Network Analysis of Prion Disease

By

KHALIQUE NEWAZ

Department of Computer science
INDRAPRASTHA INSTITUTE OF INFORMATION TECHNOLOGY

A dissertation submitted to the Indraprastha Institute of Information Technology in accordance with the requirements of the degree of MASTER OF TECHNOLOGY in the Faculty of Computer science.

MAY 2015

ABSTRACT

Prion diseases are transmissible neuro-degenerative diseases that arises due to conformational change of normal cellular prion protein (PrP^{c}) to protease-resistant isofrom ($rPrP^{Sc}$). Deposition of misfolded PrP^{Sc} proteins leads to an alteration of many signaling pathways that includes immunological and apoptotic pathways and as a result, culminates in dysfunction and death of neuronal cells. Transcriptional studies have revealed some affected pathways, but it is not clear which is (are) the prime network pathway(s) that change during the disease progression and how the prime pathways are involved in cross talk with each other from the time of incubation to clinical death. To address these issues we performed network analysis on large-scale transcriptional data of differentially expressed genes obtained from whole brain in 6 different mouse prion models. We employed the notion of differential network centrality measurement to identify the potential pathways involved in disease progression. We also used cross-talk ranking method to identify core network elements involved in the cross-talk with different pathways. We identified 148 DEGs (differentially expressed genes) potentially related to the prion disease progression. Functional association of the identified genes implicated strong involvement of immunological pathways. A bow-tie structure was extracted, suggesting that the effects of prion disease on the core elements (PI3Ks and AKT) of the structure leads to potential dysfunction of the biological pathways involved. We also modeled the bow-tie network that can be used to approximate the behavior of this network during prion disease prevalence. In this study we showed using transcriptional data that neuronal dysfunction during prion disease is strongly related to immune response. We conclude that the immunological pathways occupy the network's influential positions in the protein functional networks related to prion disease and this network-central involvement is prevalent in 5 different mouse strain-prion strain combinations.

ACKNOWLEDGEMENTS

I want to thank my advisor Dr. Debajyoti Bera, an open minded professor who supported my desire to go ahead with my project in my own way and who also trusted me with my results. I also want to express my gratitude towards Dr. K. Sriram for giving me the opportunity of exploring the interesting world of computational biology, bioinformatics and network science. I'm grateful to both of them for our interesting discussions about the problem I studied and also for their active participation. I'm thankful to my parents, who supported me emotionally and financially. Finally, I would like to thank my mother Mumtaz Begum and my friend Ekta Gupta for encouraging me in all of my pursuits and inspiring me to follow my dreams.

CERTIFICATE

This is to certify that the thesis titled Network analysis of prion disease being submitted by Khalique Newaz to Indraprastha Institute of Information Technology Delhi, for the award of Master of Technology, is an original research work carried out by him under my supervision. In my opinion, the thesis has reached the standards fulfilling the requirements of the regulations relating to the degree. The results contained in this thesis have not been submitted in part or full to any other university or institute for the award of any degree/diploma.

MAY, 2015

Dr. Debajyoti Bera

Department of Computer science
Indraprastha Institute of Information Technology
New Delhi 110020

TABLE OF CONTENTS

LIST OF TABLES

TABLE **Page**

LIST OF FIGURES

FIGURE **Page**

CHAPTER 1

INTRODUCTION

Prion diseases are transmissible neuro-degenerative diseases that arises due to conformational change of normal cellular prion protein (PrP^c) to protease-resistant isofrom ($rPrP^{Sc}$). Deposition of misfolded Prp^{Sc} proteins leads to an alteration of many signaling pathways that includes immunological and apoptotic pathways and as a result, culminates in dysfunction and death of neuronal cells. Transcriptional studies have revealed some affected pathways, but it is not clear which is (are) the prime network pathway(s) that change during the disease progression and how the prime pathways are involved in cross talk with each other from the time of incubation to clinical death. This thesis focus on the use of protein interaction networks to identify the prime pathological pathways involved in the prion disease progression.

1.1 Thesis structure

In this Chapter, we introduce the background and common terms and concepts used throughout this thesis. First, we explain the general methodology adopted in this work to explore the problem. In the subsequent section, a basic introduction about the concepts of micro-array analysis, protein-protein interaction networks and prion disease will be given. We then introduce concepts from graph theory and complex network analysis.

Chapter 2 summarizes related work and prior research on prion disease and related network analysis. We then explain the dataset used in this work.

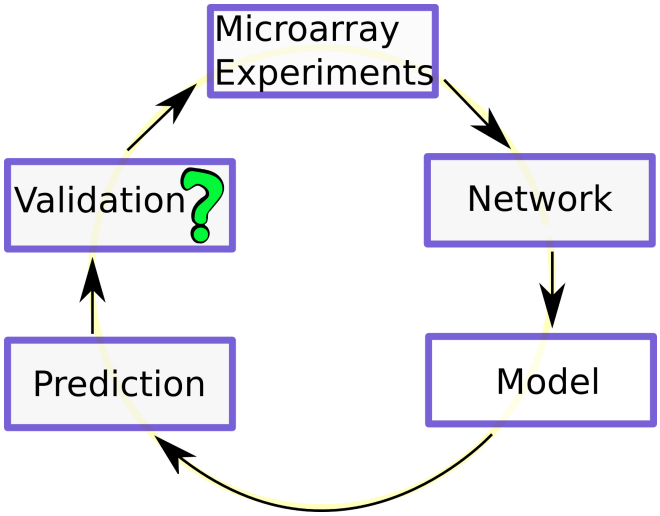

FIGURE 1.1. The figure outlines the basic methodology employed in this thesis to study prion disease

In Chapter 3 we explain our analysis pipeline, and the algorithms that we developed and implemented to identify the network central proteins related to prion disease.

Next, In chapter 4 we explain the concepts of network robustness and cross-talk between biological pathways. We then present an ODE model for capturing the dynamics of the bow-tie structure prevalent in prion disease.

Finally, in Chapter 5 we summarize our work, draw conclusions and outline possible future work.

1.2 Overview of methodology

Figure 1.1 outlines the basic procedure involved to systematically study prion disease and draw conclusions. The basic methodology employed in this work starts with the microarray experiments, which studies the differential expression of genes in a particular biological condition under study. The condition which is being investigated in this study is the onset and progress of prion disease. The results of the microarray experiments are used to identify the networks prevalent in the diseased condition. These networks can be protein interaction networks, gene regulatory networks, or metabolic networks, depending upon the structure and scale of the microarray experiments performed. In this work we have chosen to work with the protein interaction networks. The identified network is then mathematically modeled to predict the behavior of the system under the influence of the condition being studied.

The predicted behavior is generally validated by other experiments including microarray, which is then used further to modify or fine tune the model and its predictions. This work is confined to identification of the network and its mathematical modeling. Validation and refining of the mathematical model proposed is presented as a future work.

1.3 Biological background

This section provides the information about the biological terms and details needed to follow the work in this thesis. In section 1.3.1 we discuss about the prion disease and why this disease is interesting to study. Then we outline the basic concepts of microarray technology and interpretation of their outputs. Next in section 1.3.3, we define the term protein and discuss how proteins relate to DNA, mRNA, and genes. Then in section 1.3.4 , we explain the concept of protein-protein interactions and give a few examples. Furthermore, the concepts of biological pathways are elaborated in Section 1.3.5, and finally, we explain how these pathways can be involved in a cross-talk.

1.3.1 Prion disease

Prion forms of proteins are the potential disease causing agents in a group of fatal neurodegenerative diseases which affect a diverse group of species, including humans. In prion diseases, prions replicate by conversion of cellular protein (PrP^{C}) to disease specific PrP^{Sc} isoforms. Accumulation of misfolded PrP^{Sc} proteins, rich in β-sheet structure, has been hypothesized to trigger a series of events which leads to neuronal dysfunction [19]. Some of the other neurodegenerative diseases such as Alzheimer 's or parkinson's diseases, are also the result of protein misfolding, and are pathologically characterized by abnormal protein deposition and formation of plaques[5, 19]. The noval pathological mechanisms related to prion disease may be relevant to other such neurological illness [13]. The pathological similarity of these wide range of neurodegenerative diseases suggest that a thorough understanding of the molecular mechanisms related to any of these diseases may prove to be of therapeutic importance to several other diseases.

Different strains of infectious prions, presumably arising from distinct structural forms of misfolded prion proteins, has different effects on the disease progression (e.g. duration of incubation time, affected sites in the brain, etc.). This variation in the prion strains and their interaction with different host genotypes leads to a variation in the pathogenesis of the

prion disease. This poses a problem of apprehending the actual core mechanisms involved in the disease progression.

Neuropathological features common to all prion diseases in mammals can be divided into following four processes: prion replication (PrP^{Sc}) and accumulation [36], synaptic degeneration [29], microglia and astrocyte activation, and neuronal death. Several hypotheses have been put forth regarding pathogenesis of prion diseases leading to neurodegeneration. Mechanisms such as synaptic degeneration, endoplasmic reticulum stress, vacoulization, microglial activation, accumulation of misfolded proteins, oxidative stress, and apoptosis have been proposed. But none of these hypotheses have been able to explain the prion pathogenesis completely [42]. Microglia are one of the earliest responder in prion disease [37] but their role in the prion disease initiation and progression is still not clear. They are the key components of the brain immune system and are responsible for active immune response in the brain. Activated microglia may be a major contributor to neuronal death in prion diseases [35]. Pro-inflammatory responses, mainly related to microglia, has been found to be a common feature in other neurodegenerative diseases such as Alzheimer's and parkinson's diseases [4, 11].

Thus, hyper-activation of microglia has been hypothesized and supported by earlier works, as one of the key players participating in the neuronal degradation and death in the onset of prion disease. This work supports the involvement of immunological functions in the prion disease progression and also identifies how the key immunological pathways interact in the disease prevalence.

1.3.2 DNA, mRNA, Proteins

DNA stands for deoxyribonucleic acid. The primary function of DNA is to store and transmit the genetic information. This genetic information is utilized by the cells to decide which protein to make and when to make them. Figure 1.2 shows the general biological sequential information transfer. DNA is a double stranded thread like structure in which each strand consists of a sequence of code consisting of four types of nucleotides: adenine(A), cytosine(C), guanine(G), and thymine(T). The bases of one strand interact with the bases of the other strand according to a set of pairing rules, such that A pairs with T and C pairs with G. Thus, if one knows the sequence of one strand, by definition, one then knows the sequence of the opposite strand. This property of complementarity is exploited for measuring gene expression on microarrays. A sub-sequence of 3 nucleotides is called a codon. Each codon translates to single amino acid (building blocks of proteins). And a group of codons makes up a gene. A particular sequence of these codons on the DNA strands decides the type of

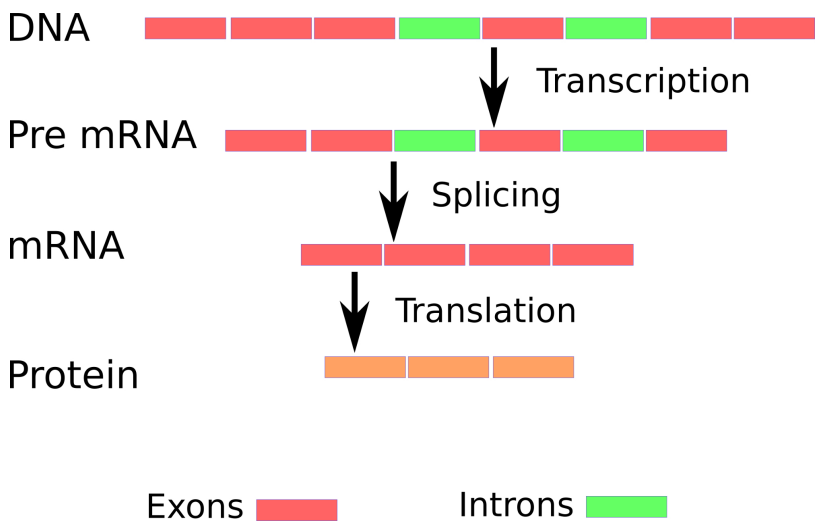

FIGURE 1.2. DNA transcripts the information to pre-mRNA by the process called transcription. The pre-mRNA then undergoes splicing in which non-coding regions (Introns) are separated from the coding regions. Finally, the mRNA is translated to protein. Exons are the coding regions which has the information about the protein to be translated.

gene to be transcribed to tRNA(pre-mRNA). The transcribed tRNA is converted to mRNA and finally mRNA is translated to a protein. Each mRNA corresponds to a particular gene and hence the concentration of a particular mRNA is proportional to the corresponding amount of gene being transcribed in this process. This fact is utilized in the transcriptomic studies, in which the relative concentration of mRNA is used to characterize a gene to be up or down-regulated corresponding to a particular condition under study.

The final product of this sequence of information transfer are proteins. Proteins are large macro-molecules which can perform a variety of biological and chemical functions. They are the core machinery of all living cells and are responsible for signaling, metabolism, the inner and outer structure of cells, transporting other proteins and substances throughout the cell, and many other tasks. All proteins consist of the same basic building blocks called amino acids, only 20 different amino acids make up all proteins[3].

The protein structure can be hierarchically organized from so-called primary structures to quaternary structures. The wide variety of conformations of proteins is due to the huge amount of different sequences of amino acids. The primary structure is the sequence of amino acids in the polypedptide chain. Secondary structure is a local regularly occurring structure in proteins and is mainly formed through hydrogen bonds between backbone

atoms. There are two types of stable secondary structures: alpha-helices and beta-sheets. In prion disease, the secondary structure of the cellular prion protein (PrP^{C}) changes its configuration from alpha-helices to beta-sheets to form disease producing prion protein (PrP^{Sc}). Tertiary structure describes the packing of alpha-helices and beta-sheets with respect to each other on the level of one whole polypeptide chain. Quaternary structure only exists, if there is more than one polypeptide chain present in a complex protein. Then quaternary structure describes the spatial organization of the chains.

1.3.3 Microarray technology

Microarray technology has provided a valuable means of inspecting bio-molecular behavior under various conditions. The purpose of a microarray is to examine expression of multiple genes simultaneously in response to some biological perturbation. Microarrays can be used to measure the gene expression in a number ways. The most common way to measure is to compare the gene expressions in the normal/healthy organism and the gene expressions of the organisms in the diseased condition (Figure 1.3). First, mRNA is extracted from the cells. Next, mRNA molecules in the extract are reverse transcribed into cDNA by using an enzyme reverse transcriptase and nucleotides labeled with different fluorescent dyes. For example, cDNA from cells grown in condition A may be labeled with a red dye and from cells grown in condition B with a green dye. Once the samples have been differentially labeled, they are allowed to hybridize (hybridization is the process of combining the two complementary strands of DNA) onto the same glass slide. At this point, any cDNA sequence in the sample will hybridize to specific spots on the glass slide containing its complementary sequence. The amount of cDNA bound to a spot will be directly proportional to the initial number of mRNA molecules present for that gene in both the samples.

Following the hybridization step, the spots in the hybridized microarray are excited by a laser and scanned at suitable wavelengths to detect the red and green dyes. The amount of fluorescence emitted upon excitation corresponds to the amount of bound nucleic acid. For instance, if cDNA from condition A for a particular gene was in greater abundance than that from condition B, one would find the spot to be red. If it was the other way, the spot would be green. If the gene was expressed to the same extent in both conditions, one would find the spot to be yellow, and if the gene was not expressed in both conditions, the spot would be black. Thus, what is seen at the end of the experimental stage is an image of the microarray, in which each spot that corresponds to a gene has an associated fluorescence value representing the relative expression level of that gene.

The relative expression level for a gene can be measured as the amount of red or green

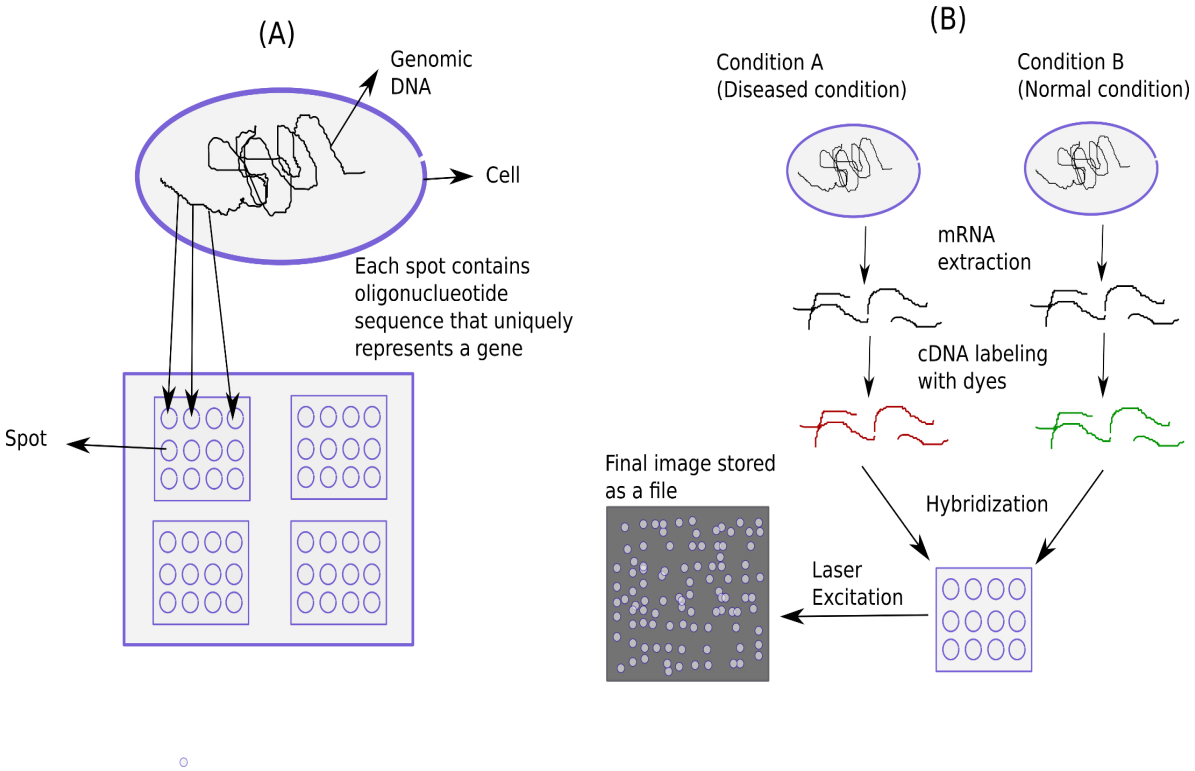

FIGURE 1.3. (A) A microarray may contain thousands of [a]Spots [o]. Each spot contains many copies of the same DNA sequence that uniquely represents a gene from an organism. (B) Schematic of the experimental protocol to study differential expression of genes. The organism is grown in two different conditions (a reference condition and a test condition). RNA is extracted from the two cells, and is labeled with different dyes (red and green) during the synthesis of cDNA by reverse transcriptase. Following this step, cDNA is hybridized onto the microarray slide, where each cDNA molecule representing a gene will bind to the spot containing its complementary DNA sequence. The microarray slide is then excited with a laser at suitable wavelengths to detect the red and green dyes. The final image is stored as a file for further analysis.

light emitted after excitation. One of the common metrics used to relate this information is called expression ratio. It is denoted here as T_k and defined as:

$$T_k = \frac{R_k}{G_k}$$

For each gene k on the array, where R_k represents the spot intensity metric for the diseased

condition and $G_k$represents the spot intensity metric for the normal condition. The expression ratio is a relevant way of representing expression differences in a very intuitive manner. But this representation may not be helpful in case of distinguishing between up and down regulation. A better transformation procedure is to take the logarithm base 2 value of the expression ratio (i.e. log2 (expression ratio)). It treats differential up-regulation and down-regulation equally, and also has a continuous mapping space. For example, if the expression ratio is 1, then log2 (1) equals 0 represents no change in expression. If the expression ratio is 4, then log2 (4) equals +2 and for expression ratio of log2 (1/4) equals -2. Thus, in this transformation the mapping space is continuous and up-regulation and down-regulation are comparable.

