes from a large set of differentially expressed genes, we have used the pathway information of differentially expressed genes and generated two types of gene interaction maps first using multipath genes and second by applying ARM. Later we have assigned an individual score to each gene in the network. The score was based on two factors; first, the number of pathways a gene belongs to and second, the number of

interactions in the network. The comparison of the results from both types of networks and top scoring genes led us to select STAT genes for the further validation process. We have validated JAK/STAT pathway and have shown that JAK-2, STAT-3, STAT-5, SOCS-1 and SOCS-3 are potential targets for curcumin to infer its anti-HIV activity which also gives weight to the fact that Curcumin-mediated transcriptional repression of JAK/STAT genes interferes with establishment and progression of viral infection. Thus, in summary, the approach we are proposing here is quite useful in identifying the relevant genes from a large set of differentially expressed genes in a microarray experiment.

Chapter 5

Construction of gene-interaction model in rotenone induced PD in VM neurons

5.1 Introduction

Parkinson's Disease (PD) is a progressive neurodegenerative disease, occurs at high frequency during ageing causing the loss of dopaminergic (DA) neurons in the substantia nigra pars compacta (SNPc) at ventral mesencephalon (VM) (Lang & Lozano, 1998b, a). The loss of DA neurons leads to tremors, slowed movement, balance problems, speech and behavior changes in PD patients. PD is sporadic as well as familial in nature, both forms of PD tend to share important feature like parkinsonism with nigrostriatal DA degeneration (Hardy *et al.*, 2003). The specific etiology of PD is not completely understood, but it is believed to involve both genetic and environmental factors (Kalia & Lang, 2015). Important insights on the contribution of environmental factors in the development of sporadic form were provided through the study of epidemiology and neuropathology of PD (Di Monte *et al.*, 2002).

Various environmental factors have been linked with a greater risk of Parkinson's disease, such as pesticide exposure, head injuries, and living in the farming territory (Noyce *et al.*, 2012, Van Maele-Fabry *et al.*, 2012, de Lau & Breteler, 2006). A study depicted that a significantly increased risk for PD was reported for banana, sugarcane and pineapple plantation workers (Van Maele-Fabry *et al.*, 2012). Heavy metals exposure, pesticides such as rotenone, paraquat, and herbicides such as Agent Orange and ziram, are considered to be a high-risk factor for PD (de Lau & Breteler, 2006, Moretto & Colosio, 2013, Tanner *et al.*, 2011). The effect of rotenone in rats has been reported to cause the development of symptoms similar to those of PD (Caboni *et al.*, 2004, Gao *et al.*, 2003). These studies do not directly propose that rotenone is the causing factor for PD in humans, but certainly suggests that chronic exposure to environmental toxins such as rotenone induces the possibility of the PD like symptoms (Goldman, 2014, Sherer *et al.*, 2002).

Rotenone is an odorless, colorless, flavonoid widely used as an insecticide and pesticide (Isman, 2006). It is naturally occurring substance in the seeds and stems of several plants, such as the jicama vine plant, and the roots of several members of Fabaceae. Rotenone induces certain features of PD both in vitro and in vivo (Betarbet *et al.*, 2000, Chaves *et al.*,

2010). Several cellular processes such as Oxidative stress, mitochondrial dysfunction and misfolded protein aggregation is reported to play a key role in the pathogenesis of PD (Moore *et al.*, 2005). Earlier studies also suggest dysfunction of mitochondrial complex-I in PD patients (Gatt *et al.*, 2016, Schapira *et al.*, 1990).

Parkinson's disease is a complex mechanism and genetic factors and pathways associated with PD are not fully understood. Several studies have been done to delineate the pathways involved in PD (Yang *et al.*, 2009, Chandrasekaran & Bonchev, 2013). In this study, we have analyzed the differentially expressed genes in rotenone-treated dopaminergic neurons (DA) in the substantia nigra at ventral mesencephalon (VM), cultured *in-vitro* and performed a network analysis based on the pathway classification using association rule mining.

5.2 Materials and Methods

5.2.1 Animals

Pregnant female Wistar rats were procured from National Institute of Nutrition, Hyderabad, India and maintained at the animal house facility according to norms of Institutional animal ethical committee (IAEC), University of Hyderabad (Proposal number LS/IAEC/AKK/10/1).

5.2.2 Isolation and culture of VM neurons

Anesthetized pregnant Wistar rats (E14) were decapitated, and embryonic pups were removed aseptically by a 'C' section. SNc tissue from embryonic mesencephalon was collected and washed with Hank's Balanced Salt Solution (HBSS) at 4^0C and incubated in 2ml of pre-warmed 2.5mg/ml trypsin solution (Gibco, NY, USA) for 15 min at room temperature. After trypsinization, the tissue was resuspended in 2ml of pre-warmed DNAse (10µg/ml), and the solution was centrifuged at 1200g for 4 min. Then Supernatant was removed, and the tissue was triturated with 0.1- 0.2ml triturating solution (1mg/ml BSA, 10µg/ml DNAse 1, 0.5mg/ml soybean trypsin inhibitor in HBSS / piece of VM tissue).

The suspension was placed on ice and triturated using a fire-polished glass pipette with a minimal number of strokes to obtain a suspension of single cells. Three pipettes of decreasing tip diameters were used, and the entire suspension was passed through each pipette tip 2-4 times and triturated slowly without any air bubbles. VM neurons were plated in the presence of Dulbecco's Modified Eagle Medium/Nutrient Mixture F-12 (DMEM F-12) containing 10% fetal bovine serum, (Gibco, NY, USA). Cultures were incubated in a humidified atmosphere of 5% CO_2 at 37°C. 1 x 10^6 cells were seeded in 1ml of medium per well in a 12 well plate coated with 0.1mg/ml Poly-D- Lysine (PDL) (Sigma Chemical Co, MO, USA). For other culture dishes, cells were seeded proportionally. Cultures were maintained by replacing half the volume of medium every alternate day. 2μM arabinosylcytosine (Ara-C, mitotic inhibitor) (Sigma Chemical Co, MO, USA) was added to cultures from the second day of culture to suppress the proliferation of mitotic non-neuronal cells and replenished every alternate day. In rotenone and curcumin treated cultures, the compounds were dissolved in culture grade DMSO (Sigma Chemical Co, MO, USA). Rotenone was added to the cultures in a single shot at specified concentrations and at specified time points and treated for 48 hrs.

5.2.3 Microarray and data analysis

VM neurons at 7th DIV were treated with 15nM rotenone, and the cells were processed at 9th DIV for microarray analysis. VM neurons at 7th DIV treated with DMSO alone and processed at 9th DIV were taken as control. The samples were collected in triplicate and processed for microarray analysis at Genotypic Technology Pvt Ltd, Bangalore, India, and using whole rat genome microarray kit provided by Agilent. Briefly, total RNA was isolated using Quiagen RNeasy mini-kit with DNAse treatment. Purity and concentration of RNA were estimated using Nanodrop spectrophotometer. Following labeling and hybridization samples were scanned for fluorescent signals. A fold change of 0.6 was used to detect the up-regulation and down-regulation in each of the treated replicates with a geomean fold of 0.8. Expression signals were normalized using Genespring GX 12.6.1 software and were presented on a log scale, where lower levels of expression were represented in colder colours and higher levels of expression was represented in warmer colours.

