# DEVELOPMENT OF COMPUTATIONAL PIPELINE AND FRAMEWORK FOR DIFFERENTIAL AND CO-EXPRESSION NETWORKS ANALYSIS

*Thesis submitted in fulfillment for the requirement of the Degree of*

## DOCTOR OF PHILOSOPHY

By

**ANKUSH BANSAL**

Department of Biotechnology and Bioinformatics

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY

WAKNAGHAT, DISTRICT SOLAN, H.P., INDIA

NOVEMBER, 2018

**Dedicated To**

**My Mom - Dad**

# TABLE OF CONTENTS

# DECLARATION

I certify that:

a. The work contained in this thesis is original and has been done by me under the guidance of my supervisors.

b. The work has not been submitted to any other organisation for any degree or diploma.

c. Wherever, I have used materials (data, analysis, figures or text), I have given due credit by citing them in the text of the thesis.

**Ankush Bansal** **Date:**

(Enrollment No. 156501)

Department of Biotechnology and Bioinformatics

Jaypee University of Information Technology, Waknaghat, India

# CERTIFICATE

This is to certify that the thesis entitled, "**DEVELOPMENT OF COMPUTATIONAL PIPELINE AND FRAMEWORK FOR DIFFERENTIAL AND CO-EXPRESSION NETWORKS ANALYSIS**" which is being submitted by **Ankush Bansal (Enrollment No. 156501)** in fulfillment for the award of degree of **Doctor of Philosophy** in **Bioinformatics** at **Jaypee University of Information Technology, India** is the record of candidate's own work carried out by him under our supervision. This work has not been submitted partially or wholly to any other University or Institute for the award of this or any other degree or diploma.

**Dr. Tiratha Raj Singh**
Supervisor – I
Associate Professor
Faculty In-Charge, Media Relations
Department of Biotechnology and Bioinformatics
Jaypee University of Information Technology
Waknaghat, Solan,
Himachal Pradesh
Email: tiratharaj.singh@juit.ac.in

**Dr. Rajinder Singh Chauhan**
Supervisor – II
Dean (Research & Consultancy)
Professor and Head
Department of Biotechnology,
School of Engineering and Applied Sciences
Bennett University, Plot No. 8-11,
Tech Zone II, Greater Noida, Uttar Pradesh
Email: rajinder.chauhan@bennett.edu.in

# ACKNOWLEDGMENT

*This Ph.D has been a long and fantastic voyage. I am indebted to lots of people I met on the way, and to those who were with me from the beginning. Although this section is rather long, it is certainly incomplete.*

*First and foremost, I owe my deepest gratitude to **Prof. Vinod Kumar** (Vice Chancellor, JUIT); **Maj. Gen. Rakesh Bassi** (Registrar, JUIT); **Prof. S.D. Gupta** (Dean, Academic and Research, JUIT); **Dr. Sudhir Kumar** (HOD, BT & BI); for providing me with the opportunity to pursue my doctorate degree. I would like to sincerely thank **Prof. R.S. Chauhan** (former Dean and HOD, BT & BI) for his encouragement and cooperation during these years and proving to be the prime example of how BT & BI department is held together.*

*I would like to express my immense gratitude to my supervisor **Dr. Tiratha Raj Singh** for his faith in me, providing me with ideas, motivation and positive criticism. I would like to especially thank him for helping me evolve to a competent computational biologist. Hir excellent guidance, persistent support and understanding of complicated issues, his sheer expertise in solving major and minor problems and invaluable advice have made this thesis possible.*

*I would like to extend a vote of thanks to the members of my thesis advisory committee; **Dr. Hemraj Saini, Dr. Harish Changotra, Dr. Raghu M. Yennamalli,** I gratefully acknowledge their constant input during my Ph.D. I also want to appreciate and thank **Prof. R.S. Chauhan** from Department of Biotechnology, Bennett University for their constant support in helping me to fulfill my Ph.D objectives.*