In this work we have utilized the large-scale transcriptomic study related to prion disease. The differentially expressed genes in the prion disease condition has been utilized to construct the protein functional networks(PFNs) and these networks were then subsequently used for the identification of the prime pathological pathways.

1.3.4 Protein interaction networks

In order to fulfill their function, proteins interact with other substances (molecules, ions, DNA, etc.) or other proteins. Proteins interact in numerous different contexts and with different outcomes. Some proteins activate or deactivate other proteins by binding to them or by (de)phosphorylating them. In the process of (de)phosphorylation, a phosphate group is (removed)added from a protein, which activates or deactivates the protein. Some proteins bind to each other, creating so-called protein complexes. These have important roles in the entire cell. Another class of proteins bind to each other to create structural complexes which give the cell its 3-dimensional structure. Yet other proteins pass on signals by interacting with source and destination proteins in so-called signaling pathways. Transcription factors are proteins that bind to DNA to activate the transcription process of a gene. This activation often requires multiple transcription factors to interact and also bind to the DNA. Thus, the elucidation of protein interactions is a central problem in biology. Unless we understand the complex interaction patterns of the tens of thousands of proteins that constitute our proteome, we cannot hope to attempt to efficiently combat some of the most important diseases.

A number of different approaches have been in use towards reconstructing the interactions between proteins. The literature comprises studies that use high throughput experiments to find if there exist pairwise interactions between a large set of query proteins. Other studies use computational modeling to determine which proteins may bind to each other

based on their (predicted) structural properties. Some of the most popular and widely used experimental and computational techniques are explained below:

1. Yeast two hybrid [18]: It is the genetic method that uses the transcriptional activity as a measure of protein-protein interaction. Two hybrid proteins are created corresponding to the proteins A and B between which the interaction is to be identified. One protein(say A) is bind with the DNA binding domain. Other protein(B) is fused with the transcription activation domain. These two hybrids are expressed in a cell containing reporter genes. The positive expression of the reporter genes identifies the interaction between the proteins A and B.

2. Co-localization: These methods work on the hypothesis that the genes which physically interact should be present in physically close proximity in the genome.

3. Co-occurrence: This method exploits the co-occurrence of homologous pairs of genes across multiple genomes. The fact that a pair of genes express together across many different species suggest that these genes are functional associated or physically interacting.

As a result of the high throughput experiments and computational techniques, many databases have been designed and setup to store the protein-protein interaction data. These databases usually are the results of integration of diverse data-sets. Some of the databases providing information about the protein-protein interactions are given below:

1. Biomolecular Interaction Network Database - BIND

2. Database of Interacting Proteins - DIP

3. Search Tool for the Retrieval of Interacting Genes/Proteins - STRING

4. General Repository for Interaction Datasets - GRID

5. Human Protein Reference Database - HPRD

6. Molecular Interactions Database - MINT

The protein-protein interaction network is generally represented as undirected graph G(V,E), where the set of nodes V are the proteins. An edge (p1, p2) $\in$ E is present iff there is an interaction between the two proteins p1 and p2. Multiple sources of protein interaction networks from different studies and databases represent the protein interaction networks differently.

In this work we have utilized the protein interaction information present in the STRING database [43]. We mapped the time-specific differentially expressed genes to protein functional networks contained in the STRING database. The edges in the functional protein networks provided by the STRING database are weighted with a score from 200 to 1000 on the basis of their confidence, with 200 being least and 1000 being the highest confidence. To minimize the false positive interactions in the retrieved protein functional association networks, a minimum confidence level of 700 was used to construct the networks. Details about how the networks were extracted and used for the identification of genes, is outlined in chapter 2.

1.3.5 Biological pathways

Biological pathways represent the biological reactions and interaction network in a cell. Each reaction is identified with its enzyme, which in turn is coded by certain genes. Identification of the pathways involved in the disease helps in better understanding of the disease pathology. In general, the identified set of genes corresponding to a particular condition is utilized to identify such pathways. Such a process of identification is called pathway enrichment. Many online tools are available which can identify the enriched set of pathways in a given set of genes. DAVID tool[27] is one of the tools which can be used for pathway enrichment. DAVID uses the KEGG [31] database to obtain the information about pathways and involved genes. KEGG is one of the well curated databases which contains detailed and updated information about the pathways and the genes involved. In the present work, the identified genes are enriched for the pathways using the DAVID tool.

1.4 Complex networks

To understand biological networks, some fundamental notions are needed. we briefly introduce some main concepts that are used in the rest of the thesis. We also introduce some of the classical network centrality measurements with possible biological meaning in the protein interaction networks. Let G = (N,E) an undirected graph, with n = |N| vertexes. deg(v), indicate the degree the vertex. dist(v,w) is the shortest path between v and w. σ_{st} is the number of shortest paths between s and t and σ_{st}(v) is the number of shortest paths between s and t passing through the vertex v.

1.4.1 Random networks

Random networks are the networks with nodes N in which the occurrence of the edges between any 2 nodes has some probability p. Consequently the total number of edges is a random variable with the expected value $E(n) = p[N(N-1)/2]$. Most of the natural networks has been identified to be non-random [1]. The random networks are often used as the reference for evaluation of properties in the natural networks.

1.4.2 Network degree

This refers to the number of nodes adjacent to a given node v, where adjacent means directly connected. Calculation of the degree allows determining the degree distribution P(k), which gives the probability that a selected node has exactly k links. P(k) is obtained by counting the number of nodes N(k) with k = 1, 2, 3 ... links and dividing by the total number of nodes N. Determining the degree distribution allows distinguishing different kind of graphs. For instance, a graph with a peaked degree distribution (Gaussian distribution) indicates that the system has a characteristic degree with no highly connected nodes (most of the nodes have average degree). This is typical of random, non-natural, networks. By contrast, a power-law degree distribution indicates the presence of few nodes having a very high degree. And this behavior has been shown to be non-random [1]. Networks displaying a degree distribution approximating a power-law, $P(k) \propto k^{-\gamma}$, where γ is degree exponent, are called scale-free networks [8]. Many networks have been identified to show the characteristics of the scale free networks [7, 8]. Nodes with high degree (highly connected) are called hubs and hold together several nodes with lower degree. This behavior gives these networks a kind of robustness against random node deletion/failure [2, 30]. Additionally it also gives an opportunity to identify few network influential nodes in disease related biological networks as potential therapeutic targets[9, 10, 20].

In biological terms, the degree allows an immediate evaluation of the regulatory relevance of the node. For instance, in signaling networks, proteins with very high degree are interacting with several other signaling proteins, thus suggesting a central regulatory role, that is they are likely to be regulatory hubs.

1.4.3 Closeness centrality

The closeness centrality of a node v is calculated by computing the sum of the shortest path between the node v and all other nodes in the graph, and then dividing by the number of nodes. Once this value is obtained, its reciprocal is calculated, so higher values assume a

positive meaning in term of node proximity. Closeness centrality of a node i in the network G is

$$C_i^c = \frac{N-1}{\sum_{j \in G} d_{i,j}}$$

where, $d_{i,j}$ is the shortest path between the nodes i and j, and N is the total number of nodes in graph G.

Notably, high values of closeness centrality should indicate that all other nodes are in proximity to node v. In contrast, low values of closeness should indicate that all other nodes are distant from node v.

The closeness of a node in a biological network, for instance a protein-signaling network, can be interpreted as a measure of the possibility of a protein to be functionally relevant for several other proteins, but with the possibility to be irrelevant for few other proteins. Thus, a protein with high closeness, compared to the average closeness of the network, will be easily central to the regulation of other proteins but with some proteins not influenced by its activity.

1.4.4 Betweenness centrality

The betweenness centrality of a node n is calculated considering couples of nodes (s,t) and counting the number of shortest paths linking s and t and passing through the node n . Then, the value is related to the total number of shortest paths linking s and t. Thus, a node can be traversed by only one path linking s and t, but if this path is the only connecting s and t the node i will score a higher betweenness value. Mathematically, betweenness centrality of a node n in the network can be represented as:

$$C_n^b = \sum_{s \neq n \neq t} \left(\frac{\sigma_{st}(n)}{\sigma_{st}} \right)$$

where, s and t are nodes from the network different from n, σ_{st} denotes the number of shortest path from s to t, and $\sigma_{st}(n)$ denotes the number of shortest path from s to t passing through n.

The betweenness centrality of a node in a biological network, for instance a protein interaction network, can indicate the relevance of a protein as functionally capable of holding together different communicating proteins. The higher the value the higher the relevance of the protein as organizing regulatory component. Betweenness centrality of a protein effectively indicates the capability of a protein to bring together communication distant proteins. In signaling modules, proteins with high betweenness centrality are likely to be crucial in maintaining functionality and coherence of signaling mechanisms.

CHAPTER 2

RELATED WORK AND DATASET

The scope and application of the experiments depends upon the data completeness and its applicability. To study and explore the biological data, we need to tackle biological noise in addition to the measurement noise in the data. The data used in this work has been taken from the microarray experiments performed in the works of Hwang et al. 2009 [28], which to some extent, dealt with the above mentioned noise. As outlined in section 1.3.1, different prion strains have different effects depending upon the genotypes of the host. This contributes in different disease pathologies. To identify the core mechanisms in the prion disease progression, Hwang et al. performed large scale microarray experiments on different mouse-prion models. The subsequent sections will explain the methods employed by them to prepare and filter the data. Section 2.1 will give information about the experimental design, data preparation and data used in the their work. It will also give the details about the data used in present work. Next, in section 2.2 we will give a brief overview of the procedure used in Hwang et al. to identify the core 333 DEGs prevalent in the prion disease progression. Finally, section 2.3 gives details about the methodology used in this study to identify disease related genes.

2.1 Data preparation and filtering

In this section, we will discuss about the procedure followed in the works of Hwang et al. to obtain the data for studying the prion disease progression.

2.1.1 Mice and prion infection

For studying the effects of different prion strains on varying genotypes of the host, Hwang's work focused on 8 different mouse strain-prion strain combinations. B6, B6.I, FVB, 0/+, 0/0, and Tg4053 mice were used. These mice were inoculated intracerebrally at 5 weeks of age with $10\mu l$ of 10% brain homogenate from clinically ill mice or normal control mice. The 2 prion strains used were: RML(mouse adapted scrapie prions) and 301V(mouse adapted BSE prions). The inoculated mice were killed at certain intervals for RNA preparation.

2.1.2 Data generation and normalization

Affymetrix GeneChip mouse array 430 2.0 containing about 45,000 probesets corresponding to ~ 25,000 annotated genes was used. Whole brains were collected from three mice at each time point in each mouse-prion combination. Total RNA isolated from each mouse was then hybridized individually according to Affymetrix manual. The total dataset from 450 chips represents ~ 20 million data points. The array intensities in each group were normalized using gcRMA algorithm.

2.1.3 Identification of DEGs in individual mouse-prion combinations

For the time-course intensities from each mouse-prion combination differing in time intervals and numbers of time points, they used a single statistical method which can be outlined as given below.

1. For each gene, a log2 median ratio(x) of the mRNA intensities from infected brains to those of the normal(control) brains was computed at each time point. This step was basically done to cancel out the responses caused by the intracerebral inoculation or ageing.
$$x = \log_2(\frac{median(I_{infected})}{median(I_{normal})})$$
where, median(I) indicates the median value of the intensity in the spot of the particular probeset.

2. The log2 ratios of every gene was then integrated numerically (trapezoidal integration) over the entire incubation time to obtain the area under the ratio profile as a statistical measure.
$$T = \int_{t=0}^{Tl} x dt$$
where, Tl is the life span for the given prion-mouse combination.

3. To perform the statistical hypothesis testing, two control sets of data ($SI_{infected}$ & $SI_{control}$) were generated for each gene at each time point. The following steps were taken.

   a) $SI_{infected}$ and $SI_{control}$ were sampled from $N(m_{control}, s_{control})$ to have the same size as the size of the matrix containing the intensity values for all the genes at a particular time, where $m_{control}$ and $s_{control}$ are the mean and standard deviation of $\log_2(I_{control})$.

   b) For each gene, the log2 median ratio was computed for $SI_{control}$ and $SI_{infected}$ at every time point.

   $$Sx = \log_2\left(\frac{median(SI_{infected})}{median(SI_{control})}\right) \tag{2.1}$$

   c) The log2 ratios of every gene was then integrated numerically (trapezoidal integration) over the entire incubation time to obtain the area under the ratio profile.

   $$ST = \int_{t=0}^{Tl} Sx\,dt \tag{2.2}$$

4. A null hypothesis distribution (probability density function of $H_0 - pdf(H_0)$) was then empirically driven using Gaussian kernel density estimation for such areas.

5. A p-value for each gene was then calculated using pdf(H_0) by two-tailed test of the corresponding area in step 2.

   $$P_i = 2 \times min(1 - cdf(ST_i = T_i), cdf(ST_i = T_i)) \tag{2.3}$$

   where T_i is the area under the ratio profile computed in step 2 for gene i. A gene with a small p-value, thus having a large area under the temporal profile, shows a clear temporal differential pattern from an early time point. In this method, pdf(H_0) was directly modeled from control data, thus representing characteristics of various technical and biological sources of variability in data. As a result, this method captures significant temporal patterns more effectively than those using pair-wise comparisons of infected and control samples at each time point and then combining p-values from the individual comparisons. The p-value at each time point was also computed to determine whether a gene was differentially expressed at that time point or not.

2.1.4 Dataset used in present work

The data used in this study were downloaded from prion disease database (PDDB) [23]. This database contains all the results and data corresponding to the experiments performed in the works of Hwang et al . PDDB also provides the p-values for the differential expression of the genes at different time points corresponding to different mouse strain-prion strain combinations. We downloaded a list of 13,822 genes that were expressed in 5 mouse-prion combinations in the works of Hwang et al. In particular, these genes with their corresponding time specific p-values in 6 different mouse strain-prion strain combinations were used in this study. Table 2.1 lists the details about the mouse-prion combinations used in this study.

Table 2.1: Mouse strain-prion strain combinations

Mouse strain	Prnp genotype[1]	Prion strain	End point[2]	Harvest interval
C57BL/6J (BL6)	a/a	RML	23	2
C57BL/6J (BL6)	a/a	301V	41	4
C57BL/6I-1 (B6I)	b/b	RML	48	4
C57BL/6I-1 (B6I)	b/b	301V	18	2
FVB/Ncr (FVB)	a/a	RML	22	2
FVB.129-prnp (FVB)	0/0	RML	28	4

The table lists different mouse-prion combinations used in this study.

[1]Host genotype for PrP^C protein.

[2]Represents the clinical death of mouse.

2.2 Hwang's work

The study performed in Hwang et. al. employed several mouse-prion combinations to identify prion related genes. They tracked the global gene expression in the brains of eight distinct mouse strain-prion strain combinations throughout the progression of the disease to capture different effects of prion strain, host genetics, and PrP^c concentration on disease incubation time. The two different strains of prions used were, Rocky Mountain Lab (RML) mouse-adapted scrapie prions and 301V mouse-adapted BSE prions which showed drastic difference in the incubation times when inoculated in the genetically same hosts. They exploited 5 mouse-prion models to identify a core of 333 differentially expressed genes (DEGs) that appeared central to prion disease. These DEGs were then mapped to the biological pathways and networks to identify the modules that may be involved in the disease

progression. These common DEGs showed similar expression patterns over the incubation times in 5 different mouse strain-prion strain combinations.

The shared genes identified in their study were used to obtain dynamic protein interaction networks corresponding to different time-stamps in the disease progression. The dynamic nature of the network was captured via changes in the differential expression of the genes present as nodes in the networks. These dynamic networks only capture the dynamics of the nodes and neglects the edge dynamics which may be present at different time-stamped networks related to the disease. The present work focus on identification of core disease related genes, through exploitation of the protein functional networks which show both node and edge dynamics with time.

2.3 Method employed in present work

One of the aims of this study was to highlight the potential common DEGs involved the progression of the disease. The 5 mouse-prion combinations were used to detect 148 DEGs. The general criterion employed was to identify the DEGs whose protein products are present at the network influential positions in the disease related temporal protein functional networks. These DEGs represent the genes which are potentially involved in the prion disease progression and present centrally in the time-stamped protein functional networks. The central involvement of these genes in the disease network increases their potential disease involvement with respect to the ease of communication between them and the other network DEGs.

Mapping the DEGs to protein functional networks resulted in loss of many DEGs (many DEGs in the network were isolated from the main central network) which can be attributed to the following reasons: (1) the lack of complete knowledge about the mapping of the genes to its functional proteins, and (2) the incomplete knowledge about the protein interactions. These limitations resulted in relatively incomplete protein functional network which may have missed some important DEGs involved in the disease network. Even though the network is incomplete, this study revealed many DEGs which were previously related to the pathogenesis of prion disease. Hence we can hypothesize that the novel disease related DEGs discovered in this study should also be considered important from the disease network's point of view. Newly identified DEGs which were not highlighted in the Hwang's study reflects the fact that these genes, although doesn't show expression correlation with time, are present centrally in the disease related protein functional networks and should be considered equally important. DEGs found to be highly correlating in their expression

should be considered important as also the DEGs which are present centrally in the protein association network.

2.3.1 Disease related protein functional networks

For every mouse strain-prion strain combinations at a particular time-stamp, the DEGs were identified by taking the genes having the p-values for differential expression as less than a defined threshold value (< 0.05). Using the STRING database, the time specific DEGs corresponding to a particular mouse strain-prion strain combination were mapped to the protein functional interaction network. STRING database maps a gene to only one of its protein products. In this work, the term gene and protein have been used interchangeably. A time-stamped network signify the disease related network for a particular mouse strain-prion strain combination at a specific point of time. The edges in the functional protein networks provided by the STRING database are weighted with a score from 200 to 1000 on the basis of their confidence, with 200 being least and 1000 being the highest confidence. To minimize the false positive interactions in the retrieved protein functional association networks, a minimum confidence level of 700 was used to construct the networks. The obtained networks mostly consisted of a single large component and several small disconnected components. The largest connected component of the networks were used for further analysis. The following tables provide details about the network structures. The protein functional networks were extracted from the STRING database. The data highlights the lack of knowledge about the functional association of genes to its protein products (number of mapped proteins are less than the query genes). Also the incomplete knowledge corresponding to the functional relations between different proteins is evident by the large difference between the largest connected component and the number of mapped proteins. Even though the current knowledge about the protein interaction networks is incomplete, we can still get valuable insights about the diseased condition by using this type of network studies.