Overall Summary

As mentioned in the Introduction part we have divided the whole work into three parts with a total of four objectives (chapter 2 to 5 respectively). Following are the summarized points from each objective.

Analysis of differentially expressed genes during neuronal senescence and identification of potential markers.

- We have studied two different methods to reconstruct gene networks from microarray data
    - Differential equation model
    - Supervised learning method
- Differential equation models are very much sensitive to the dimension of the data and do not work for high dimension data where a number of genes are more than the number of experiments.
- Knowledge-driven Approach: We used biological categorization of genes to overcome the dimensional problem of microarray data. *(Dholaniya et. al. 2015)*
- We have established Npy and Slit2 as potential markers for senescence in cultures CGN's. *(Gupta and Dholaniya et. al. 2015)*

Construction of hypothetical model showing the role of top2b in neuronal maturation and axon guidance.

- A database was created using available literature information about topoisomerase 2 beta.
- A hypothetical model explaining the role of top2 beta in axon guidance and neuron maturation were developed.
- A hypothesis was established, showing how Npy and Slit2 inhibition may increase the expression of Top2b and Top2b expression was found significantly increased in Npy and Slit2 silenced CGNs.
- The database can also be used to develop such top2b centric models for any other particular disease/function.

Identification of crucial pathways involved in the anti-HIV activity of curcumin.

- A pathway-based scoring approach with Association Rule Mining was developed to identify relevant genes from DNA microarray of Human SUP-T1 cells treated with HIV and nano-curcumin
- Analysis of network revealed several crucial pathways which may have a significant role in the anti-HIV activity of curcumin. Such as Jak/Stat pathway, Wnt signaling pathway and pathways involved in regulation of actin cytoskeleton.
- It was validated that the anti-HIV response of curcumin could be the result of its suppressive action on Jak/Stat signaling pathway by the up-regulation of SOCS genes (such as SOCS3 and SOCS6) and down-regulation of JAK and STAT genes.

Construction of gene interaction model in rotenone-induced PD in VM neurons.

- Rotenone-induced Parkinson's disease was studied in DA neurons.
- Differentially expressed genes were selected and gene network was constructed using Association Rule Mining.
- Association rule mining find the relationships among items based on their appearance in transactions, we used have used the same concept, replacing genes as items and pathways as transactions to identify the relationship between genes.
- The functional analysis of the network suggests that a large number of genes belong to chemokine and cytokine-mediated inflammation.
- Few genes such as complement-3, interleukin-1 beta, Tnf, and Toll-Like Receptors observed to have a high score in the network and are very well shown to be involved in PD from previous studies.

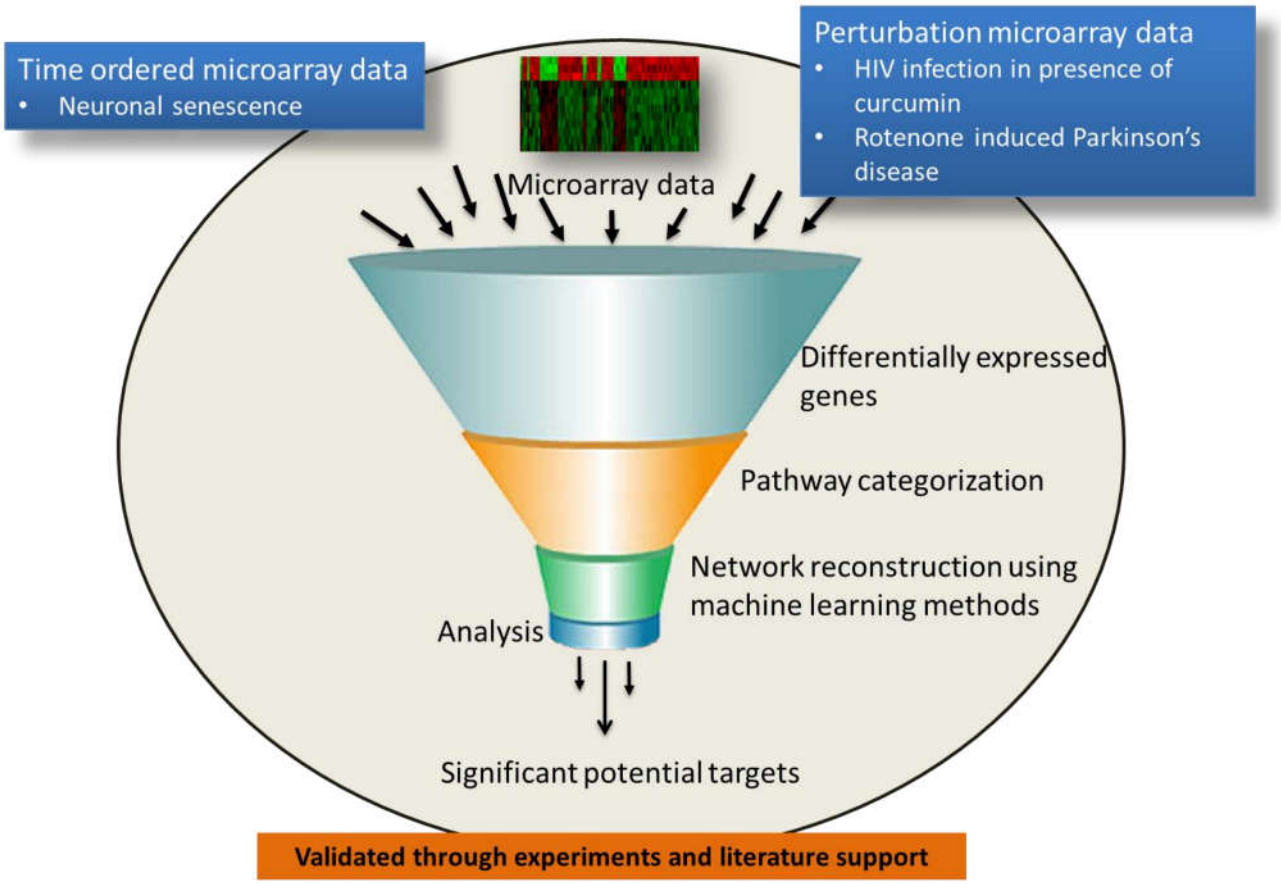

Figure 6.1: *The overall approach of the study to list out significant potential marker from a large set of gene expression data*

Overall conclusion:

The goal of this study was to study different approaches to reduce the search space to identify the probable marker of any microarray expression data. As shown in figure 6.1, we have used the biological information along with the machine learning methods to fetch the relationship among genes, which can be depicted in the form of networks. Further, we have also shown that how the network analysis can be used to identify the significant potential targets corresponding to a particular microarray experiment.

References

Aggarwal, B. B., Kumar, A. & Bharti, A. C. (2003). Anticancer research 23, 363-398.

Agrawal, R., Imielinski, T. & Swami, A. (1993). Proceedings of the 1993 ACM SIGMOD Conference Washington DC, USA.

Akaike, H. (1974). IEEE Transactions on Automatic Control 19, 716 - 723.

Bahbouhi, B., Landay, A. & Al-Harthi, L. (2004). Blood 103, 4581-4587.

Barthelemy, S., Vergnes, L., Moynier, M., Guyot, D., Labidalle, S. & Bahraoui, E. (1998). Research in virology 149, 43-52.

Bellman, R. (1966). Science 153, 34-37.