*I would like to thank everyone in the Bioinformatics Lab along with **Arvind, Kusum, Imran, Sonia, Smriti, Nadia, Nupur and Tarun**. Sincere thanks to interdepartment friends **Amit Sharma, Neelam Gupta Shah, Ankur Choudhary, Ashutosh Sharma, Prabhat Thakur, Kirti Tripathi, Ekta Thakur and Jyotsana Dogra** for their constant support. Thanks to **Pulkit, Mittu, Hitesh and Sid** for being in support and encouragement. I am indebted to be surrounded by such great friends who helped me celebrate and survive all the ups and downs of my research journey. I would especially like to thank **Ashwani Kumar and Ankita Shukla** and also deeply acknowledge the assistance of the **Somlata Ma'am Sonika Ma'am**, **Ira Ma'am, teaching**,* **administration** *and* **server room** *staff who were always willing to help.*

# ABSTRACT

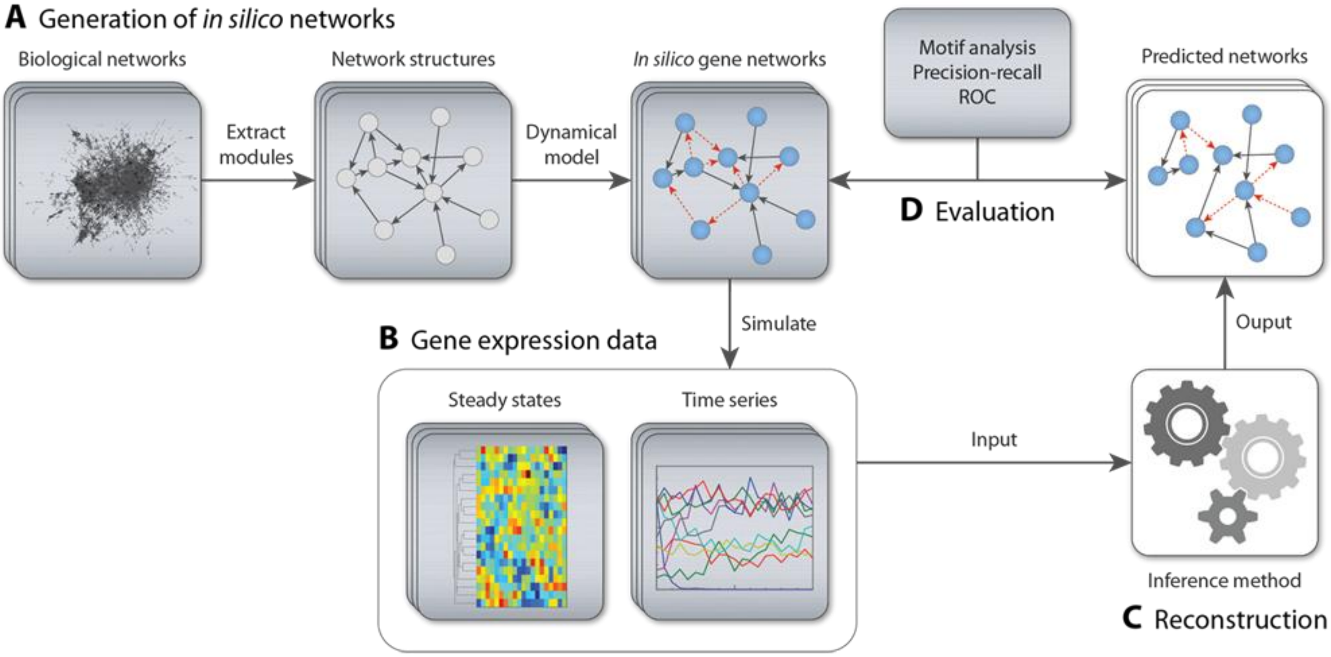
Understanding the general principles governing the functioning of biological networks is a major objective of the work presented in this thesis. Functionality of biological networks can be viewed from two aspects, *viz.*, differential and co-expression network. All possible modes of operations of biological networks are confined by the differential expression analysis. Regulation then imposes additional constraints that determine which of the numerous possible phenotypes is observed under a given condition. It is relatively easy and intuitive to understand and interpret differential aspects of biological function in light of information flow in terms of gene expression. In contrast, operation of regulatory circuits and their effects on cellular operations has been to large extent a descriptive science. Nonetheless, new opportunities to uncover and test principles of regulation are being opened through the availability of large amount of molecular abundance data on genome-wide scale. Several of the existing approaches in this direction, however, are data-driven and thus lack potential to be generalized and extrapolated to different species. Thus new pipelines and platform need to be built on hypotheses and data rather than data-only are necessary for integration of omics data and discovery of biologically meaningful patterns. To overcome the research gap between data and hypothesis driven analysis we have developed three pipelines and framework which focus on differential and co-expression network construction. First objective of work focuses on miRNA network analysis framework which specifically focuses on identification of miRNA, prediction of miRNA targets and then constructing a co-expression network using gene ontology and correlation analysis. For this analysis we have used differential transcriptome datasets of healthy and diseased conditions in *Jatropha curcas*. Second objective focuses on deciphering pathway of interest after constructing a global network. For this analysis we have used gene dataset which are specifically involved in cancer metastasis. Third objective of deals with drug target interaction networks analysis by using Picroside molecules as potent inhibitory molecules. Overall, the work presented in this thesis will add value to existing differential transcriptomic or genomic datasets by constructing networks to indentify key genes or proteins of interest. Data analysis as well as different hypothesis testing can be done using proposed pipelines.