Table 2.2: Network information for B6.I-Prnp(b/b)-301V combination:

Dataset type	**W2**	**W4**	**W6**	**W8**	**W10**	**W12**	**W14**	**W16**	**W18**
Number of DEGs	1590	1557	1664	1609	1690	1655	1715	1775	1779
Number of mapped proteins	539	465	539	475	599	520	620	614	699
Number of nodes in LCC	415	386	432	421	498	435	526	516	588
Number of edges in LCC	1119	1078	1104	1001	1497	1188	1732	1839	2124

Mouse strain = C57BL/6I-1, Prion strain = 301V mouse adapted BSE prions, Host genotype = b/b
For filtering the DEGs at a particular time-stamp, we have used a p-value threshold (< 0.05). Same criteria has been used for the DEGs identification across different mouse strain-prion strain combinations.

Table 2.3: Network information for B6.I-Prnp(b/b)-RML combination:

Dataset type	**W4**	**W8**	**W12**	**W16**	**W20**	**W24**	**W28**	**W32**	**W36**	**W48**
Number of DEGs	1545	1575	1584	1658	1636	1649	1673	1697	1689	1654
Number of mapped proteins	461	506	500	537	393	498	479	572	589	615
Number of nodes in LCC	379	412	404	433	277	407	364	478	517	527
Number of edges in LCC	960	1069	1023	1014	590	1004	999	1625	1570	1988

Mouse strain = C57BL/6I-1, Prion strain = Rocky mountain lab (RML) mouse adapted scrapie prions, Host genotype = b/b

Table 2.4: Network information for BL6-Prnp(a/a)-301V combination

Dataset type	**W6**	**W8**	**W12**	**W16**	**W20**	**W24**	**W28**	**W32**	**W36**	**W48**
Number of DEGs	1738	1767	1610	1593	1655	1715	1729	1571	1721	1668
Number of mapped proteins	556	580	446	522	509	585	619	449	447	546
Number of nodes in LCC	467	480	389	421	426	495	541	378	324	438
Number of edges in LCC	1164	1323	964	976	1076	1528	1649	1199	717	1411

Mouse strain = C57BL/6J, Prion strain = 301V mouse adapted BSE prions, Host genotype = a/a

Table 2.5: Network information for BL6-Prnp(a/a)-RML combination

Dataset type	**W2**	**W4**	**W6**	**W8**	**W10**	**W12**	**W14**	**W16**	**W18**	**W20**	**W22**	**W23**
DEGs	135	41	49	48	36	4085	122	241	1760	897	648	1981
proteins*	17	5	9	4	12	1035	61	109	496	339	241	700
nodes	7	3	3	4	7	925	59	105	431	290	216	641
edges	6	3	2	4	9	2962	250	382	1292	1007	834	2112

Mouse strain = C57BL/6J, Prion strain = Rocky mountain lab (RML) mouse adapted scrapie prions, Host genotype =a/a

*The small number of nodes in the early time-stamped networks is the cumulative effect of both incomplete knowledge and less number of DEGs at these time-points.

Table 2.6: Network information for FVB-Prnp(a/a)-RML

Dataset type	**W2**	**W4**	**W6**	**W8**	**W10**	**W12**	**W14**	**W16**	**W18**	**W20**	**W22**
Number of DEGs	582	657	586	466	635	1194	1077	935	950	3734	1852
proteins	135	166	156	84	165	407	397	339	355	1283	666
Number of nodes in LCC	91	73	30	46	117	316	322	288	287	1191	587
Number of edges in LCC	124	117	36	66	208	861	854	796	871	4311	2101

Mouse strain = FVB/NCr, Prion strain = Rocky mountain lab (RML) mouse adapted scrapie prions, Host genotype = a/a

Table 2.7: Network information for FVB-Prnp(0/0)-RML

Dataset type	**W4**	**W8**	**W12**	**W16**	**W20**	**W24**	**W28**	**W51**
Number of DEGs	1659	1714	1714	1754	1699	1823	1630	1595
Number of mapped proteins	514	556	565	552	527	599	510	540
Number of nodes in LCC	442	472	452	420	443	501	400	461
Number of edges in LCC	1168	1153	1203	1051	1073	1360	1009	1185

Mouse strain = FVB.129-prnp, Prion strain = Rocky mountain lab (RML) mouse adapted scrapie prions, Host genotype = 0/0

This combination was used as a control combination for the differential network study of all the other combinations.

2.3.2 Pathway enrichment

We used the DAVID tool [27] for the identification of the enriched KEGG pathways and GO terms in the obtained set of 148 genes. DAVID employs a modified version of Fischer's test to score the identified GO terms or KEGG pathways in the set of input genes. A p-value cutoff of 0.05 was used for the modified Fischer's test and affymetrix mouse 430 2.0 genome was used as the background for the purpose of enrichment.

CHAPTER 3

NETWORK ANALYSIS

The components inside a cell work in a coordinated fashion to maintain a homeostatic environment. This homeostatic environment is necessary for the cell to perform optimally, so as to contribute to the survival of the organism as a whole. Proteins, as one the types of components inside the cell, interact and communicate among themselves, to perform necessary tasks. Proteins also show changes in inter-protein interaction behavior in response to external perturbations, intra-cellular dis-balance and many other factors. At the network level, these changes corresponding to a particular protein occurs via change in its interacting neighbors. These topological changes in the biological networks can be utilized to identify the nodes involved in a certain perturbed condition. Protein functional network related to the differentially expressed genes (DEGs) should be considered important since it signifies a dysregulated network due to onset/progression of the disease. Identification of the core elements of such structure may help in capturing the main regulatory components which can prove to be potential therapeutic targets.

This chapter will explain the network properties shown by the protein functional networks related to prion disease and how the networks were utilized to identify potential disease related genes present in the core of the networks. Section 3.1 will explain the global network properties of the temporal PFNs. Next, in section 3.2 we explain the studies performed to capture sub-graph distribution in temporal networks. Then, we explain the concept of differential network analysis. In section 3.3 we give the details about the results and section 3.4 ends with discussion and conclusion.

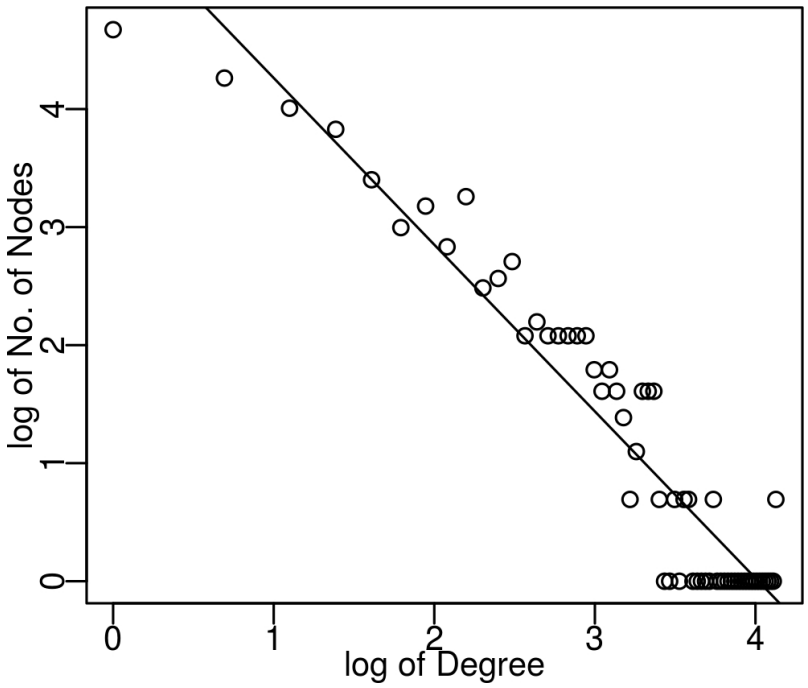

FIGURE 3.1. Typical degree distribution of the network at a particular time point. The dots represent the observed degree distribution. And the line approximates the behavior. This approximates a power law distribution with the degree distribution constant $\gamma \sim 1.5$.

3.1 Global topological properties

It is anticipated that the effects of differential expression of the genes in different mouse strain-prion strain combinations will be reflected in the temporal protein functional networks. While the local changes were evident, there were negligible changes in the global properties of the temporal networks corresponding to any of the mouse-prion combinations. The degree distributions of most of the temporal networks related to a particular mouse strain-prion strain combinations showed power law nature (Figure 3.1) with degree exponent of ~ 1.5. For the mouse strain-prion strain combinations, BL6-RML and FVB-RML, the time specific networks corresponding to the first half of the incubation period (here, incubation period is defined as the time between the prion inoculation and mouse death) were small (due to less number of DEGs in that time point) and disconnected (mainly because of the incomplete knowledge about the PPI networks). Hence their degree distribution showed deviations from power law. For the rest of the combinations including the control combination, the temporal networks showed scale free nature. The global properties of the

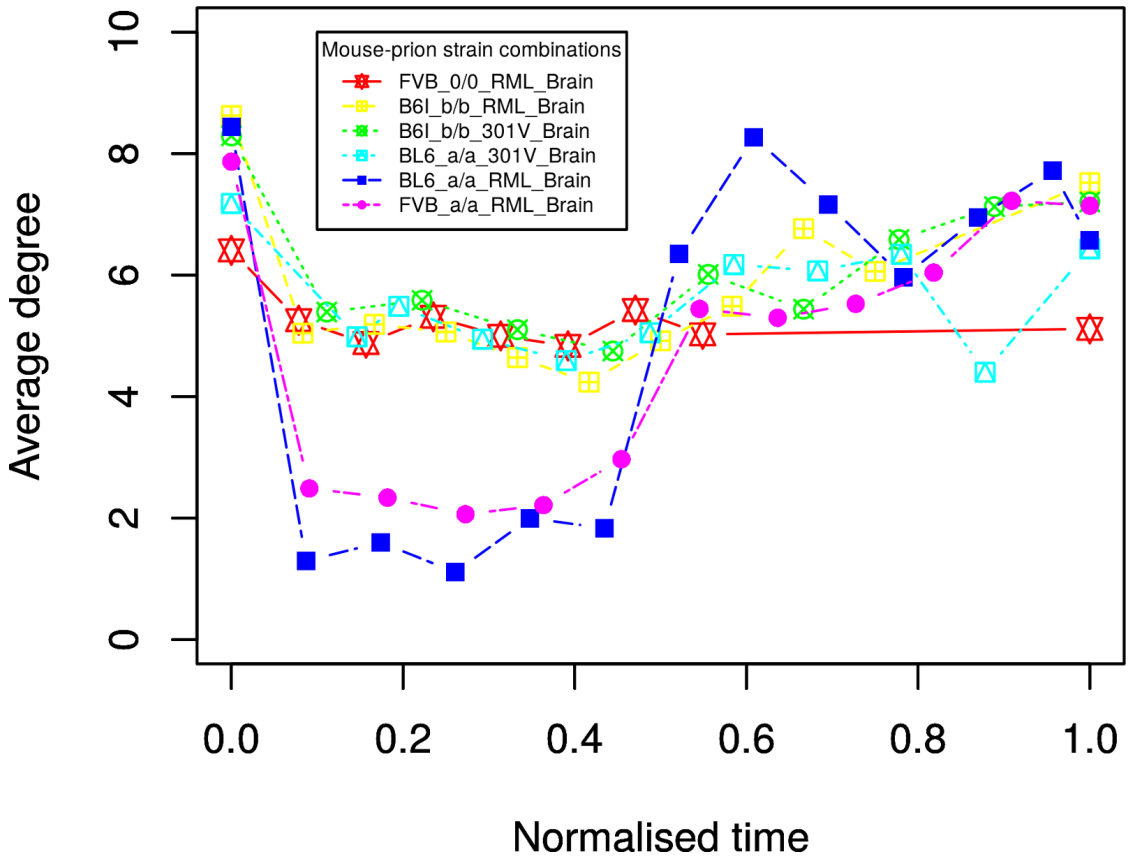

FIGURE 3.2. Change in the average degree of the networks with time. No considerable difference shown between different mouse-prion combinations.

time specific protein functional networks were analyzed using Cytoscape[39].

The scale free property of the identified networks suggests robustness of these networks to the random node failures. Also the power law distribution of the obtained networks is maintained till the end of the disease, suggesting the robust nature of the network. Also, we do not find considerable change in the global properties of the disease related networks in comparison to the control networks (time specific protein functional networks related to FVB-RML-0/0 combination.). This negligible difference suggest that the differential network study using the global properties may not prove to be fruitful in terms of capturing the difference between the disease and control network topology. Thus, our focus for the differential network study is limited to only local changes in the network topology. The tables below provide details about the time-specific global properties of the protein functional networks corresponding to different mouse-prion combinations.

Table 3.1: Global properties details for B6.I-Prnp(b/b)-301V combination:

Property	**W2***	**W4**	**W 6**	**W8**	**W10**	**W12**	**W14**	**W16**	**W18**
Avg. clustering coeff.	0.38	0.35	0.36	0.38	0.40	0.38	0.39	0.41	0.38
Centralization	0.08	0.09	0.07	0.07	0.09	0.08	0.09	0.11	0.09
Avg. shortest path	4.56	5.00	5.31	4.99	4.61	4.35	4.99	4.22	4.28
Density	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Network heterogeneity	0.99	0.99	1.05	0.95	1.05	1.03	1.02	1.07	1.06
Power-law coefficient	1.51	1.48	1.48	1.48	1.61	1.51	1.48	1.41	1.45

* Wi represents the network at week i

Table 3.2: Global properties details for B6.I-Prnp(b/b)-RML combination:

Property	**W4**	**W8**	**W12**	**W16**	**W20**	**W24**	**W28**	**W32**	**W36**	**W48**
Avg. clustering coeff.	0.38	0.34	0.36	0.31	0.33	0.35	0.36	0.37	0.34	0.37
Centralization	0.06	0.08	0.09	0.07	0.04	0.06	0.06	0.08	0.06	0.09
Avg. shortest path	4.65	4.53	4.80	5.03	5.48	4.76	4.62	4.36	4.79	4.26
Density	0.01	0.01	0.01	0.01	0.02	0.01	0.02	0.01	0.01	0.01
Network heterogeneity	0.93	0.98	0.93	1.01	0.81	0.94	0.97	1.02	1.01	1.05
Power-law coefficient	1.47	1.53	1.54	1.58	1.44	1.57	1.37	1.45	1.48	1.40

Table 3.3: Global properties details for BL6-Prnp(a/a)-301V combination

Property	**W6**	**W8**	**W12**	**W16**	**W20**	**W24**	**W28**	**W32**	**W36**	**W48**
Avg. clustering coeff.	0.36	0.35	0.31	0.36	0.38	0.36	0.36	0.40	0.38	0.38
centralization	0.06	0.06	0.07	0.07	0.06	0.07	0.06	0.10	0.08	0.10
Avg. shortest path	4.69	4.63	4.88	5.54	4.97	4.14	4.65	4.56	5.54	4.46
density	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.01	0.01
Network heterogeneity	0.95	0.97	1.00	0.89	0.94	1.03	0.95	1.04	0.95	0.95
Power-law coefficient	1.63	1.53	1.52	1.66	1.55	1.46	1.57	1.39	1.55	1.46

Table 3.4: Global properties details for BL6-Prnp(a/a)-RML combination

Property	**W12**	**W14**	**W16**	**W18**	**W20**	**W22**	**W23**
Avg. clustering coeff.	0.40	0.58	0.49	0.43	0.40	0.43	0.40
Centralization	0.05	0.30	0.26	0.08	0.14	0.18	0.06
Avg. shortest path	4.96	2.62	3.36	4.84	3.99	3.48	4.55
Density	0.01	0.14	0.07	0.01	0.02	0.04	0.01
Network heterogeneity	1.05	0.68	0.89	1.02	1.02	0.98	1.00
Power-law coefficient	1.63	1.01	1.09	1.45	1.26	1.13	1.54

For the time-points earlier than 12 weeks, the number of DEGs were very less resulting in sparse disconnected network.

Calculating the global properties for these networks didn't make sense.

Table 3.5: Global properties details for FVB-Prnp(a/a)-RML

Property	**W12**	**W14**	**W16**	**W18**	**W20**	**W22**
Avg. clustering coeff.	0.37	0.42	0.38	0.38	0.40	0.38
Centralization	0.08	0.10	0.09	0.10	0.05	0.08
Avg. shortest path	4.84	4.87	4.52	4.13	4.64	4.27
Density	0.02	0.02	0.02	0.02	0.01	0.01
Network heterogeneity	0.93	0.97	0.98	0.94	1.08	1.10
Power-law coefficient	1.41	1.43	1.36	1.34	1.66	1.41

For the time-points earlier than 12 weeks, the number of DEGs were very less resulting in sparse disconnected network.

Table 3.6: Global properties details for FVB-Prnp(0/0)-RML

Property	**W4**	**W8**	**W12**	**W16**	**W20**	**W24**	**W28**	**W51**
Avg. clustering coeff.	0.34	0.40	0.38	0.36	0.38	0.40	0.38	0.36
centralization	0.08	0.06	0.06	0.07	0.06	0.06	0.07	0.09
Avg. shortest path	4.65	4.88	4.92	4.61	5.01	4.91	4.68	4.78
density	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Network heterogeneity	1.03	0.88	0.91	0.99	0.93	0.90	0.95	1.02
Power-law coefficient	1.53	1.66	1.59	1.54	1.62	1.63	1.54	1.52

3.2 Subgraph ratio profile

Complex networks related to biology can be of varying sizes and complexities. This makes it difficult to compare them. Milo et al. [33] proposed a method to compare and systematically study similarity in the local structure of networks, based on the significance profile of small sub-graphs in the network compared to randomized networks.