Ben-Dor, A., Shamir, R. & Yakhini, Z. (1999). Journal of computational biology : a journal of computational molecular cell biology 6, 281-297.

Betarbet, R., Sherer, T. B., MacKenzie, G., Garcia-Osuna, M., Panov, A. V. & Greenamyre, J. T. (2000). Nature neuroscience 3, 1301-1306.

Beutler, B. & Bazzoni, F. (1998). Blood cells, molecules & diseases 24, 216-230.

Bhanu, M. U., Mandraju, R. K., Bhaskar, C. & Kondapi, A. K. (2010). Toxicology in vitro : an international journal published in association with BIBRA 24, 1935-1945.

Bhattacharyya, S., Md Sakib Hossain, D., Mohanty, S., Sankar Sen, G., Chattopadhyay, S., Banerjee, S., Chakraborty, J., Das, K., Sarkar, D., Das, T. & Sa, G. (2010). Cellular & molecular immunology 7, 306-315.

Bishop, N. A., Lu, T. & Yankner, B. A. (2010). Nature 464, 529-535.

Boka, G., Anglade, P., Wallach, D., Javoy-Agid, F., Agid, Y. & Hirsch, E. C. (1994). Neuroscience letters 172, 151-154.

Bovolenta, C., Camorali, L., Lorini, A. L., Ghezzi, S., Vicenzi, E., Lazzarin, A. & Poli, G. (1999). Blood 94, 4202-4209.

Burda, Z., Krzywicki, A., Martin, O. C. & Zagorski, M. (2011). Proceedings of the National Academy of Sciences of the United States of America 108, 17263-17268.

Caboni, P., Sherer, T. B., Zhang, N., Taylor, G., Na, H. M., Greenamyre, J. T. & Casida, J. E. (2004). Chemical research in toxicology 17, 1540-1548.

Capranico, G., Tinelli, S., Austin, C. A., Fisher, M. L. & Zunino, F. (1992). Biochimica et biophysica acta 1132, 43-48.

Champoux, J. J. (2001). Annual review of biochemistry 70, 369-413.

Chandrasekaran, S. & Bonchev, D. (2013). Computational and structural biotechnology journal 7, e201304004.

Chang, Y. H., Gray, J. W. & Tomlin, C. J. (2014). BMC bioinformatics 15, 400.

Chaves, R. S., Melo, T. Q., Martins, S. A. & Ferrari, M. F. (2010). BMC neuroscience 11, 144.

Chen, G., Cairelli, M. J., Kilicoglu, H., Shin, D. & Rindflesch, T. C. (2014). PLoS computational biology 10, e1003666.

Chen, T., He, H. L. & Church, G. M. (1999). Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing, 29-40.

Ciftci, O., Tanyildizi, S. & Godekmerdan, A. (2010). Immunopharmacology and immunotoxicology 32, 99-104.

Contestabile, A. (2002). Cerebellum 1, 41-55.

Culmsee, C. & Mattson, M. P. (2005). Biochemical and biophysical research communications 331, 761-777.

de Hoon, M. J., Imoto, S., Kobayashi, K., Ogasawara, N. & Miyano, S. (2003). Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing, 17-28.

de Lau, L. M. & Breteler, M. M. (2006). The Lancet. Neurology 5, 525-535.

De, R., Kundu, P., Swarnakar, S., Ramamurthy, T., Chowdhury, A., Nair, G. B. & Mukhopadhyay, A. K. (2009). Antimicrobial agents and chemotherapy 53, 1592-1597.

Di Monte, D. A., Lavasani, M. & Manning-Bog, A. B. (2002). Neurotoxicology 23, 487-502.

Droge, W. & Schipper, H. M. (2007). Aging cell 6, 361-370.

Eisen, M. B., Spellman, P. T., Brown, P. O. & Botstein, D. (1998). Proceedings of the National Academy of Sciences of the United States of America 95, 14863-14868.

Farooqui, T. & Farooqui, A. A. (2011). Parkinson's disease 2011, 247467.

Federico, M., Percario, Z., Olivetta, E., Fiorucci, G., Muratori, C., Micheli, A., Romeo, G. & Affabris, E. (2001). Blood 98, 2752-2761.

Fischer, R. & Maier, O. (2015). Oxidative medicine and cellular longevity 2015, 610813.

Friedman, N., Linial, M., Nachman, I. & Pe'er, D. (2000). Journal of computational biology : a journal of computational molecular cell biology 7, 601-620.

Gallagher, M., Stocker, A. M. & Koh, M. T. (2011). ILAR journal / National Research Council, Institute of Laboratory Animal Resources 52, 32-40.

Gandapu, U., Chaitanya, R. K., Kishore, G., Reddy, R. C. & Kondapi, A. K. (2011). PloS one 6, e23388.

Gao, H. M., Liu, B. & Hong, J. S. (2003). The Journal of neuroscience : the official journal of the Society for Neuroscience 23, 6181-6187.

Gatt, A. P., Duncan, O. F., Attems, J., Francis, P. T., Ballard, C. G. & Bateman, J. M. (2016). Movement disorders : official journal of the Movement Disorder Society 31, 352-359.

Giudici, P. (2005). Applied Data Mining: Statistical Methods for Business and Industry, 1 ed.: Wiley;.

Goldman, S. M. (2014). Annual review of pharmacology and toxicology 54, 141-164.

Gu, Z., Kaul, M., Yan, B., Kridel, S. J., Cui, J., Strongin, A., Smith, J. W., Liddington, R. C. & Lipton, S. A. (2002). Science 297, 1186-1190.

Gupta, K. P., Swain, U., Rao, K. S. & Kondapi, A. K. (2012). Mech Ageing Dev 133, 203-213.

Gupta, S. C., Prasad, S., Kim, J. H., Patchva, S., Webb, L. J., Priyadarsini, I. K. & Aggarwal, B. B. (2011). Natural product reports 28, 1937-1955.

Han, J. W., Pei, J. & Yin, Y. W. (2000). Sigmod Record 29, 1-12.

Han, J. W., Pei, J., Yin, Y. W. & Mao, R. Y. (2004). Data Min Knowl Disc 8, 53-87.

Hardy, J., Cookson, M. R. & Singleton, A. (2003). The Lancet. Neurology 2, 221-228.

Hayflick, L. (1965). Experimental cell research 37, 614-636.

Hecker, M., Lambeck, S., Toepfer, S., van Someren, E. & Guthke, R. (2009). Bio Systems 96, 86-103.

Herbein, G., Gras, G., Khan, K. A. & Abbas, W. (2010). Retrovirology 7, 34.

Hoch, R. V. & Soriano, P. (2003). Development 130, 4769-4784.

Imoto, S., Goto, T. & Miyano, S. (2002). Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing, 175-186.

Isman, M. B. (2006). Annual review of entomology 51, 45-66.

Jellinger, K. A. & Attems, J. (2013). Dialogues in clinical neuroscience 15, 29-43.

Jiawei Han, M. K., Jian Pei (2011). Data Mining: Concepts and Techniques, 3 ed., pp. 243-278: Elsevier Inc.

Jurenka, J. S. (2009). Alternative medicine review : a journal of clinical therapeutic 14, 141-153.

Kalia, L. V. & Lang, A. E. (2015). Lancet 386, 896-912.

Kanehisa, M. & Goto, S. (2000). Nucleic acids research 28, 27-30.