# LIST OF FIGURES

# LIST OF TABLES

# CHAPTER 1

## INTRODUCTION


A Generation of *in silico* networks
Biological networks
Extract modules
Network structures
Dynamical model
*In silico* gene networks
Motif analysis
Precision-recall
ROC
Predicted networks
D Evaluation
Simulate
B Gene expression data
Steady states
Time series
Input
Ouput
Inference method
C Reconstruction

## 1.1 Definition of “System” and “Systems Biology”

In various courses of physics, mathematics, and thermodynamics, I have studied about the well-defined system which is used for predicting and analyzing the behavior of physical systems. Still, it took me a while to realize that the implementation of a system is more worth in analyzing a scientific question rather than simply defining it. Thus, a system under investigation may be defined as a particular set of objects and boundaries which is being described or predicted as a phenomenon. A transparent and completely defined system under investigation is always subjective and depends upon the nature of the problem. Besides, these problems and ambivalences ensure certain assumptions to solve the problem conceptually and define a system up to a certain extent. Systems are speculated in various ways in different research fields as defined below:

**System:** The word system (derived from the Latin (systēma) is a collection of elements in which each element is related to other. If any of the elements are not related to the other one, it cannot be considered as a part of that particular system. Thus a Subsystem is defined as a set of elements having an accurate subset of the whole system.

**System (Biological):** A system which consists living conditions and can maintain homeostasis even change in internal and external environment. Bioinformatics aids in understanding biological systems through various approaches like systems biology or network biology.

**System (Thermodynamic):** A system is said to be thermodynamic if it is the part of the universe that is under consideration. Environment, surroundings or reservoir is the real or fictional borderline that separates the system from the rest of the universe. The character of the periphery and the quantities flowing through it, such as work, heat matter, entropy, and energy are some of the parameters on which the classification of a thermodynamic system depends. Anything can be considered a system such as a mixture inside a test tube, a piston, a planet or a living organism etc. Figure 1.1 represents the general concept of a thermodynamical system.