3.2.1 Basic concept

This methodology is a step towards understanding the design principles of the complex networks. The approach is based on comparing the sub-graph profiles (SP) of the networks to be compared. A similar profile postulates similarity in the origin of the networks concerned. To calculate the SP of a network, it is compared to an ensemble of randomized networks with the same degree distribution. The comparison to randomized networks compensates for effects due to network size and degree distribution. For each sub-graph i, its abundance relative to the randomized networks is taken as

$$\delta_i = \frac{Nreal_i - < Nrand_i >}{Nreal_i + < Nrand_i > + \epsilon}$$

where, $Nreal_i$ is the frequency of subgraph i in the concerned network, $Nrand_i$ is the average frequency of subgraph i in the randomized set of networks, ϵ is a constant which ensures that $|\delta|$ is not misleadingly large when the subgraph appears very few times in both the real and random networks. The subgraph ratio profile (SRP) of a network is then given by an array containing the normalized values of the subgraph abundances ($\delta_i' s$):

$$SRP_i = \frac{\delta_i}{(\sum \delta_i^2)^{1/2}}$$

3.2.2 Application on prion PPI networks

We hypothesize that the time-specific protein functional networks related to the control mouse-prion combination will show different SRP in comparison to the any of the disease reproducing combinations. And that the differential gene expression with respect to the control and disease reproducing combination, follows different constraints and should give different profiles for the local structures corresponding to their time-stamped networks. We also hypothesize that the SRP of the networks should change as the disease progress. Figure 3.3 shows the SRP of the time-stamped networks corresponding to the control combination. And figure 3.4 shows the SRP of time-stamped networks corresponding to one of the diseased

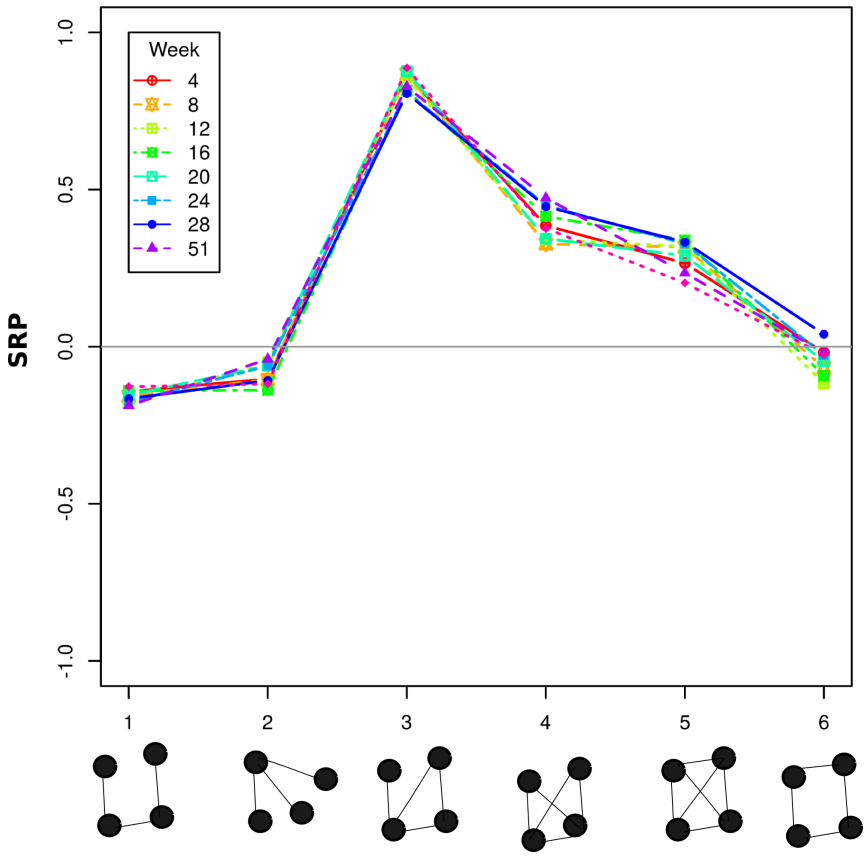

FIGURE 3.3. The plots show the SRP for the 4 node subgraphs. For the undirected networks there are only 6 configurations of 4 node subgraphs. Significance ratio profile of the networks related to the control combination (FVB-00-RML).

reproducing combinations. The results show that there is negligible difference between the SRPs corresponding the networks of diseased and control combinations. Also it can be seen that there is no deviation with time, in the behavior of SRPs belonging to a particular mouse-prion combination. All the time-stamped networks seems to have similar SRP. The results conclude that although there is slight changes in the subgraph profiles, this study of local structure distribution doesn't show much promise in distinguishing protein networks corresponding to diseased and control conditions.

For the calculation of the sub-graphs present in the networks a tool called FANMOD[44] was used. This tool gives the frequency of different sub-graphs in a given network and also the mean frequency of the same sub-graph corresponding to the ensemble of random networks.

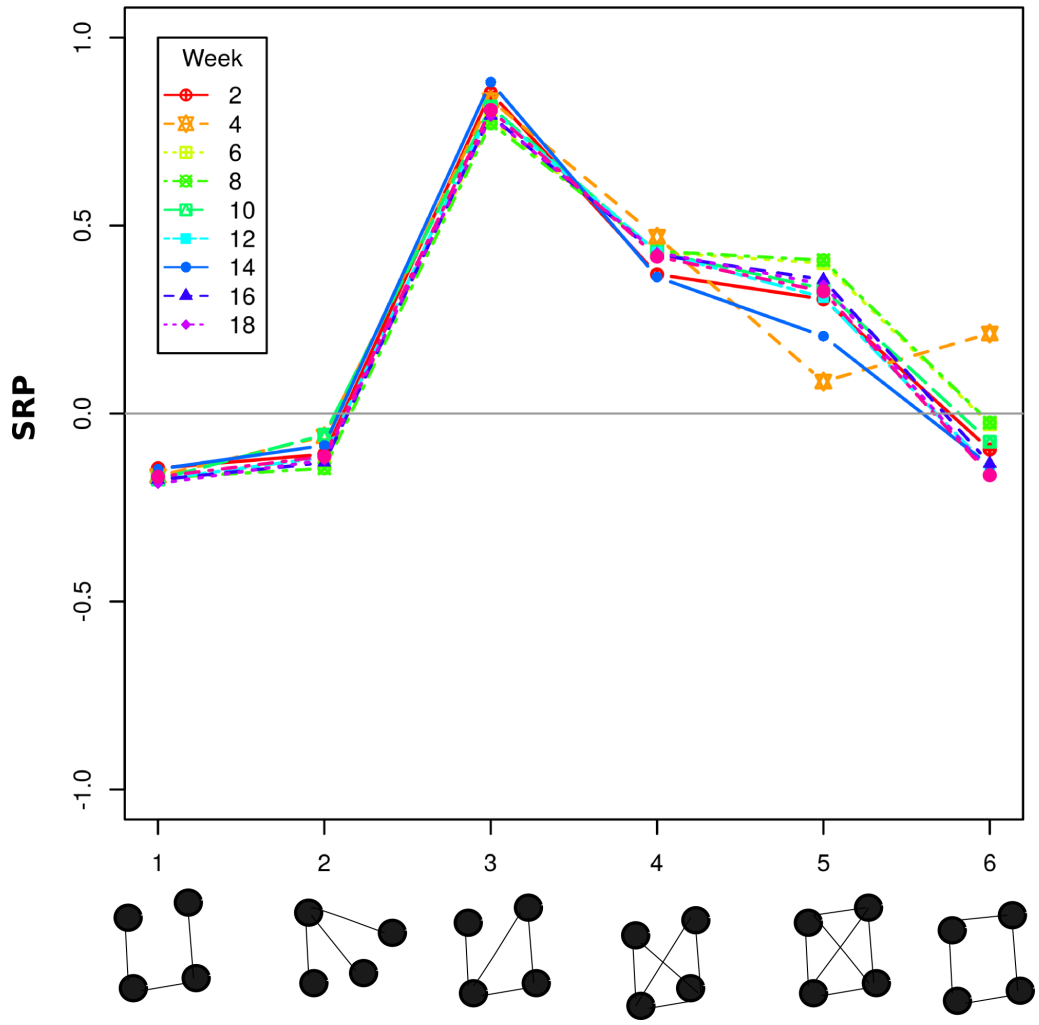

FIGURE 3.4. The plots show the SRP for the 4 node subgraphs. For the undirected networks there are only 6 configurations of 4 node subgraphs. Significance ratio profile corresponding to the disease reproducing combination (B6I-bb-301V).

3.3 Differential network analysis

We used FVB.129-prnp(0/0) mouse strain-prion strain combination which lacks PrP^{c}, as the control combination. This combination cannot be infected with prions [15]. Although this combination doesn't develop disease, initial inoculation of prion into these models may lead to differential expression of the genes involved in prion disease (due to the presence of the inoculated PrP^{Sc}). In the disease developing combinations also, these genes will be differentially expressed. For capturing the important genes among these kind of common DEGs in control and disease developing combinations, the local topological properties of

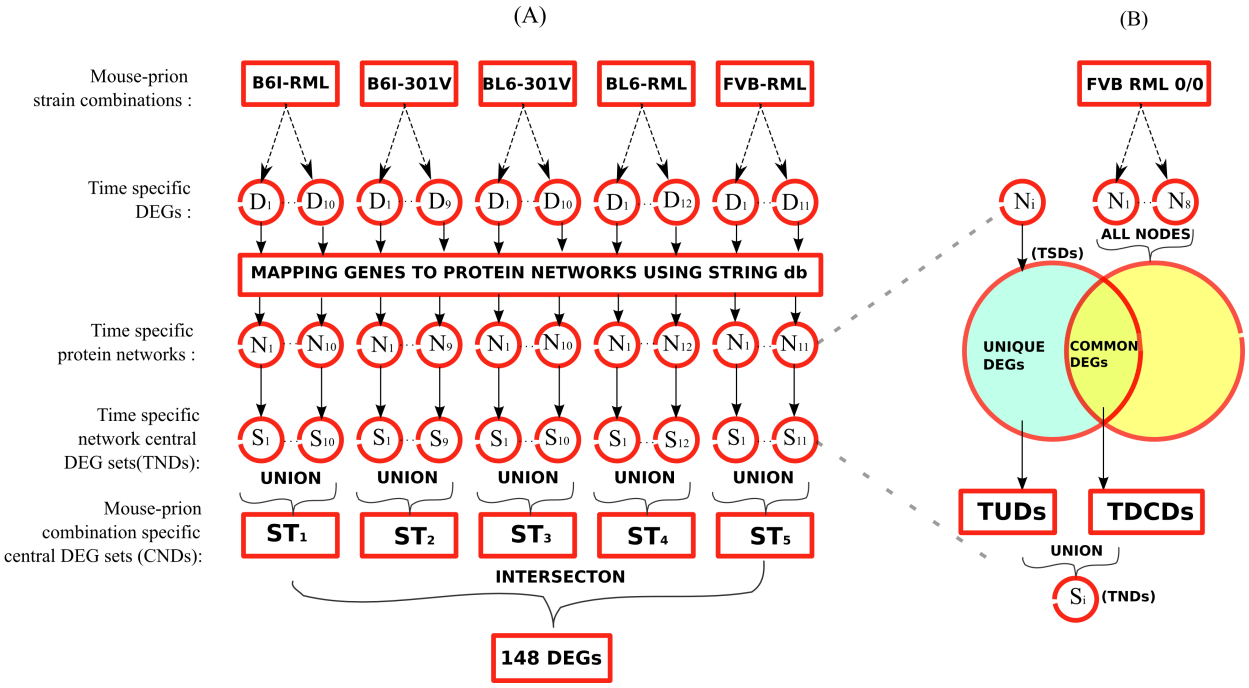

FIGURE 3.5. (A) Time-stamped PFNs (N_i) were created by mapping the time-stamped DEGs (D_i) to functional PPI networks. Then network central DEGs (S_i) were used to obtain network central DEGs (ST_i). (B) Each time specific PFNs (N_i) corresponding to any of the 5 diseased mouse strain-prion strain combinations were compared with the network DEGs of the mouse strain-prion strain combination which lacks PrP gene. First set consists of the hub DEGs which are unique at a particular timestamp. Second set consists of the DEGs which are common to both the diseased and control combinations but have relatively high centrality measures in diseased network compared to any of the time-specific networks corresponding to the control combination. Finally these two sets were combined to give time specific network central DEGs (S_i) for each diseased DEGs network.

the protein functional networks have been utilized. The hypothesis is that with the disease progression, the important disease related common DEGs will show considerable change in their network topological properties in the disease developing combination as compared to the control combination. For a particular disease developing mouse strain-prion strain combination, we compared their time-specific DEGs (TSDs) with all the DEGs (expressed at any point of time) of the control combination.

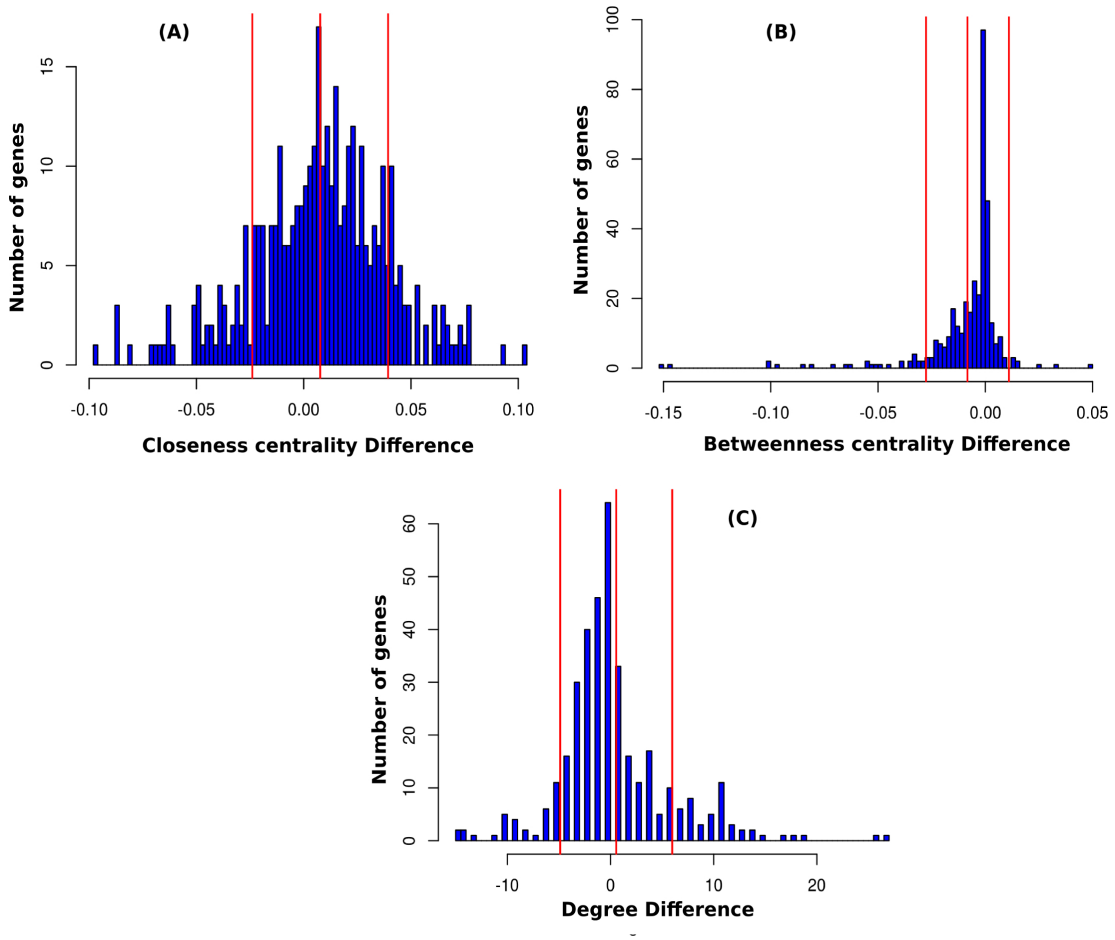

FIGURE 3.6. The left, middle and right red vertical lines in the plots shows the marks for the left standard deviation, mean and right standard deviation respectively. The genes present to the right of the right standard deviation has been taken as genes playing crucial roles in the disease progression. (A) Closeness centrality measures the closeness of a node to other nodes in the network. The graph here shows that for most of the genes in the disease network as compared to the control network, the difference in centrality is marginal. (B) Betweenness centrality measures the number of paths being traversed through a node. More paths through a node in the network make a node important from the communication point of view. This also shows the normal behavior supporting the initial hypothesis. (C) Degree measures the local communication and activity of a node in the network. The behavior of the graph is similar to other measurements.

Figure 3.5 shows the outline of the procedure used to identify the common network central DEGs related to the prion disease. The common DEGs present in TSDs (corresponding to a particular disease reproducing combination) and in the set of DEGs corresponding to control combination were taken for comparison on the basis of network centrality measurements. This comparison have been used to identify the DEGs which show incremental change in their network centrality behavior (incremental change in diseased network in comparison to control network). we call such DEGs which shows incremental network centrality behavior, as time-specific differential common DEGs or TDCDs.

We plotted the histogram of the calculated centrality difference against the number of DEGs (see Figure 3.6). We found that majority of the DEGs showed negligible change in the their centrality values and few showed considerable change. The centrality measurement difference is approximately normally distributed and hence we have chosen to include those proteins whose centrality measure difference is more than the standard deviation of the distribution (see Algorithm 1 for the procedure used to identify TDCDs). The DEGs which were present in the timestamped disease developing combination and not present in the set of DEGs corresponding to the control combination were referred as time-specific unique DEGs (TUDs). The TUDs were identified and then filtered on the basis of the degree values corresponding to the time-stamped protein functional network. We define a node in the time-stamped protein functional network as a hub if the node's degree is greater than 10 (more than the average degree in any time-stamped protein functional network). Only the DEGs fitting this hub definition were retained in the TUDs group. Both the TUDs and TDCDs were then combined to get time specific network central DEGs(TNDs). These TNDs were then combined to get combination specific network central DEGs(CNDs) (corresponding to any of the 5 disease reproducing mouse-prion models). All the 5 CNDs belonging to 5 different mouse strain-prion strain combinations were then pooled to identify DEGs which were present in at-least 4 of these CNDs groups.

The network centrality measures used were degree, betweenness centrality and closeness centrality. The definitions of these centrality measures are as follows:
Closeness centrality of a node i in the network G is

$$C_i^c = \frac{N-1}{\sum_{j \in G} d_{i,j}}$$

where, $d_{i,j}$ is the shortest path between the nodes i and j, and N is the total number of nodes in graph G.
Betweenness centrality of a node i in the network is

$$C_i^b = \sum_{s \neq n \neq t} \left(\frac{\sigma_{st}(n)}{\sigma_{st}} \right)$$

where, s and t are nodes from the network different from n, σ_{st} denotes the number of shortest path from s to t, and $\sigma_{st}(n)$ denotes the number of shortest path from s to t passing through n.

Algorithm 1 outlines the procedure used to identify TDCDs (The R implementation of the given algorithm can be downloaded from `https://sites.google.com/a/iiitd.ac.in/khalique-research/codes/`, along with other scripts to reproduce the results found in this study). The initialization of the pseudo-code is as follows:

1. N_i = Network at time-stamp i corresponding to a particular mouse-prion combination.
2. D_i = Set of DEGs of a particular mouse-prion strain combination at time-stamp i.
3. C_i = Common DEGs between D_i and all the DEGs of FVB-prnp-0/0.
4. M = Set of all time-stamped networks of FVB-prnp-0/0 combination.
5. $Diff = \phi$ = Set storing the centrality difference.
6. $Nodes()$ = Function returning the set of nodes of the input network.
7. $N_i map()$ = Maps the DEGs at time stamp i to their centrality values corresponding to the network at time stamp i.
8. $S_i = \phi$ = Set of DEGs selected as central for the time stamp i.