Khalid Raza, R. P. (2012). Journal of Advanced Bioinformatics Applications and Research 3, 271-280.

Kim, H. Y., Park, E. J., Joe, E. H. & Jou, I. (2003). J Immunol 171, 6072-6079.

Kohler, J. J., Tuttle, D. L., Coberley, C. R., Sleasman, J. W. & Goodenow, M. M. (2003). Journal of leukocyte biology 73, 407-416.

Kondapi, A. K., Mulpuri, N., Mandraju, R. K., Sasikaran, B. & Subba Rao, K. (2004). International journal of developmental neuroscience : the official journal of the International Society for Developmental Neuroscience 22, 19-30.

Kuo, M. L., Huang, T. S. & Lin, J. K. (1996). Biochimica et biophysica acta 1317, 95-100.

Kutluay, S. B., Doroghazi, J., Roemer, M. E. & Triezenberg, S. J. (2008). Virology 373, 239-247.

Lang, A. E. & Lozano, A. M. (1998a). The New England journal of medicine 339, 1044-1053.

Lang, A. E. & Lozano, A. M. (1998b). The New England journal of medicine 339, 1130-1143.

Laplante, M. & Sabatini, D. M. (2012). Cell 149, 274-293.

Liu, M. Q., Zhou, D. J., Wang, X., Zhou, W., Ye, L., Li, J. L., Wang, Y. Z. & Ho, W. Z. (2012). PloS one 7, e35902.

Loeffler, D. A., Camp, D. M. & Conant, S. B. (2006). Journal of neuroinflammation 3, 29.

Mallik, S., Mukhopadhyay, A. & Maulik, U. (2015). IEEE Transactions on NanoBioscience, 59 - 66.

Martin, P., Mora, I., Cortes, M. A., Calleros, L., Garcia-Jerez, A., Ortiz, A., Rodriguez-Puyol, M., Rodriguez-Puyol, D. & Olmos, G. (2014). American journal of physiology. Renal physiology 307, F75-85.

Martins, C. V., da Silva, D. L., Neres, A. T., Magalhaes, T. F., Watanabe, G. A., Modolo, L. V., Sabino, A. A., de Fatima, A. & de Resende, M. A. (2009). The Journal of antimicrobial chemotherapy 63, 337-339.

Mattson, M. P. & Magnus, T. (2006). Nature reviews. Neuroscience 7, 278-294.

Mazumder, A., Raghavan, K., Weinstein, J., Kohn, K. W. & Pommier, Y. (1995). Biochemical pharmacology 49, 1165-1170.

McCoy, M. K. & Tansey, M. G. (2008). Journal of neuroinflammation 5, 45.

Miller, R. C., Schlaepfer, E., Baenziger, S., Crameri, R., Zeller, S., Byland, R., Audige, A., Nadal, D. & Speck, R. F. (2011). European journal of immunology 41, 1058-1069.

Misra, J., Chanda, D., Kim, D. K., Li, T., Koo, S. H., Back, S. H., Chiang, J. Y. & Choi, H. S. (2011). The Journal of biological chemistry 286, 41972-41984.

Moghadamtousi, S. Z., Kadir, H. A., Hassandarvish, P., Tajik, H., Abubakar, S. & Zandi, K. (2014). BioMed research international 2014, 186864.

Mogi, M., Harada, M., Riederer, P., Narabayashi, H., Fujita, K. & Nagatsu, T. (1994). Neuroscience letters 165, 208-210.

Moore, D. J., West, A. B., Dawson, V. L. & Dawson, T. M. (2005). Annual review of neuroscience 28, 57-87.

Moretto, A. & Colosio, C. (2013). Toxicology 307, 24-34.

Nikoletopoulou, V. & Tavernarakis, N. (2012). Frontiers in genetics 3, 200.

Noyce, A. J., Bestwick, J. P., Silveira-Moriyama, L., Hawkes, C. H., Giovannoni, G., Lees, A. J. & Schrag, A. (2012). Annals of neurology 72, 893-901.

Olmos, G. & Llado, J. (2014). Mediators of inflammation 2014, 861231.

Osheroff, N., Zechiedrich, E. L. & Gale, K. C. (1991). BioEssays : news and reviews in molecular, cellular and developmental biology 13, 269-273.

Piwocka, K., Zablocki, K., Wieckowski, M. R., Skierski, J., Feiga, I., Szopa, J., Drela, N., Wojtczak, L. & Sikora, E. (1999). Experimental cell research 249, 299-307.

Przedborski, S. (2010). Movement disorders : official journal of the Movement Disorder Society 25 Suppl 1, S55-57.

Qin, Z. S. (2006). Bioinformatics 22, 1988-1997.

Rechtman, M. M., Har-Noy, O., Bar-Yishay, I., Fishman, S., Adamovich, Y., Shaul, Y., Halpern, Z. & Shlomai, A. (2010). FEBS letters 584, 2485-2490.

Rocha-Perugini, V., Gordon-Alonso, M. & Sanchez-Madrid, F. (2014). Trends in microbiology 22, 379-388.

Saiki, S. (2014). Rinsho shinkeigaku = Clinical neurology 54, 1125-1127.

Schapira, A. H., Cooper, J. M., Dexter, D., Clark, J. B., Jenner, P. & Marsden, C. D. (1990). Journal of neurochemistry 54, 823-827.

Schena, M., Shalon, D., Davis, R. W. & Brown, P. O. (1995). Science 270, 467-470.

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B. & Ideker, T. (2003). Genome research 13, 2498-2504.

Shehzad, A. & Lee, Y. S. (2013). Biofactors 39, 27-36.

Sherer, T. B., Betarbet, R., Stout, A. K., Lund, S., Baptista, M., Panov, A. V., Cookson, M. R. & Greenamyre, J. T. (2002). The Journal of neuroscience : the official journal of the Society for Neuroscience 22, 7006-7015.

Si, X., Wang, Y., Wong, J., Zhang, J., McManus, B. M. & Luo, H. (2007). Journal of virology 81, 3142-3150.

Singh, S. & Aggarwal, B. B. (1995). The Journal of biological chemistry 270, 24995-25000.

Soinov, L. A., Krestyaninova, M. A. & Brazma, A. (2003). Genome biology 4, R6.

Spellman, P. T., Sherlock, G., Zhang, M. Q., Iyer, V. R., Anders, K., Eisen, M. B., Brown, P. O., Botstein, D. & Futcher, B. (1998). Molecular biology of the cell 9, 3273-3297.

Sui, Z., Salto, R., Li, J., Craik, C. & Ortiz de Montellano, P. R. (1993). Bioorganic & medicinal chemistry 1, 415-422.

Szalay-Beko, M., Palotai, R., Szappanos, B., Kovacs, I. A., Papp, B. & Csermely, P. (2012). Bioinformatics 28, 2202-2204.

Taher, M. M., Lammering, G., Hershey, C. & Valerie, K. (2003). Molecular and cellular biochemistry 254, 289-297.

Tanner, C. M., Kamel, F., Ross, G. W., Hoppin, J. A., Goldman, S. M., Korell, M., Marras, C., Bhudhikanok, G. S., Kasten, M., Chade, A. R., Comyns, K., Richards, M. B., Meng, C., Priestley, B., Fernandez, H. H., Cambi, F., Umbach, D. M., Blair, A., Sandler, D. P. & Langston, J. W. (2011). Environmental health perspectives 119, 866-872.