Algorithm 1 Common genes selection

for each DEG $'g'$ in C_i **do**
    temp = $N_i map(g)$
    value = 0
    for each network N_j in M **do**
        if $g \in Nodes(N_j)$ **then**
            value = MAX(value,$N_j map(g)$
        end if
    end for
    $Diff$ = $Diff \cup$ (temp - value)
end for
Std = standard-deviation($Diff$)
Mea = mean($Diff$)
for each DEG $'g'$ in C_i **do**
    temp = $N_i map(g)$
    value = 0
    for each network N_j in M **do**
        if $g \in Nodes(N_j)$ **then**
            value = MAX(value,$N_j map(g)$
        end if
    end for
    $Diff$ = $Diff \cup$ (temp - value)
    if $Diff >$ (Std+Mea) **then**
        $S_i = S_i \cup g$
    end if
end for

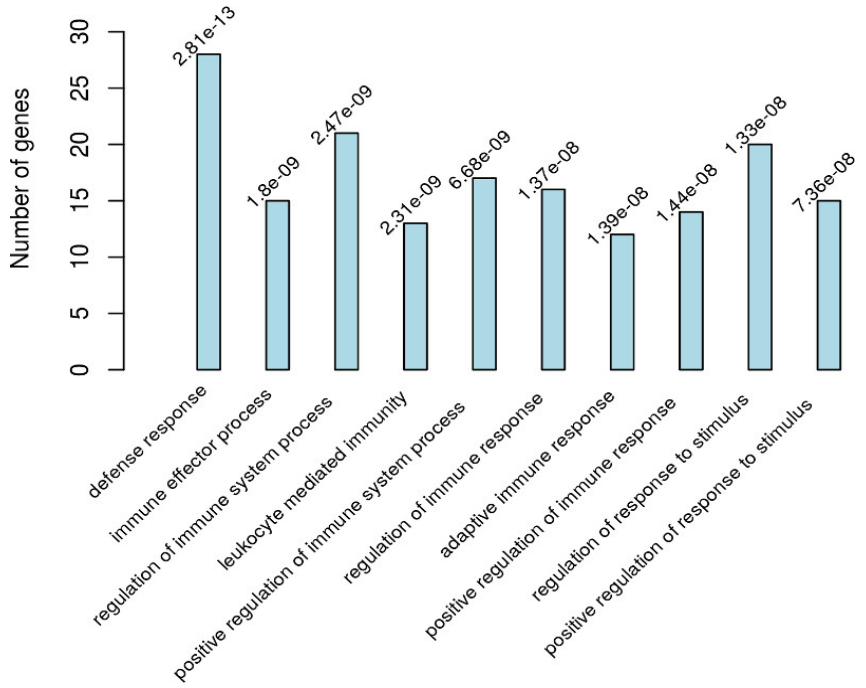

FIGURE 3.7. Top 10 GO enriched terms related to the 148 DEGs identified in this work. Gene ontology enrichment of the 148 DEGs shows that the immune response related functions are over-expressed. The bottom of the bars show the Gene Ontology terms (p-values are reported on the top of the bars)

3.4 Results and Discussion

3.4.1 Network central DEGs

To identify the genes involved in core processes and present centrally in the networks related to prion pathogenesis, we focused our study on the five mouse strain-prion strain combinations that reflect prion strain and $Prnp$ allele dependent incubation times with normal levels of cellular prion protein. We used the concept of differential network centrality to identify the genes that showed incremental behavior in their network centrality measurements as compared to the networks of control combination. Also, the unique genes (as compared to the networks of control combination) present as hubs in the time specific networks corresponding the five disease prone combinations were extracted. Overall, we identified 148 DEGs present centrally in the protein functional networks in at least four of the core mouse-prion combinations. These identified DEGs are one of the major findings of this study and has been used as a base to further explore the problem. To highlight the

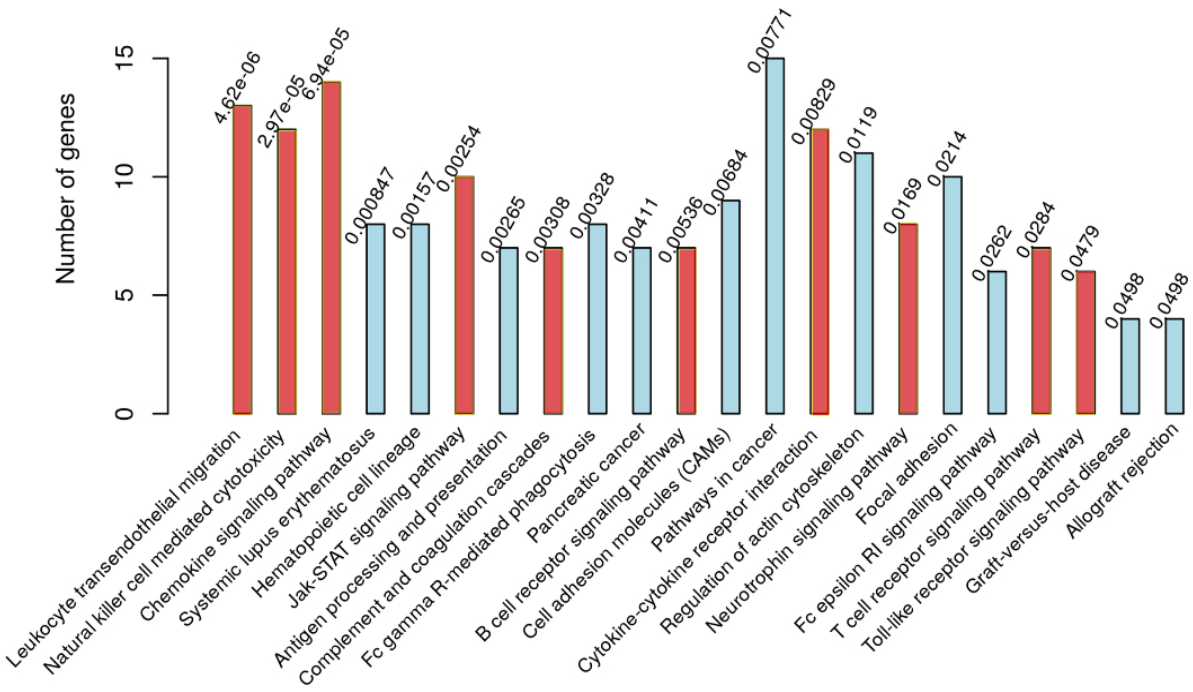

FIGURE 3.8. Results of the KEGG pathway enrichment done on all 148 central DEGs identified. Red bars shows the pathways taken for cross-talk analysis in this work (explained in chapter 4). The bottom of the bars show the KEGG pathway terms (p-values are reported on the top of the bars)

biological functions potentially represented by the identified 148 DEGs, we performed the GO enrichment analysis using DAVID annotation tool. Majority of the GO terms enriched are related to the immune response (Figure 3.7), again suggesting the influence of the immune response DEGs in the networks involved in prion disease progression.

As an infectious agent, PrP^{Sc} induces inflammatory responses by activating innate immunity through glial cells in the brain [34]. The presence of genes related to the microglial/astrocytic activation (eg; $Cd68, Emr1$) in the shared network central 148 DEGs, suggest that microglia not only proliferate and hypertrophy in prion disease [45] but are also positioned as network influential nodes in the related protein functional networks. Activation of endothelial cells through cytokines released from microglia and astrocytes would enhance migration of leukocytes from blood.

Presence of glial markers such as $Gfap$ and $P2ry$ receptors in addition to $Csf1$, in 148 shared central DEGs indicates the increment in the activities of astrocytes and leukocytes. Indeed, the pathway enrichment of these central DEGs highlighted the participation of

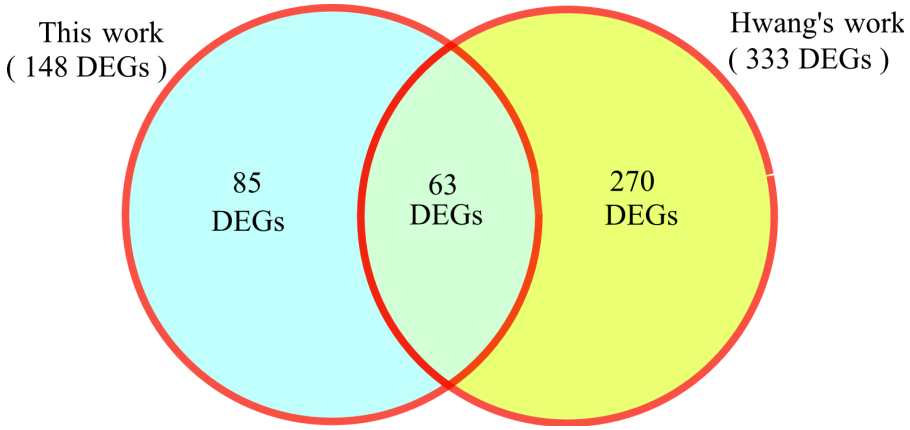

FIGURE 3.9. The 148 DEGs identified are essentially the DEGs which are present centrally in the protein functional interaction network related to the prion disease at some time point during the disease progression. The experiment performed in Hwang's work identified 333 DEGs, hypothesized to be involved in essential biological processes involved the the prion disease. The overlap of 63 DEGs can be observed in these studies, suggesting the importance of these 63 genes.

pathways such as leukocyte extravasation, cytokine-cytokine receptors and few others (Figure 3.8), which can be related directly or indirectly to the microglial/astrocytic activation.

3.4.2 Comparison with other works

Many DEGs have been identified in previous studies on the differential gene expression in the mice infected with prion [12, 41]. Out of 148 DEGs identified in this study, 63 DEGs overlap (see Figure 3.9) with the findings of Hwang et al.[28], suggesting that these DEGs are not only present at influential positions in the protein functional networks but also show correlated behavior in their expression patterns during prion disease progression. Many of these 63 DEGs are confined to the functions related to microglial activation and immune response, hypothesizing the central role of these functions in the disease progression. The central involvement of these genes in the disease network increases their potential disease involvement with respect to the ease of communication between them and the other network proteins. This study revealed many DEGs which were previously related to the pathogenesis of the prion disease. The novel disease related genes identified in this study should be considered important from the disease network's point of view. Identification of 85 DEGs which were not highlighted in the Hwang's study reflects the fact that these genes, although

didn't show correlation in their expression, is present at network central positions. The DEGs found to be highly correlating in their expression should be considered important as also the DEGs which are present centrally in the protein association network. Further, the overlapped 63 DEGs should be given more importance because of the fact that in addition to their behavior of high correlation, they are also present centrally in the disease protein functional networks.

3.4.3 Immunological response and prion disease

Neuropathological features common to all prion diseases in mammals can be divided into following four processes: prion replication (PcP_{Sc}) and accumulation [36], synaptic degeneration [29], microglia and astrocyte activation, and neuronal death. The inflammatory response in the prion disease occurs mostly via the activation of microglia. Microglia are the immune cells of the brain that are activated in response to infection or injury. Activation of T cells can lead to the activation of microglia. The activation of microglia results in the upregulation of proteins including complement factors, proteins of major histicompatibility complex, proinflammatory cytokines and interleukins [14]. Excessive and chronic activation of these factors produces oxidative stress, which leads to neurotoxicity and subsequently to neurodegeneration. Role of astrocyte activation during prion neuroinflammation has been reported in [38], in which astrocytes appear to be regulated by the interleukin-1 (IL-1) inducing astrocyte activation through CXCR3 ligand. Detectable changes in the expression of CXC ligands even in the asymptotic stages of the disease suggest that the chemokines might play a pivotal role in promoting neurodegeneration in prion diseases [16]. These findings support the role of microglial hyper-activation in neuro-degeneration related to prion disease.

The presence of DEGs related to the microglial activation, chemokines, cytokines, and other immunological modules, in the present set of identified 148 DEGs supports the hypothesis of the role of immunological system in neurodegeneration during prion disease. This study also highlights that these DEGs are present at the central positions in the disease network. The network influential positions of these DEGs make them possible drug targets as targeting these genes will have most influential effect on the disease network.

Table 3.7: List of 63 common genes also found in the works of Hwang et al.

Gene Names	Gene Names	Gene Names
A2M	CSF1R	LGALS3
ADCY7	CTSS	MPEG1
AIF1	CXCL10	MS4A6B
AQP4	EMR1	MS4A6D
ARHGDIB	FCGR3	MSN
B2M	GBP2	MT2
C1QA	GFAP	NCKAP1L
C1QB	H2-AA	OASL2
C1QC	H2-AB1	OSMR
C3	H2-D1	P2RY6
C3AR1	H2-K1	PLCE1
CCL9	IFIH1	PTGS1
CD14	IFIT2	PTPN6
CD44	IGTP	PTPRC
CD52	INPP5D	RHOJ
CD53	IRF8	SDC4
CD68	ITGB2	SERPING1
CLEC7A	ITGB5	SOCS3
CSF1	LAPTM5	STAT1
STAT3	THBS2	UCP2
TGFB1	TYROBP	VIM

List of genes identified in this work can be downloaded from `https://sites.google.com/a/iiitd.ac.in/khalique-research/research`. This site also provides the data corresponding to all the results outlined in this work.

Table 3.9: List of enriched pathways with corresponding genes

Pathway Names	Gene Names
Leukocyte transendothelial migration	ICAM1, NCF2, GNAI2, NCF4, ACTN1, CTNND1, ITGB2, VAV1, CXCL12, VCAM1, PLCG2, MSN, PIK3R1
Natural killer cell mediated cytotoxicity	H2-K1, ICAM1, PTPN6, PLCG2, H2-D1, ITGB2, FAS, VAV1, PIK3R1, LCP2, FCGR3, TYROBP
Chemokine signaling pathway	GNAI2, ADCY7, CCL9, CX3CL1, STAT1, VAV1, CXCL12, STAT3, CXCL10, GSK3B, CXCL16, CRK, PIK3R1, AKT2
Systemic lupus erythematosus	C1QA, C1QB, C3, H2-AA, ACTN1, C1QC, FCGR1, FCGR3
Hematopoietic cell lineage	CD44, CSF1, H2-AA, KITL, FCGR1, CD14, CSF2RA, CSF1R
Jak-STAT signaling pathway	IRF9, PTPN6, OSMR, SOCS3, JAK1, STAT1, PIK3R1, STAT3, CSF2RA, AKT2
Antigen processing and presentation	H2-K1, H2-D1, H2-AA, HSPA4, CTSS, CD74, B2M
Complement and coagulation cascades	C1QA, C1QB, C3AR1, A2M, C3, SERPING1, C1QC
Fc gamma R-mediated phagocytosis	PTPRC, PLCG2, INPP5D, CRK, VAV1, FCGR1, PIK3R1, AKT2
Pancreatic cancer	TRP53, JAK1, STAT1, PIK3R1, STAT3, TGFB1, AKT2
B cell receptor signaling pathway	PTPN6, GSK3B, PLCG2, INPP5D, VAV1, PIK3R1, AKT2
Cell adhesion molecules (CAMs)	H2-K1, VCAM1, ICAM1, PTPRC, ITGAV, H2-D1, H2-AA, ITGB2, SDC4
Pathways in cancer	TRP53, STAT1, KITL, STAT3, TGFB1, GSK3B, ITGAV, PLCG2, JAK1, FAS, CRK, CSF2RA, PIK3R1, AKT2, CSF1R
Cytokine-cytokine receptor interaction	OSMR, CXCL16, CSF1, CCL9, CX3CL1, FAS, CXCL12, KITL, CSF2RA, TGFB1, CSF1R, CXCL10
Regulation of actin cytoskeleton	ITGAV, CHRM1, ITGB5, ACTN1, NCKAP1L, ITGB2, MSN, CRK, VAV1, PIK3R1, CD14

Continued on next page

Table 3.9 – continued from previous page

Pathway Names	Gene Names
Neurotrophin signaling pathway	TRP53, GSK3B, PLCG2, CRK, RAPGEF1, PIK3R1, AKT2, ARHGDIB
Focal adhesion	ITGAV, GSK3B, ITGB5, ACTN1, CRK, VAV1, THBS2, RAPGEF1, PIK3R1, AKT2
Fc epsilon RI signaling pathway	PLCG2, INPP5D, VAV1, PIK3R1, LCP2, AKT2
T cell receptor signaling pathway	PTPN6, PTPRC, GSK3B, VAV1, PIK3R1, LCP2, AKT2
Toll-like receptor signaling pathway	IRF3, STAT1, PIK3R1, CD14, AKT2, CXCL10
Graft-versus-host disease	H2-K1, H2-D1, H2-AA, FAS
Allograft rejection	H2-K1, H2-D1, H2-AA, FAS

Table 3.8: 85 novel genes identified in this study

Gene Names	Citation	Gene Names	Citation	Gene Names	Citation
JAK1		CX3CL1		IRF3	
PLCG2		FCGR1		NCF4	
RGS4		NCF2		TRP53	
LPAR1		IRF9		SPARC	[12]
GRP	[24]	ANXA2		CD2AP	
ICAM1		CSF2RA		CDC27	
POLR2E		CHKA		CLASP2	
NCOA2		KITL		DAB2	[12]
B3GALT6		CHPT1		LPAR6	
CHRM1		DCX		CDC40	
PARP9		RAB13		CACNG2	
RSAD2		HSPA4		ACTN1	
CXCL16		PSME4		IGFBP3	
LCP2		GNAI2		RAB14	
TRHR	[24]	CXCL12		GSK3B	
NCOR1		PIK3R1		AKT2	
MYO1C		PCYT1B		ACE	
PLOD2		S1PR3	[24]	GALC	
CPSF2		ITGAV		USP15	
IRGM1		FAS		CKLF	
LY86	[12] [41]	VCAM1		S1PR2	
LYZ2		CTNND1		ATR	
CD74		USP14		TELO2	
COL9A3		VAV1		NUP54	
IRGM2		RAPGEF1		GBP3	
ESF1		DGKA		DNAJC6	
CRK		KCNMA1		GTPBP4	
RTP4		PLOD1			
USP18		PPM1A			

Some of the novel genes identified in this study have been reported in earlier studies, to be involved in prion disease.

CHAPTER 4

ROBUSTNESS AND CROSS-TALK

Universal and highly organized structures underlying the biological networks mediate different essential functions including robustness and evolvability. One of the structures employed are the bow-tie architectures which are prevalent in the biological systems. These architectures also provide predictable fragile points which can be exploited to understand disease pathogenesis. The robustness of bow-tie structure can be perceived from the fact that this structure provides a common gateway which facilitates control over fluctuating inputs and perturbations on different time and spatial scales.

This chapter discusses about the bow-tie structure within the biological pathways possibly involved in the prion disease progression. Section 4.1 gives the details about the procedure used to identify the potential genes involved in the cross-talk. Then, in section 4.2 we outline the proposed ODE model for the bow-tie structure extracted from the cross-talk study and also discuss different predictions of the model with implications on the prion disease progression and possible therapeutic targets. Finally, in section 4.3 we conclude about the implications of dysregulation of the bow-tie structure on the progression of prion disease.

4.1 Cross-talk candidate genes

In section 3.4.1, we identified the pathways which were enriched in the set of 148 shared genes. Figure 3.6 shows the list of pathways used for the cross-talk study. In section 4.1.1, we explain the procedure used to identify the protein functional networks(PFNs) related to the

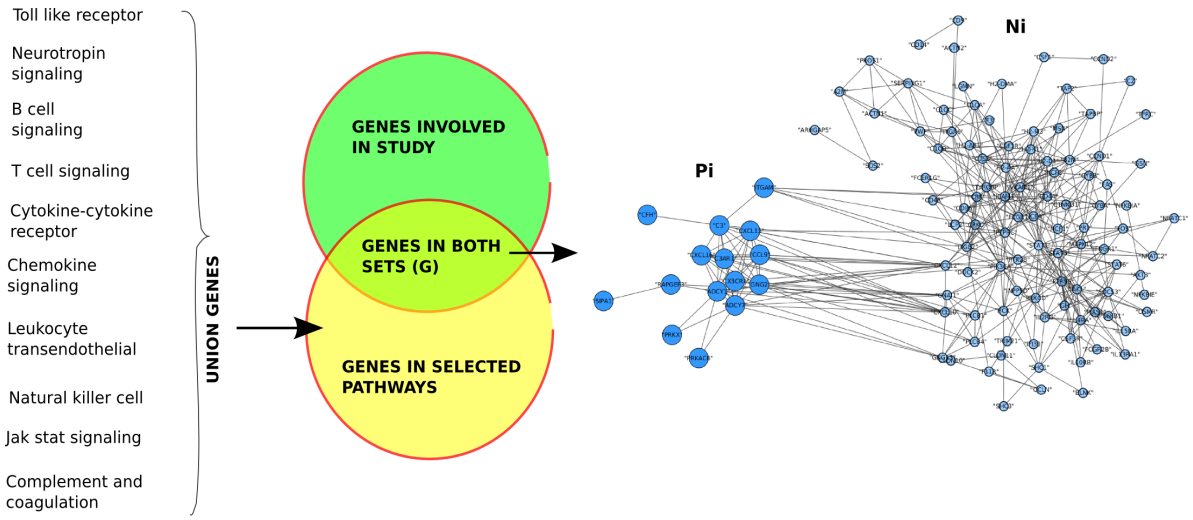

FIGURE 4.1. The left column shows the enriched pathways used. The genes involved in these pathways were pooled to get a set of genes involved in at-least one of these pathways. The time-stamped protein functional network ($Ni+Pi$) belonging to one of the mouse strain-prion strain combinations were used to extract the time-stamped PFN (Pi) related to these pathway related genes (G). This scheme was applied to every time-stamped PFNs ($Ni+Pi$) of a particular mouse strain-prion strain combination and the Pis obtained were further used to identify the genes possibly involved in the cross-talk between the pathways used.

genes involved in these pathways. In section 4.1.2, we outline the criterion used to rank the genes in the PFNs for the their possible involvement in cross-talk.

4.1.1 Extraction of PFNs

Figure 4.1 outlines the method used to extract the protein functional networks related to the genes corresponding to the identified pathways. We took genes involved in the identified pathways from the KEGG database and pooled all the genes together to get a set of genes involved in at least one of these pathways. These genes were then intersected with 13,822 genes used in this study. This was basically done to get the genes which are involved in the selected pathways and also present in our study. The time-stamped PFNs belonging to a particular mouse-prion combination were used to extract the time-stamped sub-graphs related to these genes. This procedure was repeated with every time-stamped PFNs of the

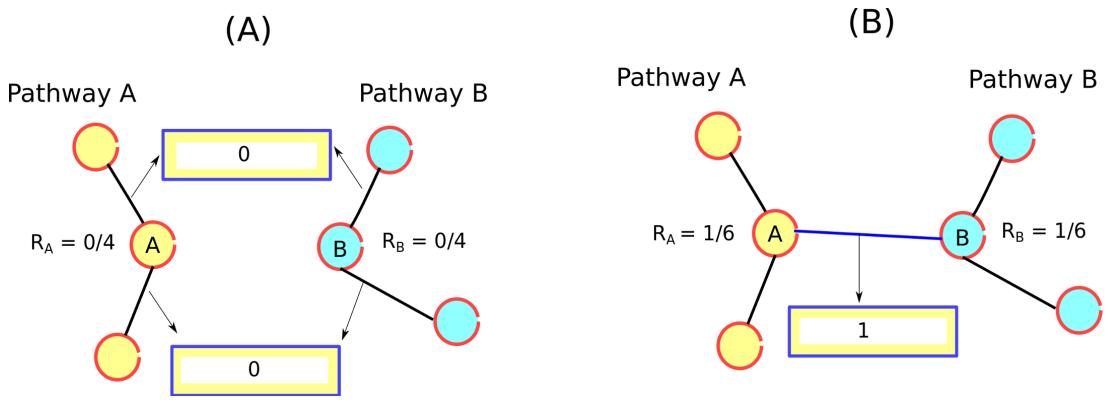

FIGURE 4.2. CTS of a node in a pathway reflects the potential of that node to participate in a cross-talk. For a particular protein, the relative cross-talk score is calculated as the ratio of inter-pathway edge's scores and the normalizing factor (n*(n-1)). where, n is the number of pathways considered in the study. In this example n is equal to 2. (A) Gene A and gene B are connected to the genes of their respective pathways. This results in relative cross-talk score for both protein/gene A and protein/gene B to be zero. (B) Gene A is connected to gene B present in different pathways. In this case, there is only one edge between pathway A and pathway B. The cross-talk score (M) corresponding to this edge is 1. Hence the relative cross-talk score is 1/6 for both the gene A and gene B.

selected mouse-prion model and the related sub-graphs obtained were then used to identify the genes possibly involved in the cross-talk.

4.1.2 Cross-talk ranking criterion

To determine if a gene is possibly involved in a cross-talk, an approach similar to the one used in [46] has been proposed with a modification in the computation of Cross-Talk Scores (CTSs). We clustered every protein in the time-stamped sub-graphs obtained via the procedure outlined in figure 4.1, into different pathways. For every protein in a particular time-stamped sub-graph, we then calculated the relative cross-talk scores. This gave us the relative cross-talk score for every gene at a particular time-stamp. These relative scores in different time-steps were then summed up, resulting in CTSs. The relative score corresponding to every gene i at time-stamp t, is calculated using the following formula.

$$R_i^t = \frac{\sum_{e \in E(i)} M_e}{|E(i)|(n(n-1))}$$

where, R_i^t is the relative score of gene i corresponding to the network at time-stamp t, M_e is the cross-talk score corresponding to the edge e of node i, $E(i)$ is the set edges adjacent on node i, and $n(n-1)$ is the normalizing factor, where n is the number of pathways.

As an example for the calculation of the relative cross-talk scores at every time-step, consider a protein G_A in a hypothetical pathway A which is only connected to other proteins of the same pathway, then its relative score will be 0 (see Figure 4.2A). But if this protein G_A connects with a protein G_B of another hypothetical pathway B then the cross-talk scores of both the proteins G_A and G_B will have a non-zero score (see Figure 4.2B). In general, if a gene/protein has more functional associations to connect with several other pathways in comparison to its own pathway, then it would be considered to have more possibility to participate in a cross-talk. Algorithm 1 outlines the procedure used to rank proteins according to their cross-talk scores. The initialization of the pseudo-code is as follows:

1. P = Set of protein functional sub-graphs corresponding to a particular mouse-prion model.
2. $Path(g)$ = Returns the set of pathways to which the gene g belongs.
3. $Score_p[g]$ = Stores the cross-talk score of gene g corresponding to a particular time-stamped sub-graph p in the set of protein functional networks P.
4. $Edge_p(g)$ = Stores the normalization factor of gene g for a particular time-stamp.
5. $CTS[g]$ = Stores the cross-talk scores of the gene g. Initialized to 0 for every gene.
6. Paths = Set of all pathways used in the procedure to identify the possible cross-talk proteins.

The identification procedure for cross-talk genes were performed on two mouse-prion combinations. B6.I-301V mouse-prion combination, which shows short life span (18 weeks), and B6.I-RML combination which shows longer life span (48 weeks). The variation in the life spans of these models may be the result of differing pathology of the prion disease. The overlapping high-scoring cross-talk genes in both the combinations should be considered important. This will imply that these genes, even though have been expressed via presumingly different pathologies, are ranked as top scoring candidates in cross-talk gene scoring criterion. Table 4.1 list the top 10 genes having high cross-talk scores corresponding to the combination B6I-RML. And Table 4.2 list the top 10 genes having high cross-talk scores corresponding to the combination B6I-301V.

Algorithm 2 Outline of the procedure used to identify the proteins possibly involved in the cross-talk.

for each time-stamped network $'p'$ in P **do**
  for each gene g in p **do**
    $Score_p[g] = 0$
    $Edge_p[g] = 0$
  end for
  for every edge e in p **do**
    $g1$ = node 1 of e
    $g2$ = node 2 of e
    for every pathway t in Paths **do**
      for every pathway k in Paths **do**
        if $t! = k$ **then**
          $Edge_p[g1] = Edge_p[g1] + 1$
          $Edge_p[g2] = Edge_p[g2] + 1$
        end if
        if $t! = k$ AND $t \in Path(g1)$ AND $k \in Path(g2)$ **then**
          $Score_p[g1] = Score_p[g1] + 1$
          $Score_p[g2] = Score_p[g2] + 1$
        end if
      end for
    end for
  end for
  for every gene g in p **do**
    $Score_p[g] = Score_p[g]/Edge_p[g]$
    $CTS[g] = Score_p[g] + CTS[g]$
  end for
end for

Table 4.1: Top 10 high cross-talk scoring genes (B6I-RML)

Gene	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	CTSs
PIK3CA	0.00	0.00	0.15	0.00	0.00	0.15	0.00	0.13	0.12	0.15	0.70
IFNAR2	0.00	0.00	0.05	0.09	0.03	0.09	0.08	0.09	0.08	0.11	0.62
IKBKG	0.17	0.00	0.14	0.00	0.07	0.06	0.10	0.00	0.00	0.04	0.58
VAV3	0.00	0.18	0.00	0.00	0.00	0.00	0.00	0.11	0.13	0.14	0.57
NFKBIA	0.11	0.14	0.15	0.00	0.00	0.00	0.09	0.05	0.00	0.00	0.55
AKT2	0.00	0.21	0.22	0.11	0.00	0.00	0.00	0.00	0.00	0.00	0.54
GSK3B	0.14	0.12	0.00	0.11	0.00	0.12	0.00	0.00	0.00	0.00	0.50
PIK3CD	0.19	0.00	0.00	0.00	0.11	0.19	0.00	0.00	0.00	0.00	0.48
PTPN11	0.00	0.10	0.00	0.09	0.00	0.12	0.00	0.00	0.00	0.13	0.44
PTPN6	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.07	0.08	0.09	0.34

Table 4.2: Top 10 high cross-talk scoring genes (B6I-301V)

Gene	T1	T2	T3	T4	T5	T6	T7	T8	T9	CTSs
MAP2K2	0.31	0.00	0.00	0.00	0.39	0.00	0.00	0.00	0.00	0.70
AKT3	0.00	0.00	0.23	0.00	0.00	0.00	0.00	0.15	0.24	0.62
NFKBIA	0.00	0.00	0.11	0.11	0.10	0.06	0.06	0.00	0.11	0.56
IFNAR2	0.09	0.00	0.06	0.06	0.09	0.07	0.07	0.09	0.00	0.53
NFATC1	0.06	0.07	0.00	0.07	0.12	0.00	0.06	0.06	0.07	0.51
PIK3CD	0.15	0.00	0.18	0.00	0.18	0.00	0.00	0.00	0.00	0.51
GSK3B	0.11	0.16	0.00	0.09	0.10	0.00	0.00	0.00	0.00	0.46
SOS1	0.00	0.17	0.00	0.00	0.20	0.08	0.00	0.00	0.00	0.45
NFATC2	0.00	0.08	0.00	0.00	0.00	0.07	0.05	0.07	0.08	0.35
PIK3R1	0.00	0.15	0.18	0.00	0.00	0.00	0.00	0.00	0.00	0.33

We mapped the high-scoring genes to KEGG pathway database. The KEGG database provides the information about the biological pathways to which the genes belong, along with the information about the kind of interaction present between different genes within a pathway. Mapping the genes to this database identified a bow-tie structure, with some of the high cross-talk scoring genes as the core of bow-tie. Figure 4.3 shows the sub-graph extracted from KEGG database. The genes in the identified bow-tie structure are part of the time-stamped PFNs extracted corresponding to the pathways considered in this study.

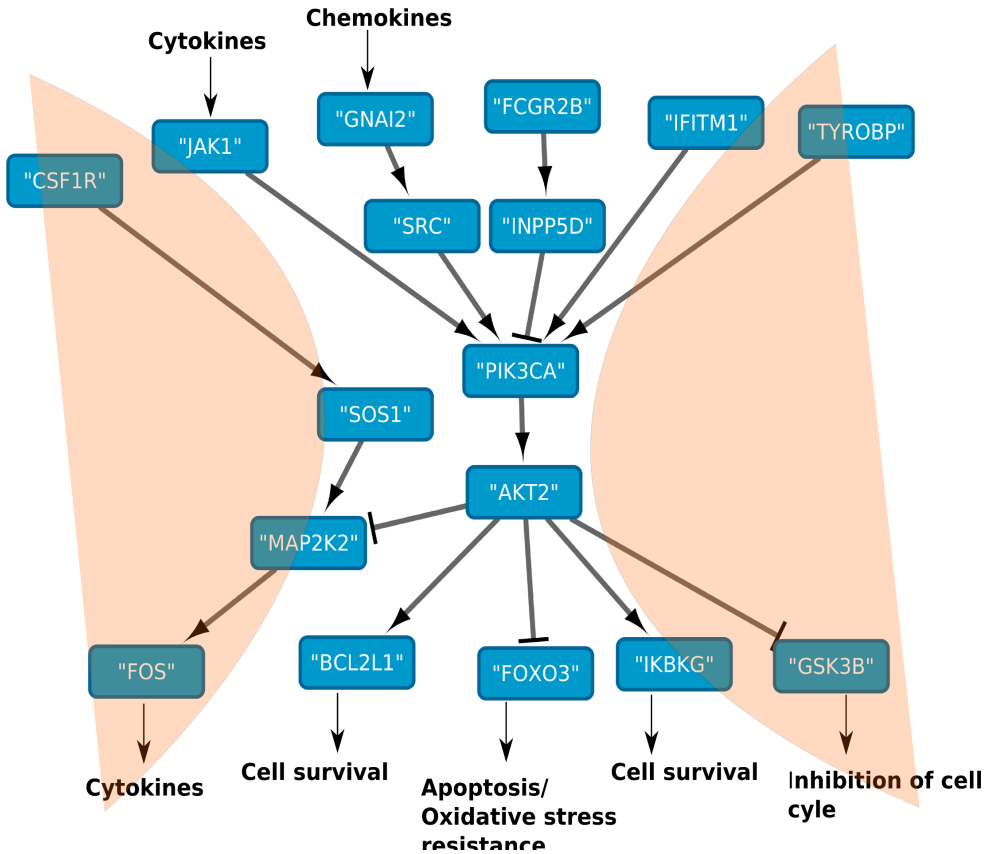

FIGURE 4.3. The bow-tie structure identified from the protein functional networks. The pointed arrows represent the activation action by source onto the target gene and the flattened end arrows suggest inhibitory action.

4.2 Modeling the identified network structure

4.2.1 Model derivation and assumptions

The total amount of active component X_n produced when the cell is exposed to a given signal input can be represented as $\bar{X}_n = \int_0^\infty X_n(t)dt$.

The assumption taken here for the modeling is that the signaling cascades are weakly activated and hence can be modeled as linear systems [17, 26] (In the presented model,there is exception to the inactivation action which has been modeled as having a non-linear behavior). Let X_{i-1} and X_i denote the active concentrations of the upstream and present component respectively. Also, let $\widehat{X}_i$ represent the inactive concentration of the present component. Let α_i and β_i denote the rate constants for component i activation and inactivation respectively. Then the dynamics of the signaling cascade can be modeled as follows:

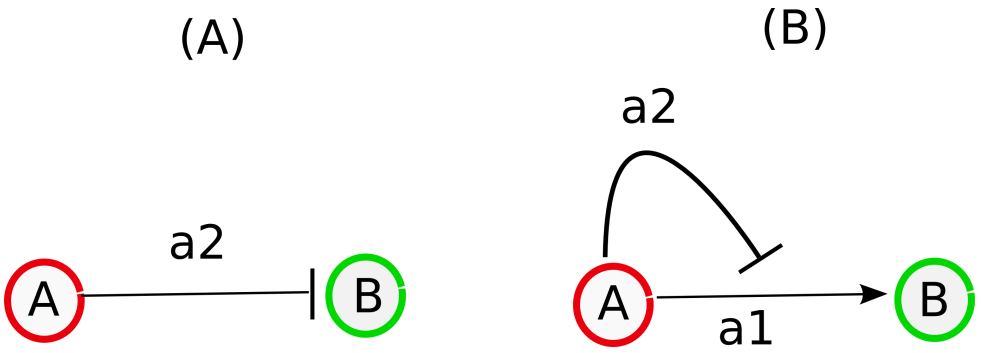

FIGURE 4.4. Gene A inactivates gene B. (A) The actual inhibition regulation network structure obtained from the KEGG database. (B) Assumed regulation network structure. The inactivation of gene B occurs only when its concentration cross some predefined threshold. a1 and a2 are the kinetic constants for activation and inactivation respectively.

$$\frac{dX_i}{dt} = \alpha_i X_{i-1} \widehat{X}_i - \beta_i X_i$$

With the assumption that the component i remains constant, ie. $X_i + \widehat{X}_i = X_{total}$, the above equation can be rewritten as

$$\frac{dX_i}{dt} = \widehat{\alpha}_i X_{i-1}(1 - \frac{X_i}{X_{total}}) - \beta_i X_i$$

where, $\widehat{\alpha_i} = \alpha_i X_{total}$

Again assuming that the signaling cascades are weakly activated, $X_i << \widehat{X}_i$, equation 2 can be modeled as linear ODE.

$$\frac{dX_i}{dt} = \widehat{\alpha}_i X_{i-1} - \beta_i X_i$$

The above explained assumption has been utilized to model the activation components of the model. For the negative regulation a different model has been assumed. It is assumed that the inactivation of genes by other regulating genes occurs when the activated concentration of the regulated gene goes above a predefined threshold. Figure 4.4 shows the assumption used in modeling the inactivation phenomenon. here, we assume that there is always some production of the inhibited component. The inactivation occurs only when a certain concentration threshold level of the regulated gene is crossed.