Taylor, M. P., Koyuncu, O. O. & Enquist, L. W. (2011). Nature reviews. Microbiology 9, 427-439.

Thomas, P. D., Campbell, M. J., Kejariwal, A., Mi, H., Karlak, B., Daverman, R., Diemer, K., Muruganujan, A. & Narechania, A. (2003). Genome research 13, 2129-2141.

Thomas, S. A. & Jin, Y. (2014). Evolutionary Intelligence 7, 29-47.

Tomita, M., Kawakami, H., Uchihara, J. N., Okudaira, T., Masuda, M., Matsuda, T., Tanaka, Y., Ohshiro, K. & Mori, N. (2006). Retrovirology 3, 22.

Tomita, M., Matsuda, T., Kawakami, H., Uchihara, J. N., Okudaira, T., Masuda, M., Ohshiro, K. & Mori, N. (2006). Cancer science 97, 322-327.

Tsutsui, K., Sano, K., Kikuchi, A. & Tokunaga, A. (2001). The Journal of biological chemistry 276, 5769-5778.

Van Maele-Fabry, G., Hoet, P., Vilain, F. & Lison, D. (2012). Environment international 46, 30-43.

Wang, J. C. (1991). The Journal of biological chemistry 266, 6659-6662.

Wang, J. C. (1996). Annual review of biochemistry 65, 635-692.

Wang, Y., Joshi, T., Zhang, X. S., Xu, D. & Chen, L. (2006). Bioinformatics 22, 2413-2420.

Watson, J. D. & Crick, F. H. (1953a). Cold Spring Harbor symposia on quantitative biology 18, 123-131.

Watson, J. D. & Crick, F. H. (1953b). Nature 171, 737-738.

Wessels, L. F., van Someren, E. P. & Reinders, M. J. (2001). Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing, 508-519.

Yang, J. O., Kim, W. Y., Jeong, S. Y., Oh, J. H., Jho, S., Bhak, J. & Kim, N. S. (2009). BMC genomics 10 Suppl 3, S32.

Zaki, M. J. (2000). Ieee T Knowl Data En 12, 372-390.

Zhang, H. S., Ruan, Z. & Sang, W. W. (2011). Journal of cellular physiology 226, 3385-3391.

Appendix

Appendix I

Decision rules

Explaining Gene		Predicting Gene
-cpt1b	->	-Acsl6
+Acsl6	->	+cpt1b
+Acsl6	->	+Cpt2
+Npr2	->	+Adcy8
+Npr2	->	+Entpd1
-Npr2	->	-Gda
+Gda	->	+Npr2
-Hsd3b7	->	-Dhrs3
-dhrs3	->	-Hsd3b7
-dhrs3	->	-Pon2
+fuca2	->	+Aga
-Aga	->	-Fuca2
-Aga	->	-Hexa
-sphk1	->	-Asah1
-sphk1	->	-Smpd2
-Smpd2	->	-Sphk1
-Nmnat3	->	-Bst1
+nnt	->	+Nmnat3
-Nmnat3	->	-Nnt
-Fuca2	->	-Aga
-Aga	->	-Fuca2
-Aga	->	-Hexa
-Gja1	->	-Col1a1
+Gja1	->	+Col1a2
-Gja1	->	-Col3a1
+Col3a1	->	+Col5a1
+Col3a1	->	+Col5a2
+Gja1	->	+Col5a3
+Gja1	->	+Col6a2
+Gja1	->	+Fn1
-col1a1	->	-Gja1
+col1a1	->	+Gja5
-Gja1	->	+Gjb1
-Gja1	->	-Gjb2
+Gja1	->	+Lama5
-Gja1	->	-Lamb2
+Gja1	->	+Lamc1
+Gja1	->	+Thbs2
-Gja1	->	-Vim
-abcc8	->	-Abcb9

Explaining Gene		Predicting Gene
+edg2	->	+Gria1
+edg2	->	+Grin2c
-gabra6	->	-Grm3
-gabra6	->	-Grm4
-gabra6	->	-Grm8
+cckbr	->	+Lhb
-Lhb	->	-P2rx2
-Lhb	->	+Uts2r
-igf1	->	-Ccnd1
+ccnd1	->	+Igf1
-igf1	->	+Igfbp3
+Nfatc4	->	+Ccnd1
+Sfrp4	->	+Nfatc4
-nfatc4	->	-Sfrp4
+nfatc4	->	+Wif1
-bmp6	->	-Bmp4
+Bmp4	->	+bmp6
-bmp6	->	-Bmp4
-Bmp4	->	-bmp6
-thbs2	->	-Dcn
+bmp4	->	+Thbs2
-nfatc4	->	+Ablim2
+nfatc4	->	+Ephb6
+slit2	->	+Nfatc4
-nfatc4	->	-Slit2
-nfatc4	->	-Slit3
-igf1	->	-Cav3
+pdgfrb	->	+Ccnd1
-col5a2	->	-Col1a1
-col5a2	->	-Col1a2
-col5a2	->	-Col3a1
-cav3	->	-Col5a1
+col1a1	->	+Col5a2
+ccnd1	->	+Col5a3
-col5a2	->	-Col6a2
-col5a2	->	-Fn1
+pdgfrb	->	+Igf1
-col5a2	->	-Itga1
-col5a2	->	-Lama5
-cav3	->	-Lamb2

Explaining Gene		Predicting Gene
+a2m	->	+F5
+plau	->	-Kng1
-f3	->	-Masp1
-c1r	->	+Plau
+plau	->	-Serping1
+Tap2	->	+Cd74
+Cd74	->	+Hspa1a
-Cd74	->	-Tap2
-Tlr3	->	-Lbp
-lbp	->	-Tlr3
+Il11ra1	->	+Ccnd1
-Il11ra1	->	-Il15
+Il11ra1	->	+Stat5a
-Il11ra1	->	-Cd1d1
+Il11ra1	->	+Csf1
+cd1d1	->	+Il11ra1
+Cd1d1	->	+Itga1
+Il11ra1	->	+Kitl
-Icam1	->	-Cd48
+cd48	->	+Icam1
-icam1	->	+Lat
-icam1	->	-Nfatc4
-Mmp2	->	-Cldn11
+Vcam1	->	+Icam1
-cldn11	->	-Mmp2
+icam1	->	+Ocln
-ocln	->	+Thy1
-icam1	->	-Vcam1
+gria1	->	+Adcy8
-grin2c	->	-Gria1
+gria1	->	+Grin2c
-Grin2c	->	-Itpr1
+itpr1	->	+Gria1
-itpr1	->	+Igf1
-Igf1	->	+Itpr1
+Fgfr2	->	+Fgf2
-Fgf2	->	-Fgfr1
+Fgfr1	->	+Fgfr2
-Fgfr2	->	-Fn1
-Fgfr1	->	-Gsn