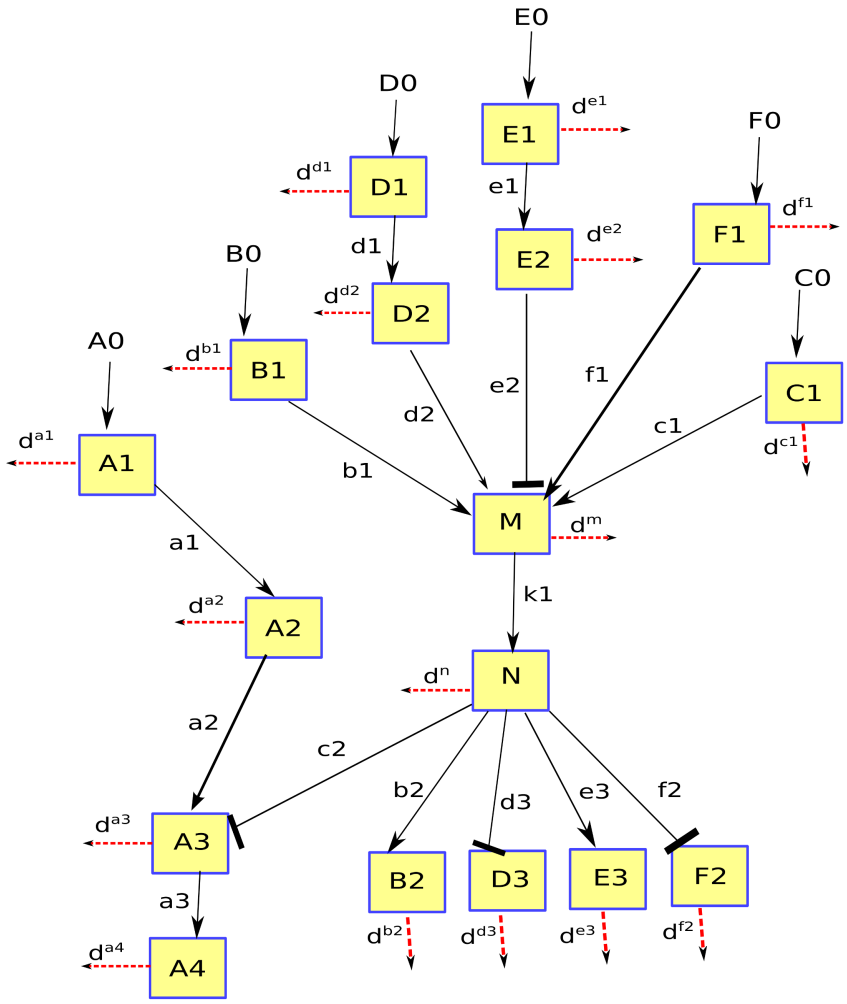

FIGURE 4.5. Schematic representation of the structure used for ODE modeling.

4.2.2 Bow-tie ODE model

The genes involved in the immunological pathways identified in this study have been found to be part of different signaling cascades with common components. The common members provides a gateway for different signaling cascades to cross-talk with each other. To understand and predict the behavior of these interacting signaling cascades, we have proposed an ODE model of the system. The directed network to be modeled has been taken from the KEGG database. The directed edges signify the activation/inhibition of the downstream components in the signaling cascade. The pointed arrow signifies the activation and the blunt ends of the arrows signify the inhibitory action (see Figure 4.3). Following equations define the model.

$$\frac{dA_1}{dt} = A_0 - d^{a1} A_1$$

$$\frac{dA_2}{dt} = a1 * A_1 - d^{a2} A_2$$

$$\frac{dA_3}{dt} = a2 * A_2 - c2 * N * A_3 + c\_2 * N - d^{a3} A_3$$

$$\frac{dA_4}{dt} = a3 * A_3 - d^{a4} A_4$$

$$\frac{dB_1}{dt} = B_0 - d^{b1} B_1$$

$$\frac{dB_2}{dt} = b2 * N - d^{b2} B_2$$

$$\frac{dC_1}{dt} = C_0 - d^{c1} C_1$$

$$\frac{dD_1}{dt} = D_0 - d^{d1} D_1$$

$$\frac{dD_2}{dt} = d1 * D_1 - d^{d2} D_2$$

$$\frac{dD_3}{dt} = -d3 * N * D_3 + d\_3 * N - d^{d3} D_3$$

$$\frac{dE_1}{dt} = E_0 - d^{e1} E_1$$

$$\frac{dE_2}{dt} = e1 * E_1 - d^{e2} E_2$$

$$\frac{dE_3}{dt} = e3 * N - d^{e3} E_3$$

$$\frac{dF_1}{dt} = F_0 - d^{f1} F_1$$

$$\frac{dF_2}{dt} = -f2 * N * F_2 + f\_2 * N - d^{f2} F_2$$

$$\frac{dN}{dt} = k1 * M - d^{n} N$$

$$\frac{dM}{dt} = b1 * B_1 + c1 * C_1 + d2 * D_2 - e2 * E_2 * M + e\_2 * E_2 + f1 * F_1 - d^{m} M$$

The basic structure of the derived network has been outlined in Figure 4.3(B). The schematic of the signaling network shows that this network is composed of six pathways, A, B, C, D, E, F. Each pathway in the network has a unique input and a unique output. For example, In pathway D, the input is designated as D_0, and can be taken to represent both the signal itself and its receptor. The activated species of the downstream elements of pathway A is designated A_1, A_2, A_3 and A_4. Similarly, activated components of other pathways are designated. M and N represents the common elements in the network structure. When any of the input signals is present, then it activates component M which in-turn activates N and ultimately N leads to the activation of other downstream components. a_1,b_1,c_1,d_1, e_1 and f_1 represents the upstream rate constants of the structure. And a_2, b_2, c_2, d_2, e_2 and f_2 represents the downstream rate constants of the structure. d^{a1}, d^{a2}, d^{a3}, d^{a4}, d^{b1}, d^{b2}, d^{c1}, d^{d1},d^{d2}, d^{d3}, d^{e1}, d^{e2}, d^{e3}, d^{f1}, d^{f2}, d^{m}, and d^{n} represent the decay rate constant of the activated components A_1, A_2, A_3, A_4, B_1, B_2, C_1, D_1, D_2, D_3, E_1, E_2, E_3, F_1, F_2, M, and N respectively. The constants e_2, d_3, c_2, f_2 represents the activation kinetic constants for basal production of M, D_3, A_3, and F_2 respectively.

Table 4.3: Units and values of the parameters used

Parameter	Units	Value	Parameter	Units	Value
d^{a1}	S^{-1}	1	d^{a4}	S^{-1}	1
d^{b1}	S^{-1}	1	d^{e1}	S^{-1}	1
d^{a2}	S^{-1}	1	d^{e2}	S^{-1}	1
d^{b2}	S^{-1}	1	d^{e3}	S^{-1}	1
d^{c1}	S^{-1}	1	d^{f1}	S^{-1}	1
d^{a3}	S^{-1}	1	d^{f2}	S^{-1}	1
d^{d1}	S^{-1}	1	d^{m}	S^{-1}	1
d^{d2}	S^{-1}	1	d^{n}	S^{-1}	1
d^{d3}	S^{-1}	1	a_1	S^{-1}	1
e_2	$S^{-1}nM^{-1}$	1	c_2	S^{-1}	0.1
f_2	S^{-1}	0.1	d_3	S^{-1}	0.1
f_2	$S^{-1}nM^{-1}$	1.8	f_1	S^{-1}	1
k_1	S^{-1}	1	d_1	S^{-1}	1
a_2	S^{-1}	1	d_2	S^{-1}	1
a_3	S^{-1}	1	d_3	$S^{-1}nM^{-1}$	1
b_1	S^{-1}	1	e_1	S^{-1}	1
b_2	S^{-1}	1.5	e_2	S^{-1}	0.1
c_1	S^{-1}	1	e_3	S^{-1}	1
c_2	$S^{-1}nM^{-1}$	1			

Parameter values have been assigned arbitrarily.

4.2.3 Results

The proposed model can be used to study the behavior of this network system in disease condition. The relevance of this network system can be realized by the fact that previous studies have highlighted the involvement of PI3k/AKT pathway in prion disease. Dysregulation of $PI3K\text{-}AKT$ pathway in prion disease, mainly due to down-regulation of $PI3Ks$ and $AKTs$, has been reported in [40]. In this work the kinetic constants of the model are arbitrary and the results should be interpreted accordingly.

Figure 4.6 shows the simulation results corresponding to the diseased condition of the modeled network. The outputs show the behavior of some of the network components when the input signals corresponding to the diseased state is applied. In the modeled structure, the input components comprises of $CSF1R$, $JAK1$, $GNAI2$, $FCGR2B$, $IFITM1$,

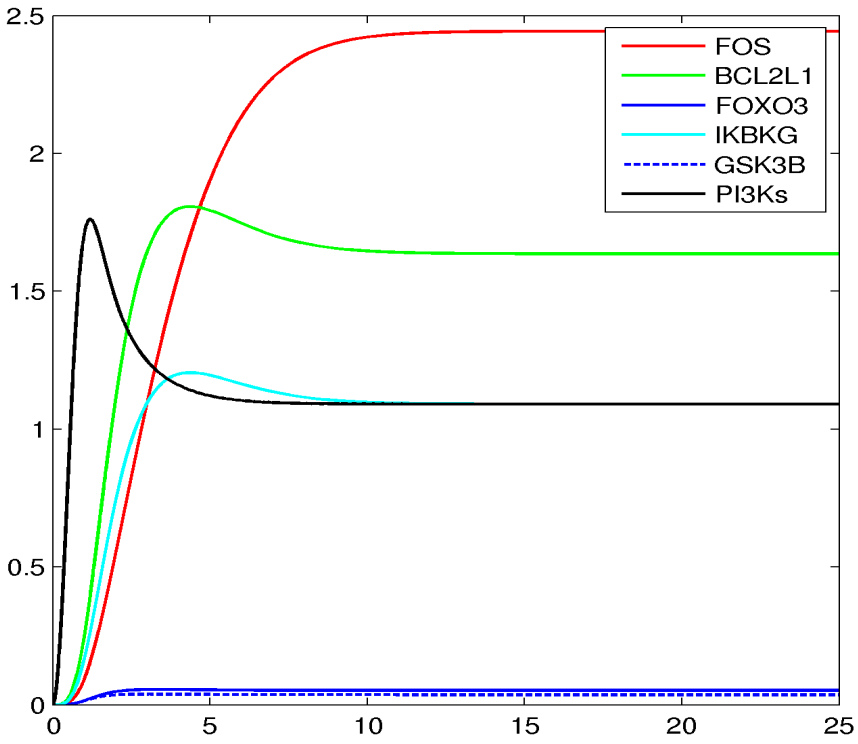

FIGURE 4.6. Activation response of various output components of the structure in the diseased condition. The input signals taken here are as follows: A_0=5, B_0=1, C_0=5, D_0=1, E_0=10, F_0=4. These values are arbitrary and are partially guided by the differential expression patterns of the input genes in the diseased condition.

and $TYROBP$. The input signals related to the diseased condition has been taken from the differential gene expression pattern shown by the corresponding genes of the input components in the network. High arbitrary input was given to the network input component whose corresponding gene expression pattern was upregulated, and comparatively low input signal was given to the component whose corresponding gene expression pattern was relatively downregulated. The simulation shows various output signals behavior in response to the applied input signals. The model predicts high activity of FOS gene, which is the end component in the $ERK1/2$ pathway. Also, there is comparatively low activity of the $BCL2L1$ and $IKBKG$ genes which has been shown to take part in cell survival [6, 25]. It has been reported in works of Alessandro et al. [22], that the prion replication leads to

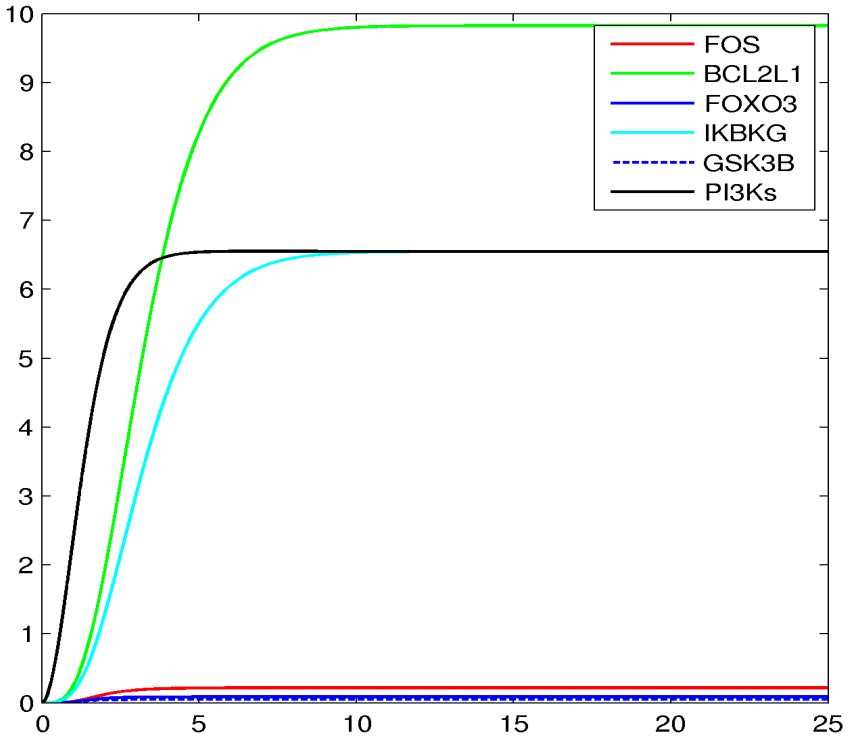

FIGURE 4.7. Activation response of the output components in the normal condition. The input signals taken here are as follows: A_0=1, B_0=5, C_0=1, D_0=5, E_0=1, F_0=2. These values are arbitrary and are partially guided by the differential expression patterns of the input genes in the diseased condition.

hyper-activation of $ERK1/2$ pathway. Activation of the $PI3Ks$ are also decreased in the diseased condition. As evident from the network structure, the $PI3K$-AKT components inhibit the $ERK1/2$ pathway via a possible cross-talk. This cross-talk behavior is inhibited in the prion disease condition, possibly leading to hyper-activation of the $ERK1/2$ pathway. These predictions suggest that the onset of prion disease is related to the dysfunction of these interacting pathways. Hence, we can hypothesize that the dysregulation of the core components (ie; $PI3Ks$ and $AKTs$) of this structure results in the possible dysregulation of other interacting downward components. Figure 4.7 shows the behavior of the same network components in response to the input signals in the normal condition. It predicts that the signals to the genes $IKGKB$ and $BCL2L1$ are high, predicting high activity of the cell survival functions. Also, the response of FOS components is low, indicating a possible

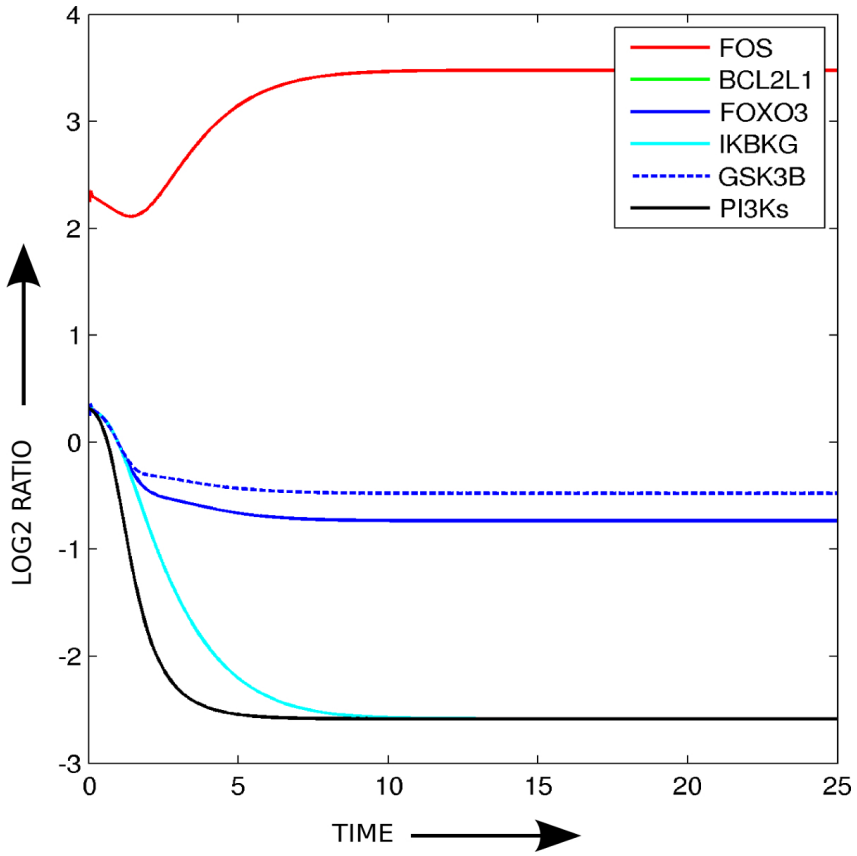

FIGURE 4.8. Differential expression pattern predicted via proposed ODE model for the identified structure. This depicts the expression pattern in the prevalence of prion disease condition.

cross-talk between the $ERK1/2$ and $PI3K\text{-}AKT$ pathways. Figure 4.8 shows the differential expression of the output components. The differential expression is captured via log_2 ratio of the diseased and normal outputs. The plots clearly indicates the downregulation of $PI3Ks$ and other components. It also shows the predicted upregulation of the FOS component.

4.3 Discussion

Biological systems maintain homeostasis in the presence of varied perturbations due to environmental changes. It has been proposed that the robustness in biological systems are inherent property and that it is necessary for the evolution of the system in the changing environment. It provides a shield to the normal functioning of the system against various perturbations, including pathogens. The robustness as a property is the effect of the or-

ganization of the biological systems as complex networks. The structure presumed gives the stability against the fluctuations in the input of the system. This bow-tie structure facilitates heterogeneity that allows for the robust regulation of the biological system. This robust design as a bow-tie structure is also inherently susceptible to fragility [21]. The core elements of the structure can be targeted for the dysregulation of the whole system. Hence, although the bow-tie architecture provides the buffer for the different input signal intensities providing stability, it is also prone to system failure due to possible attacks on the core elements. Disease as a perturbation is not likely to disrupt such a system unless it results in the dysfunction of the core elements. Identifying the structure being affected by the onset of a particular disease gives an opportunity to better understand and develop effective therapeutic strategies.

$PI3K$ (phosphoinositide-3 kinase) has important functions in the immune system regulation. Receptors such as CD28 on T cells and CD19 on B cells activates the $PI3Ks$. The products of PI3Ks, namely phosphatidylinositol3monophosphate ($PIP3$) and others govern many cellular events including cell growth and survival. The activation of PI3K also downregulates the immune activity related to both innate and adaptive immune system [32], implying the crucial role of PI3ks in controlling the pro-inflammatory activity. The core elements of the bow-tie structure identified in our study comprises of $PI3Ks$ and AKT. The participating pathways in this structure can be identified to be mostly belonging to innate and adaptive immune system. The pathways including Jak-stat signaling, B cell signaling, Natural killer cell cyto-toxicity, and Neurotropin signaling pathways converge at PI3Ks. The PI3K-AKT pathway then in turn regulates several pathways including $NF\text{-}\kappa\text{-}B$, $ERK\text{-}1/2$, and $GSK\text{-}3$. This structure resembles bow-tie which relies on the core elements to regulate any fluctuations in the input signals. Relative downregulation of PI3Ks and AKT in prion disease shows the effect of the disease on the core of this bow-tie structure, leading to the dysregulation of the output pathways. The $PI3K\text{-}AKT$ pathway regulates many biological functions including apoptosis, Erk1/2 signaling, Foxo signaling, and NF-kB signaling. Some of these pathways has been reported to be dysregulated in prion disease. In [22], the phosphorylation levels of Src, MEK 1/2 and ERK 1/2 signaling molecules, both before and after prion infection were assessed with the conclusion that prion replication leads to a hyperactivation of Erk1/2 pathway. In [40], the possibility of dysregulation of $PI3K/AKT/GSK\text{-}3$ pathway was explored. It was found that the prion strain altered PI3K-mediated signaling, evidenced by AKT inhibition and GSK-3 activation, were the common features of both in culture and in-vivo prion mediated neurotoxicity.

CHAPTER 5

CONCLUSION AND FUTURE WORK

In this study, we have carried out network analysis based on a recent study of prion disease in mouse [28] in order to gain insights into disease progression. Time-specific differentially expressed genes were used to construct time-specific protein functional networks corresponding to 6 different mouse strain-prion strain combinations. The main aim of the study was to identify the core functional modules related to the progression of prion disease. We extracted protein functional networks consisting of genes differentially expressed due to prion disease infection and analyzed network topology and dynamics to identify key network structures in disease progression.

In particular, we identified 148 genes in this study, which have been found to be mainly related to immunological system. These genes have also been found to show higher network centralities corresponding to the time-specific networks related to prion disease progression. This suggests, that the proteins related to the immunological system is present in the core of the disease related protein functional networks. Another inference which can be made here is that these identified core genes may have potential regulating effect on the other genes present in the disease related protein networks which are differentially expressed in several mouse prion models.

We performed a cross-talk analysis method to identify potential genes participating in cross-talk during disease progression. The highly ranked genes according to this criterion were mapped to KEGG database to identify the network being dysregulated. This cross-talk analysis of the time-stamped protein functional networks revealed a bow-tie structure, potentially dysregulated in the prion disease condition. We proposed a mathematical

model for the identified bow-tie structure. Numerical simulations of the modeled bow-tie structure showed that dysregulation of $PI3K$-AKT pathway during prion disease leads to dysregulation of many other pathways, supporting the role of pro-inflammatory effects in neurodegeneration. We hypothesize that this dysregulation may be the cause of sustained immune response observed in prion disease, which could ultimately lead to prominent neuronal cell death and the manifestation of clinical symptom.

APPENDIX A

Matlab code for simulating the bow-tie architecture identified in this study. The script files can be downloaded from: `https://sites.google.com/a/iiitd.ac.in/khalique-research/codes`

```
1
2  %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
3  % Function file simulating the ODE model corresponding to the
4  % bow-tie architecture identified.
5  %
6  %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
7
8
9  function f = bowTie(t,x,Fo,Eo,Ao,Bo,Co,Do,da1, a1, da2, a2, da3,...
10             a3,da4,k1,dn,b1,d2,c1,f1,e2,db1,b2,db2,dc1,c2,de3,df1,...
11             f2,df2,dd1,d1,dd2,d3,dd3,de1,de2,e1,e3,dm,c_2,f_2,e_2,d_3)
12
13
14    f(1) = Ao - da1 * x(1); % A1 = x(1)
15    f(2) = a1 * x(1) - da2 * x(2); % x(2) = A2,  x(3)=A3
16    f(3) = a2 * x(2) - c2 * x(4) * x(3) - da3 * x(3) + c_2 * x(4);
17    f(6) = a3 * x(3) - da4*x(6); % x(6) = A4
18
19    f(4) = k1 * x(17) - dn * x(4); % x(4) = N
20    f(17) = b1 * x(7) + d2 * x(10) + c1 * x(8) + f1 * x(15) - e2 * x(13)...
```

```
21            * x(17) - dm * x(17) + e_2 * x(13); % x(17) = M
22
23    f(7) = Bo - db1*x(7); % x(7) = B1
24    f(5) = b2 * x(4) - db2 * x(5); % x(5) = B2
25
26
27
28    f(8) = Co - dc1 * x(8); % x(8) = C1
29
30
31    f(9) = Do - dd1 * x(9); % x(9) = D1
32    f(10) = d1 * x(9) - dd2 * x(10); % x(10) = D2
33    f(11) = -d3 * x(4) * x(11) - dd3 * x(11) + d_3 * x(4); % x(11) = D3
34
35
36    f(12) = Eo - de1 * x(12); % x(12) = E1
37    f(13) = e1 * x(12) - de2 * x(13); % x(13) = E2
38    f(14) = e3 * x(4) - de3 * x(14); % x(14) = E3
39
40    f(15) = Fo - df1 * x(15); % x(15) = F1
41    f(16) = -f2 * x(4) * x(16) - df2 * x(16) + f_2 * x(4); % x(16) = F2
42
43
44    f = f';
```

```
1  %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
2  % This script simulates and plots the effect on the output signals of
3  % change in the concentration of input elements of the
4  % bow-tie-architecture modeled.
5  %
6  %
7  %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
8
9
10
11 clear all;
12 clc;
13
14 t = 0:0.0001:25;
15
16
```

```
Ao = 5; % Input to component gene CSF1R
Bo = 1; % Input to component gene JAK1
Co = 5; % Input to component gene TYROBP
Do = 1; % Input to component gene GNAI2
Eo = 10; % Input to component gene FCGR2B
Fo = 4; % Input to component gene IFITM1

%% Parameters definition

da1 = 1; a1 = 1; da2 = 1; a2 = 1; da3 = 1;
a3 = 1; a4 = 1; da4 = 1; k1 = 1; dn = 1;
b1 = 1; d2 = 1; c1 = 1; f1 = 1; e2 = 1;
db1 = 1; b2 = 1.5; db2 = 1; dc1 = 1; c2 = 1;
dd1 = 1; d1 = 1; dd2 = 1; d3 = 1;
dd3 = 1; de1 = 1; de2 = 1; e1 = 1; e3 =1;
de3 = 1; df1 = 1; f2 = 1.8; df2 = 1; dm = 1;
c_2 = 0.1; e_2 = 0.1; f_2 = 0.1; d_3 = 0.1;



%% Numerical simulation and plotting the time series of the outputs

[time x] = ode45(@(t,x) bowTie(t,x,Fo,Eo,Ao,Bo,Co,Do,da1, a1, da2,...
          a2, da3,a3,da4,k1,dn,b1,d2,c1,f1,e2,db1,b2,db2,dc1,c2,de3,...
            df1,f2,df2,dd1,d1,dd2,d3,dd3,de1,de2,e1,e3,dm,c_2,f_2,...
           e_2,d_3), t, [0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]);

figure();

plot(time,x(:,6),'r'); % Output of component gene FOS
hold on;
plot(time,x(:,5),'g'); % Output of component gene BCL2L1
plot(time,x(:,11),'b'); % Output of component gene FOXO3
plot(time,x(:,14),'c'); % Output of component gene IKKs
plot(time,x(:,16),'--'); % Output of component gene GSK3B
plot(time,x(:,17), 'k'); % Output of component gene PI3Ks

legend('FOS','BCL2L1','FOXO3','IKBKG','GSK3B','PI3Ks')
```

BIBLIOGRAPHY

[1] R. ALBERT AND A. BARABASI, *Statistical mechanics of complex networks*, Reviews of Modern Physics, 74 (2002), pp. 47–97.

[2] R. ALBERT, H. JEONG, AND A. BARABASI, *Error and attack tolerance of complex networks*, Nature, 406 (2000), pp. 378–482.

[3] B. ALBERTS, A. JOHNSON, J. LEWIS, M. RAFF, K. ROBERTS, AND P. WALTER, *Molecular Biology of the Cell 5TH edition*, Garland Science, 2007.

[4] S. AMOR, L. A. N. PEFEROEN, D. Y. S. VOGEL, M. BREUR, P. V. D. VALK, AND D. BAKER, *Inflammation in neurodegenerative diseases - an update*, Immunology, 142 (2014), pp. 151–166.

[5] S. B. PRUSINER, *Cell biology. a unifying role for prions in neurodegenerative diseases*, Science, 336 (2012), pp. 1511–1513.

[6] A. S. BALDWIN, *Regulation of cell death and autophagy by ikk and nf-kb: critical mechanisms in immune function and cancer*, Immunological Reviews, 246 (2012), pp. 327–345.

[7] A. BARABASI, *Scale-free networks: A decade and beyond*, Science, 325 (2009), pp. 412–413.

[8] A. BARABASI AND E. BONABEAU, *Scale-free networks*, Scientific American, 288 (2003), pp. 50–59.

[9] A. BARABASI, N. GULBAHCE, AND J. LOSCALZO, *Network medicine: a network-based approach to human disease*, Nature Reviews Genetics, 12 (2011), pp. 56–68.

[10] A. BARABASI AND Z. N. OLTVAI, *Network biology: understanding the cell's functional organization*, Nature Reviews Genetics, 5 (2004), pp. 101–113.

[11] D. BOCHE, V. H. PERRY, AND J. A. R. NICOLL, *Review: Activation patterns of microglia and their identification in the human brain*, Neuropathology and Applied Neurobiology, 39 (2013), pp. 3–18.

[12] S. BOOTH, C. BOWMAN, R. BAUMGARTNER, G. SORENSEN, C. ROBERTSON, M. COULTHART, C. PHILLIPSON, AND R. L. SOMORJAI, *Identification of central nervous system genes involved in the host response to the scrapie agent during preclinical and clinical infection*, J Gen Virol, 85(11) (2004), pp. 3459–3471.

[13] N. F. BORGES, H. ERANA, S. R. ELEZGARAI, C. HARRATHI, M. GAYOSSO, AND J. CASTILLA, *Infectivity versus seeding in neurodegenerative diseases sharing a prion-like mechanism*, International Journal of Cell Biology, 2013 (2013).

[14] D. R. BROWN, *Microglia and prion disease*, Microscopy Research and Technique, 54 (2001), pp. 71–80.

[15] H. BUELER, A. AGUZZI, A. SAILER, R. GREINER, P. AUTENRIED, M. AGUET, AND C. WEISSMANN, *Mice devoid of prp are resistant to scrapie*, Cell, 73 (1993), pp. 1339–1347.

[16] M. BURWINKEL, C. RIEMER, A. SCHWARZ, J. SCHULTZ, S. NEIDHOLD, T. BAMME, AND M. BAIER, *Role of cytokines and chemokines in prion infections of the central nervous system*, International Journal of Developmental Neuroscience, 22 (2004), pp. 497–505.

[17] M. CHAVES, E. D. SONTAG, AND R. J. DINERSTEIN, *Optimal length and signal amplification in weakly activated signal transduction cascades*, J. Phys. Chem. B, 108 (2004), pp. 15311–15320.

[18] C. T. CHIEN, P. L. BARTEL, R. STERNGLANZ, AND S. FIELDS, *The two hybrid system: A method to identify and clone genes for proteins that interact with a protein of interest*, Proc. Nati. Acad. Sci., 88 (1991), pp. 9578–9582.

[19] D. W. COLBY AND S. B. PRUSINER, *Prions*, Cold Spring Harbor Perspectives in Biology, 3 (2011).

[20] P. CSERMELY, T. KORCSMAROS, H. KISS, G. LONDON, AND R. NUSSINOV, *Structure and dynamics of molecular networks:a novel paradigm of drug discovery a comprehensive review*, Pharmacology and Therapeutics, 138 (2013), pp. 333–408.

[21] M. CSETE1 AND J. DOYLE, *Bow ties, metabolism and disease*, TRENDS in Biotechnology, 22 (2004), pp. 446–450.

[22] A. DIDONNA AND G. LEGNAME, *Aberrant erk 1/2 complex activation and localization in scrapie-infected gt1-1 cells*, Molecular Neurodegeneration, 5 (2010).

[23] N. GEHLENBORG, D. HWANG, I. Y. LEE, H. YOO, D. BAXTER, B. PETRITIS, R. PITSTICK, B. MARZOLF, S. J. DEARMOND, G. A. CARLSON, AND L. HOOD, *The prion disease database: a comprehensive transcriptome resource for systems biology research in prion diseases*, Database: The Journal of Biological Databases and Curation, 11 (2009).

[24] A. G. GOSSNER AND J. HOPKINS, *Transcriptome analysis of cns immediately before and after the detection of prpsc in ssbp/1 sheep scrapie*, Veterinary Microbiology, 173 (2014), pp. 201–207.

[25] J. M. HARDER, Q. DING, K. A. FERNANDES, J. D. CHERRY, L. GAN, AND R. T. LIBBY, *Bcl2l1 (bcl-x) promotes survival of adult and developing retinal ganglion cells*, Molecular and Cellular Neuroscience, 51 (2012), pp. 53–59.

[26] R. HEINRICH, B. G. NEEL, AND T. A. RAPOPORT, *Mathematical models of protein kinase signal transduction*, Molecular Cell, 9 (2002), pp. 957–970.

[27] D. W. HUANG, B. T. SHERMAN, AND R. A. LEMPICKI, *Systematic and integrative analysis of large gene lists using david bioinformatics resources*, Nature Protocols, 4 (2008), pp. 44–57.

[28] D. HWANG, I. Y. LEE, H. YOO, N. GEHLENBORG, J.-H. CHO, B. PETRITIS, D. BAXTER, R. PITSTICK, R. YOUNG, D. SPICER, N. D. PRICE, J. G. HOHMANN, S. J. DEARMOND, G. A. CARLSON, AND L. E. HOOD, *A systems approach to prion disease*, Molecular Systems Biology, 5 (2009).

[29] N. ISHIKURA, J. L. CLEVER, E. BOUZAMONDO-BERNSTEIN, E. SAMAYOA, S. B. PRUSINER, E. J. HUANG, AND E. A. DEARMOND, *Notch-1 activation and dendritic atrophy in prion disease*, Proceedings of the National Academy of Sciences of the United States of America, 102 (2005), pp. 886–891.

[30] H. JEONG, S. P. MASON, A. BARABASI, AND Z. N. OLTVAI, *Lethality and centrality in protein networks*, Nature, 411 (2001), pp. 41–42.

[31] M. KANEHISA AND S. GOTO, *Kegg: Kyoto encyclopedia of genes and genomes*, Nucleic Acids Research, 28 (2000), pp. 27–30.

[32] S. KOYASU, *The role of pi3k in immune cells*, Nature Immunology, 4 (2003), pp. 313–319.

[33] R. MILO, S. ITZKOVITZ, N. KASHTAN, R. LEVITT, S. SHEN-ORR, I. AYZENSHTAT, M. SHEFFER, AND U. ALON, *Superfamilies of evolved and designed networks*, Science, 303 (2004), pp. 1538–1542.

[34] V. H. PERRY, C. CUNNINGHAM, AND D. BOCHE, *Atypical inflammation in the central nervous system in prion disease.*, Curr. Opin. Neurol, 15 (2002), pp. 349–354.

[35] J. PRILLER, M. PRINZ, M. HEIKENWALDER, N. ZELLER, P. SCHWARZ, F. HEPPNER, AND A. A, *Early and rapid engraftment of bone marrow derived microglia in scrapie*, J Neurosci, 26 (2006), pp. 11753–11762.

[36] S. B. PRUSINER, *An introduction to prion biology and diseases*, Cold Spring Harbor Monograph Archive, 41 (2004).

[37] P. REZAIE AND P. LANTOS, *Microglia and the pathogenesis of spongiform encephalopathies*, Brain Res Brain Res Rev, 35 (2001), pp. 55–72.

[38] J. SCHULTZ, A. SCHWARZ, S. NEIDHOLD, M. BURWINKEL, C. RIEMER, D. SIMON, M. KOPF, M. OTTO, AND M. BAIER, *Role of interleukin-1 in prion disease-associated astrocyte activation*, The American Journal of Pathology, 165 (2004), pp. 671–678.

[39] P. SHANNON, A. MARKIEL, O. OZIER, N. BALIGA, J. T. WANG, D. RAMAGE, N. AMIN, B. SCHWIKOWSKI, AND T. IDEKER, *Cytoscape: a software environment for integrated models of biomolecular interaction networks*, Genome Research, 13 (2003), pp. 2498–2504.

[40] D. SIMON, M. E. HERVA, M. J. BENITEZ, J. J. GARRIDO, A. I. ROJO, A. CUADRADO, J. M. TORRES, AND F. WANDOSELL, *Dysfunction of the pi3k-akt-gsk-3 pathway is a common feature in cell culture and in vivo models of prion disease*, Neuropathology and Applied Neurobiology, 40 (2014), pp. 311–326.

[41] G. SORENSEN, S. MEDINA, D. PARCHALIUK, C. PHILLIPSON, C. ROBERTSON, AND S. A. BOOTH, *Comprehensive transcriptional profiling of prion infection in mouse models reveals networks of responsive genes*, BMC Genomics, 9 (2008).

[42] C. SOTO AND N. SATANI, *The intricate mechanisms of neurodegeneration in prion diseases*, Trends in Molecular Medicine, 17 (2011), pp. 14–24.

[43] D. SZKLARCZYK, A. FRANCESCHINI, M. KUHN, M. SIMONOVIC, A. ROTH, P. MINGUEZ, T. DOERKS, M. STARK, J. MULLER, P. BORK, L. J. JENSEN, AND M. C. VON, *The string database in 2011: functional interaction networks of proteins, globally integrated and scored*, Nucleic Acids Research, 39 (2011), pp. 561–568.

[44] S. WERNICKE AND F. RASCHE, *Fanmod: a tool for fast network motif detection*, Bioinformatics, 22 (2006), pp. 1152–1153.

[45] P. J. WILLIAMS, A AMD LUCASSEN, D. RITCHIE, AND M. BRUCE, *Prp deposition, microglial activation, and neuronal apoptosis in murine scrapie.*, Exp Neurol, 144 (1997), pp. 433–438.

[46] S. K. YANG, Y. C. WANG, C. C. CHAO, Y. J. CHUANG, C. Y. LAN, AND B. S. CHEN, *Dynamic cross-talk analysis among tnf-r, tlr-4 and il-1r signalings in tnf-induced inflammatory responses*, BMC Medical Genomics, 3 (2010).