-abcb9	->	-Abcc8
+tap2	->	+Abcg3
+abcb9	->	-Tap2
-Angptl4	->	-Acsl6
-Acsl6	->	-Angptl4
-cpt2	->	-Cpt1b
+Acsl6	->	-Cpt2
+nfatc4	->	+Fgf2
-pdgfrb	->	-Fgfr1
-pdgfrb	->	-Fgfr2
+pdgfrb	->	+Hspa1a
-pdgfrb	->	-Nfatc4
-Fgf2	->	-Pdgfrb
+pdgfrb	->	-Ptpn5
-gna14	->	-Adcy8
-itpr1	->	+Adora2b
-Gna14	->	-Bst1
-itpr1	->	-Cckbr
-adcy8	->	-Gna14
-gna14	->	-Grin2c
-grin2c	->	-Itpr1
-gna14	->	+P2rx2
+bst1	->	+Pdgfrb
-itpr1	->	-Slc8a3
-pdgfrb	->	-Sphk1
+Kitl	->	-Ccl5
-Kitl	->	-Csf1
-Kitl	->	-Cx3cl1
-Kitl	->	-Il11ra1
-Cx3cl1	->	-Il15
-Csf1	->	-Kitl
-Cx3cl1	->	-Pdgfrb
-kitl	->	-Tnfrsf11b
-kitl	->	-Tnfsf9
-gabra6	->	+Adora2b
- adora2b	->	-Agtrl1
-gabra6	->	-Cckbr
-Gria1	->	-Edg2
-gabra6	->	-Gabra1
-cckbr	->	-Gabra6
-gabra6	->	-Gabrb1
-gabra6	->	-Gabrd

-col5a2	->	-Lamc1
-col5a2	->	-Parva
-col5a2	->	-Pdgfrb
-col5a2	->	-Thbs2
-col5a2	->	-Col1a1
+col3a1	->	+Col1a2
-col5a1	->	-Col3a1
-Col3a1	->	-Col5a1
-col3a1	->	-Col5a2
+col6a2	->	+Col5a3
-col3a1	->	-Col6a2
+Col5a2	->	+Fn1
-col3a1	->	-Itga1
-col3a1	->	-Lama5
-col3a1	->	-Lamb2
-col3a1	->	-Lamc1
-col3a1	->	-Sdc2
-fn1	->	-Thbs2
-glycam1	->	-Cdh1
-glycam1	->	-Cldn11
-Icam1	->	-Glycam1
+sdc2	->	+Icam1
+Icam1	->	-Mag
-spn	->	-Ocln
-spn	->	-Sdc2
-ocln	->	-Spn
-Glycam1	->	-Vcam1
-rab13	->	-Cldn11
+rab3b	->	+Myh14
+myh14	->	-Ocln
-myh14	->	-Rab3b
+cldn11	->	+Rab13
+gja1	->	+Adcy8
-npr2	->	-Gja1
+pdgfrb	->	-Itpr1
-pdgfrb	->	-Npr2
+npr2	->	+Pdgfrb
+plau	->	-A2m
+plau	->	-C1r
+plau	->	-C1s
+plau	->	-Cfh
+Kng1	->	+F3

-Fgf2	->	-Itga1
-Fgfr1	->	-Myh14
+Fgfr1	->	+Pdgfrb
-Fgfr2	->	-Ssh3
+mmp2	->	+Adcy8
-itpr1	->	+Hbegf
+hbegf	->	-Itpr1
-mmp2	->	+Lhb
+adcy8	->	+Mmp2
-tyrp1	->	-Adcy8
+tyrp1	->	+Kitl
-kitl	->	-Tyrp1
-cpt1b	->	-Acsl6
+Acsl6	->	+Cpt1b
+cpt1b	->	-Cpt2
+cpt1b	->	-Npy
-igf1	->	-Ccnd1
+ccnd1	->	+Igf1
-igf1	->	-Pdgfrb
+pdgfrb	->	+Ccnd1
-pdgfrb	->	-Fgfr1
-pdgfrb	->	-Fgfr2
+pdgfrb	->	+Igf1
+ret	->	+Ccnd1
-ret	->	-Cdh1
+cdh1	->	+Ret
+Fgf2	->	+Ccnd1
-Fgf2	->	-Cdh1
-igf1	->	-Fgf2
+cdh1	->	+Fgfr1
+cdh1	->	+Igf1
+fgf2	->	+Pdgfrb
+mmp2	->	+Ccnd1
-mmp2	->	-Cdh1
-cdh1	->	-Mmp2
+lama5	->	+Ccnd1
-lama5	->	-Fn1
+lamb2	->	+Lama5
+fn1	->	+Lamb2
-Lama5	->	-Lamc1

Appendix II

Pathway wise list of the module and their genes

Module	Pathway	No. of Genes	Genes
MMP2	Adherens junction	1	Cdh1
MMP2	Bladder cancer	2	Cdh1, Mmp2
MMP2	Calcium signaling pathway	4	Adcy8, Gna14, Grin2c, P2rx2
MMP2	Cell adhesion molecules (CAMs)	3	Cdh1, Cldn11, Glycam1
MMP2	Endometrial cancer	1	Cdh1
MMP2	Gap junction	2	Adcy8, Edg2
MMP2	GnRH signaling pathway	3	Adcy8, Lhb, Mmp2
MMP2	Leukocyte transendothelial migration	2	Cldn11
MMP2	Long-term depression	1	Mmp2
MMP2	Long-term potentiation	3	Adcy8, Gria1, Grin2c
MMP2	Melanogenesis	1	Adcy8
MMP2	Melanoma	1	Cdh1
MMP2	Neuroactive ligand-receptor interaction	6	Edg2, Gria1, Grin2c, Lhb, P2rx2, Uts2r
MMP2	Purine metabolism	1	Adcy8
MMP2	Taste transduction	1	Adcy8
MMP2	Thyroid cancer	2	Cdh1, Ret
MMP2	Tight junction	2	Cldn11, Rab13
PDGFRB	Acute myeloid leukemia	1	Ccnd1
PDGFRB	Adherens junction	1	Fgfr1
PDGFRB	Antigen processing and presentation	1	Hspa1a
PDGFRB	Axon guidance	5	Ablim2, Ephb6, Nfatc4, Slit2, Slit3
PDGFRB	B cell receptor signaling pathway	1	Nfatc4
PDGFRB	Bladder cancer	1	Ccnd1
PDGFRB	Calcium signaling pathway	5	Bst1, Itpr1, Pdgfrb, Slc8a3, Sphk1
PDGFRB	Cell cycle	1	Ccnd1
PDGFRB	Chronic myeloid leukemia	1	Ccnd1
PDGFRB	Colorectal cancer	2	Ccnd1, Pdgfrb
PDGFRB	Cytokine-cytokine receptor interaction	1	Pdgfrb
PDGFRB	Endometrial cancer	1	Ccnd1
PDGFRB	ErbB signaling pathway	1	Hbegf
PDGFRB	Focal adhesion	3	Ccnd1, Igf1, Pdgfrb
PDGFRB	Gap junction	3	Itpr1, Npr2, Pdgfrb
PDGFRB	Glioma	3	Ccnd1, Igf1, Pdgfrb
PDGFRB	GnRH signaling pathway	2	Hbegf, Itpr1

PDGFRB	Jak-STAT signaling pathway	1	Ccnd1
PDGFRB	Long-term depression	3	Igf1, Itpr1, Npr2
PDGFRB	Long-term potentiation	1	Itpr1
PDGFRB	MAPK signaling pathway	7	Fgf2, Fgfr1, Fgfr2, Hspa1a, Nfatc4, Pdgfrb, Ptpn5
PDGFRB	Melanoma	5	Ccnd1, Fgf2, Fgfr1, Igf1, Pdgfrb
PDGFRB	mTOR signaling pathway	1	Igf1
PDGFRB	Natural killer cell mediated cytotoxicity	1	Nfatc4
PDGFRB	Nicotinate and nicotinamide metabolism	1	Bst1
PDGFRB	Non-small cell lung cancer	1	Ccnd1
PDGFRB	p53 signaling pathway	3	Ccnd1, Igf1, Igfbp3
PDGFRB	Pancreatic cancer	1	Ccnd1
PDGFRB	Phosphatidylinositol signaling system	1	Itpr1
PDGFRB	Prostate cancer	5	Ccnd1, Fgfr1, Fgfr2, Igf1, Pdgfrb
PDGFRB	Purine metabolism	3	Entpd1, Gda, Npr2
PDGFRB	Pyrimidine metabolism	1	Entpd1
PDGFRB	Regulation of actin cytoskeleton	6	Fgf2, Fgfr1, Fgfr2, Gsn, Pdgfrb, Ssh3
PDGFRB	Small cell lung cancer	1	Ccnd1
PDGFRB	Sphingolipid metabolism	3	Asah1, Smpd2, Sphk1
PDGFRB	T cell receptor signaling pathway	1	Nfatc4
PDGFRB	Thyroid cancer	1	Ccnd1
PDGFRB	VEGF signaling pathway	2	Nfatc4, Sphk1
PDGFRB	Wnt signaling pathway	4	Ccnd1, Nfatc4, Sfrp4, Wif1
GJA1	Cell Communication	17	Col1a1, Col1a2, Col3a1, Col5a1, Col5a2, Col5a3, Col6a2, Fn1, Gja1, Gja5, Gjb1, Gjb2, Lama5, Lamb2, Lamc1, Thbs2, Vim
GJA1	ECM-receptor interaction	12	Col1a1, Col1a2, Col3a1, Col5a1, Col5a2, Col5a3, Col6a2, Fn1, Itga1, Lama5, Lamb2, Lamc1, Thbs2
GJA1	Focal adhesion	15	Cav3, Col1a1, Col1a2, Col3a1, Col5a1, Col5a2, Col5a3, Col6a2, Fn1, Itga1, Lama5, Lamb2, Lamc1, Parva, Thbs2
GJA1	Gap junction	1	Gja1
GJA1	Hematopoietic cell lineage	1	Itga1
GJA1	Prion disease	1	Lamc1
GJA1	Regulation of actin cytoskeleton	2	Fn1, Itga1
GJA1	Small cell lung cancer	4	Fn1, Lama5, Lamb2, Lamc1
GJA1	TGF-beta signaling pathway	2	Dcn, Thbs2

KITL	Acute myeloid leukemia	1	Stat5a
KITL	Chronic myeloid leukemia	1	Stat5a
KITL	Cytokine-cytokine receptor interaction	8	Ccl5, Csf1, Cx3cl1, Il11ra1, Il15, Kitl, Tnfrsf11b, Tnfsf9
KITL	ErbB signaling pathway	1	Stat5a
KITL	Hematopoietic cell lineage	4	Cd1d1, Csf1, Il11ra1, Kitl
KITL	Jak-STAT signaling pathway	3	Il11ra1, Il15, Stat5a
KITL	Melanogenesis	2	Kitl, Tyrp1
KITL	Toll-like receptor signaling pathway	1	Ccl5
KITL	Tyrosine metabolism	1	Tyrp1
GABRA6	Calcium signaling pathway	2	Adora2b, Cckbr
GABRA6	Neuroactive ligand-receptor interaction	10	Adora2b, Agtrl1, Cckbr, Gabra1, Gabra6, Gabrb1, Gabrd, Grm3, Grm4, Grm8
GABRA6	Taste transduction	1	Grm4
ICAM1	Cell adhesion molecules (CAMs)	6	Icam1, Mag, Ocln, Sdc2, Spn, Vcam1
ICAM1	ECM-receptor interaction	1	Sdc2
ICAM1	Fc epsilon RI signaling pathway	1	Lat
ICAM1	Leukocyte transendothelial migration	4	Icam1, Ocln, Thy1, Vcam1
ICAM1	Natural killer cell mediated cytotoxicity	3	Cd48, Icam1, Lat
ICAM1	T cell receptor signaling pathway	1	Lat
ICAM1	Tight junction	1	Ocln
PLAU	Alzheimer's disease	1	A2m
PLAU	Complement and coagulation cascades	10	A2m, C1r, C1s, Cfh, F3, F5, Kng1, Masp1, Plau, Serping1
ACSL6	Adipocytokine signaling pathway	4	Acsl6, Cpt1b, Cpt2, Npy
ACSL6	Fatty acid metabolism	3	Acsl6, Cpt1b, Cpt2
ACSL6	PPAR signaling pathway	4	Acsl6, Angptl4, Cpt1b, Cpt2
TAP2	ABC transporters - General	4	Abcb9, Abcc8, Abcg3, Tap2
TAP2	Antigen processing and presentation	2	Cd74, Tap2
TAP2	Type II diabetes mellitus	1	Abcc8
MYH14	Regulation of actin cytoskeleton	1	Myh14
MYH14	Tight junction	2	Myh14, Rab3b
BMP4	Basal cell carcinoma	1	Bmp4
BMP4	Hedgehog signaling pathway	2	Bmp4, Bmp6
BMP4	TGF-beta signaling pathway	2	Bmp4, Bmp6
NMNAT3	Nicotinate and nicotinamide metabolism	2	Nmnat3, Nnt
AGA	Aminosugars metabolism	1	Hexa
AGA	Glycan structures - degradation	3	Aga, Fuca2, Hexa

AGA	Glycosaminoglycan degradation	1	Hexa
AGA	Glycosphingolipid biosynthesis - ganglioseries	1	Hexa
AGA	Glycosphingolipid biosynthesis - globoseries	1	Hexa
AGA	N-Glycan degradation	3	Aga, Fuca2, Hexa
DHRS3	1- and 2-Methylnaphthalene degradation	1	Dhrs3
DHRS3	Benzoate degradation via CoA ligation	1	Dhrs3
DHRS3	Bile acid biosynthesis	1	Hsd3b7
DHRS3	Bisphenol A degradation	3	Dhrs3, Hsd3b7, Pon2
DHRS3	Butanoate metabolism	1	Hsd3b7
DHRS3	Fructose and mannose metabolism	1	Hsd3b7
DHRS3	Galactose metabolism	1	Hsd3b7
DHRS3	gamma-Hexachlorocyclohexane degradation	2	Dhrs3, Pon2
DHRS3	Glycine, serine and threonine metabolism	1	Hsd3b7
DHRS3	Limonene and pinene degradation	1	Dhrs3
DHRS3	Linoleic acid metabolism	1	Hsd3b7
DHRS3	Naphthalene and anthracene degradation	1	Dhrs3
DHRS3	Tetrachloroethene degradation	1	Hsd3b7
LBP	Toll-like receptor signaling pathway	2	Lbp, Tlr3

Appendix III

Pathway analysis usingby ToppGene server

Name	Source	p-value	No. of Genes
Metabolism of lipids and lipoproteins	BioSystems: REACTOME	2.68E-02	16
Pathways in cancer	BioSystems: KEGG	1.82E-02	11
PDGF signaling pathway	PantherDB	1.10E-03	8
Rap1 signaling pathway	BioSystems: KEGG	2.35E-02	8
Regulation of actin cytoskeleton	BioSystems: KEGG	2.46E-02	8
Alzheimer disease-presenilin pathway	PantherDB	1.04E-02	6
Angiogenesis	PantherDB	3.62E-02	6
Neural Crest Differentiation	BioSystems: WikiPathways	6.12E-03	6
WNT ligand biogenesis and trafficking	BioSystems: REACTOME	3.75E-02	5
Aminoacyl-tRNA biosynthesis	BioSystems: KEGG	2.26E-02	4
Bacterial invasion of epithelial cells	BioSystems: KEGG	3.55E-02	4
Basal cell carcinoma	BioSystems: KEGG	1.22E-02	4
Cytosolic tRNA aminoacylation	BioSystems: REACTOME	5.64E-04	4
Phosphatidylinositol signaling system	BioSystems: KEGG	4.32E-02	4
Prolactin signaling pathway	BioSystems: KEGG	2.99E-02	4
Regulation of actin dynamics for phagocytic cup formation	BioSystems: REACTOME	4.32E-02	4
tRNA Aminoacylation	BioSystems: REACTOME	4.73E-03	4
3-phosphoinositide biosynthesis	BioSystems: BIOCYC	9.37E-03	3
Aldosterone-regulated sodium reabsorption	BioSystems: KEGG	2.54E-02	3
Benazepril Pathway	SMPDB	1.76E-02	3
Captopril Pathway	SMPDB	1.76E-02	3
Carbohydrate digestion and absorption	BioSystems: KEGG	3.68E-02	3
CD40/CD40L signaling	BioSystems: Pathway Interaction Database	1.26E-02	3
Chaperonin-mediated protein folding	BioSystems: REACTOME	4.79E-02	3
Cilazapril Pathway	SMPDB	1.76E-02	3
Enalapril Pathway	SMPDB	1.76E-02	3
Fosinopril Pathway	SMPDB	1.76E-02	3
HIF-2-alpha transcription factor network	BioSystems: Pathway Interaction Database	1.76E-02	3
Interleukin-2 signaling	BioSystems: REACTOME	3.27E-02	3
Interleukin-3, 5 and GM-CSF signaling	BioSystems: REACTOME	4.56E-02	3
Lisinopril Pathway	SMPDB	1.76E-02	3
MAP00310 Lysine degradation	GenMAPP	2.05E-03	3

MAP00970 Aminoacyl tRNA biosynthesis	GenMAPP	2.91E-03	3
Moexipril Pathway	SMPDB	1.76E-02	3
Notch Signaling Pathway	BioSystems: WikiPathways	3.89E-02	3
Notch signaling pathway	BioSystems: KEGG	4.33E-02	3
Perindopril Pathway	SMPDB	1.76E-02	3
Quinapril Pathway	SMPDB	1.76E-02	3
Ramipril Pathway	SMPDB	1.76E-02	3
Rescinnamine Pathway	SMPDB	1.76E-02	3
Spirapril Pathway	SMPDB	1.76E-02	3
Trandolapril Pathway	SMPDB	1.76E-02	3
Type II interferon signaling (IFNG)	BioSystems: WikiPathways	4.79E-02	3
Wnt signaling	Pathway Ontology	1.63E-02	3
Amlodipine Pathway	SMPDB	3.41E-02	2
Arachidonic Acid Metabolism	SMPDB	4.15E-02	2
CD28 dependent PI3K/Akt signaling	BioSystems: REACTOME	4.54E-02	2
Crosslinking of collagen fibrils	BioSystems: REACTOME	8.84E-03	2
Felodipine Pathway	SMPDB	3.41E-02	2
Gap junction assembly	BioSystems: REACTOME	3.78E-02	2
Glycosaminoglycan biosynthesis - chondroitin sulfate / dermatan sulfate	BioSystems: KEGG	4.15E-02	2
Interleukin-7 signaling	BioSystems: REACTOME	1.57E-02	2
Isradipine Pathway	SMPDB	3.41E-02	2
JAK/STAT signaling pathway	PantherDB	2.74E-02	2
Lysine degradation, lysine => saccharopine => acetoacetyl-CoA	BioSystems: KEGG	1.32E-02	2
Nifedipine Pathway	SMPDB	3.41E-02	2
Nimodipine Pathway	SMPDB	3.41E-02	2
Nisoldipine Pathway	SMPDB	3.41E-02	2
Nitrendipine Pathway	SMPDB	3.41E-02	2
NOTCH2 Activation and Transmission of Signal to the Nucleus	BioSystems: REACTOME	4.94E-02	2
Rap1 signalling	BioSystems: REACTOME	2.74E-02	2
Signaling by NOTCH3	BioSystems: REACTOME	1.32E-02	2
Signaling by NOTCH4	BioSystems: REACTOME	1.32E-02	2
superpathway of melatonin degradation	BioSystems: BIOCYC	1.57E-02	2
2-oxoglutarate decarboxylation to succinyl-CoA	BioSystems: BIOCYC	4.82E-02	1
Beta oxidation of myristoyl-CoA to lauroyl-CoA	BioSystems: REACTOME	4.82E-02	1
Beta oxidation of palmitoyl-CoA to myristoyl-CoA	BioSystems: REACTOME	4.82E-02	1
Conjugation of phenylacetate with glutamine	BioSystems: REACTOME	3.24E-02	1
Desmosterolosis	SMPDB	1.63E-02	1

diphthamide biosynthesis	BioSystems: BIOCYC	3.24E-02	1
Dopamine clearance from the synaptic cleft	BioSystems: REACTOME	4.82E-02	1
dTMP de novo biosynthesis (mitochondrial)	BioSystems: BIOCYC	4.82E-02	1
Enzymatic degradation of dopamine by COMT	BioSystems: REACTOME	3.24E-02	1
Enzymatic degradation of Dopamine by monoamine oxidase	BioSystems: REACTOME	3.24E-02	1
Fibronectin matrix formation	BioSystems: REACTOME	4.82E-02	1
Glycine biosynthesis I	BioSystems: BIOCYC	3.24E-02	1
Glycine metabolic	Pathway Ontology	3.24E-02	1
Glycine Metabolism	BioSystems: WikiPathways	3.24E-02	1
melatonin degradation II	BioSystems: BIOCYC	1.63E-02	1
Metabolism of serotonin	BioSystems: REACTOME	3.24E-02	1
Monoamines are oxidized to aldehydes by MAOA and MAOB, producing NH3 and H2O2	BioSystems: REACTOME	3.24E-02	1
Oxygen-dependent Asparagine Hydroxylation of Hypoxia-inducible Factor Alpha	BioSystems: REACTOME	4.82E-02	1
Phenylacetate Metabolism	SMPDB	4.82E-02	1
Progressive trimming of alpha-1,2-linked mannose residues from Man9/8/7GlcNAc2 to produce Man5GlcNAc2	BioSystems: REACTOME	4.82E-02	1
Regulation of thyroid hormone activity	BioSystems: REACTOME	4.82E-02	1
Serine and glycine biosynthesis	BioSystems: BIOCYC	4.82E-02	1
Serotonin clearance from the synaptic cleft	BioSystems: REACTOME	3.24E-02	1
Signaling by FGFR2 amplification mutants	BioSystems: REACTOME	1.63E-02	1
thyroid hormone metabolism I (via deiodination)	BioSystems: BIOCYC	3.24E-02	1
thyronamine and iodothyronamine metabolism	BioSystems: BIOCYC	3.24E-02	1