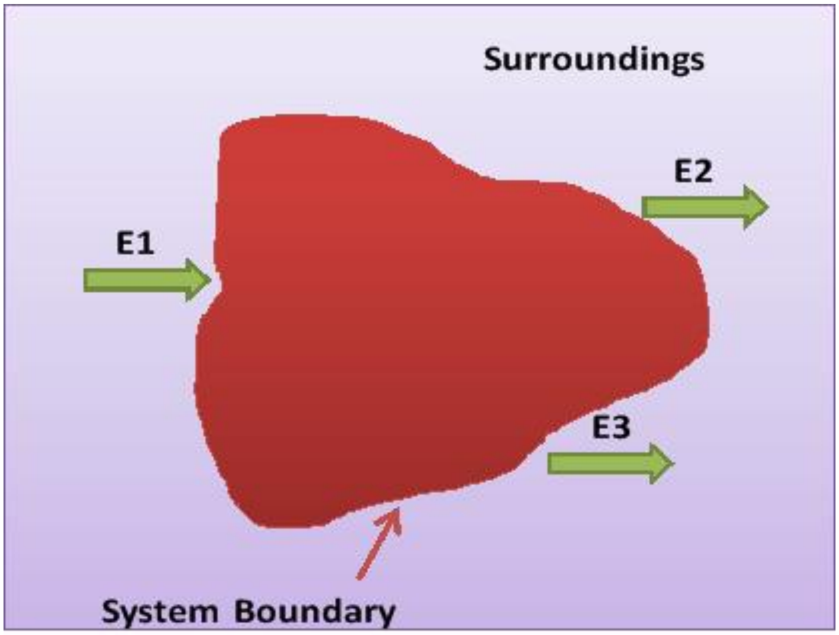


**Figure 1.1** The General Concept of a Thermodynamical System. If $E_1$, $E_2$, and $E_3$ denote the energies crossing the system boundary, then the energy balance equation for this system may be depicted as $E_1 = E_2 + E_3$.

The most signifying part of these definitions is the denoted text " a system can be anything". The above-said statement depicts the problems faced in understanding the concept of the system. Any of the definitions of the system is uncertain in nature; however, this uncertainty can be decreased with imposed speculations and limitations on the desired accuracy of the illustrations/predictions. The most remarkable difference between the above three definitions of systems is a priori condition of relatedness required by the first definition. This simply contradicts the statement "a system can be anything." Keeping these contradictions in mind, I will adapt the thermodynamic definition of the system in light of my views regarding the nature of systems. Furthermore, the first definition is very inaccurate as:

1) It does not define "element".

2) It does not clarify what type of relations is expected between the elements.

3) Although the nature of the relationship between elements is defined, it does not explain whether such relationships can/should always be established a priori?

4) Most significantly, it does not force the existence of any boundaries, and thus makes it conceptually problematic to use in a systematic way.

Although the first definition does not define a system properly but the second definition represents the system well once defined. Nevertheless, the second definition seems closer to the one which is used in this thesis.

First of all, I will clarify the meaning of “systems biology” which implies “study of biological entities as systems”. Needless to say, the repetitive nature of this definition (here due to the term entity) cannot be avoided as for the general concept of a system. In this regard, it can be declared that all biological study is systems biology and similarly all sciences are systems sciences as one cannot even define a scientific problem without possessing a definition of the underlying system. As far as definition is concerned, this argument is completely valid; still, it is possible to draw a boundary (up to certain extent) between biology and systems biology based on the application (and mode of application) of the speculations designated the systems under investigation.

*A system biology problem may be defined as any investigation of a biological system that addresses and accounts for the defined interactions between defined components under investigation (in terms of their contribution in describing/predicting the behavior/properties of the system).*

The word “defined” should be highlighted here, as it incidentally sets the boundaries of a system and thus makes the definition more relevant. The need for this limitation may not be evident in many cases. However, if we think about all potential unknown factors that might influence the properties and behavior of the systems (e.g. unknown/weak interactions), the urge for clearly defining the boundary of a system is immediately evident. This definition is diligent and other generally accepted definitions of systems biology can be seen as special examples of this definition. The definition denoted here is not targeted to classify existing biological research, but only to justify and rationalize the usage of this term in this thesis.

### 1.1.1 Implementation of System-Level Analysis

This definition of system is difficult to use in biological problems because of the inherent complexity of the components and interactions. The true association of using these kinds of definitions can be illustrated by analogy with other fields of science where it has proved to be tremendously useful. I hereby admit that identical attempts in systems biology will help us in the future to compile different levels of knowledge to get a holistic perspective of the cellular machinery. Systemic integration is well explained by the demonstration of an ancient Indian story of an elephant and blindfolded men.

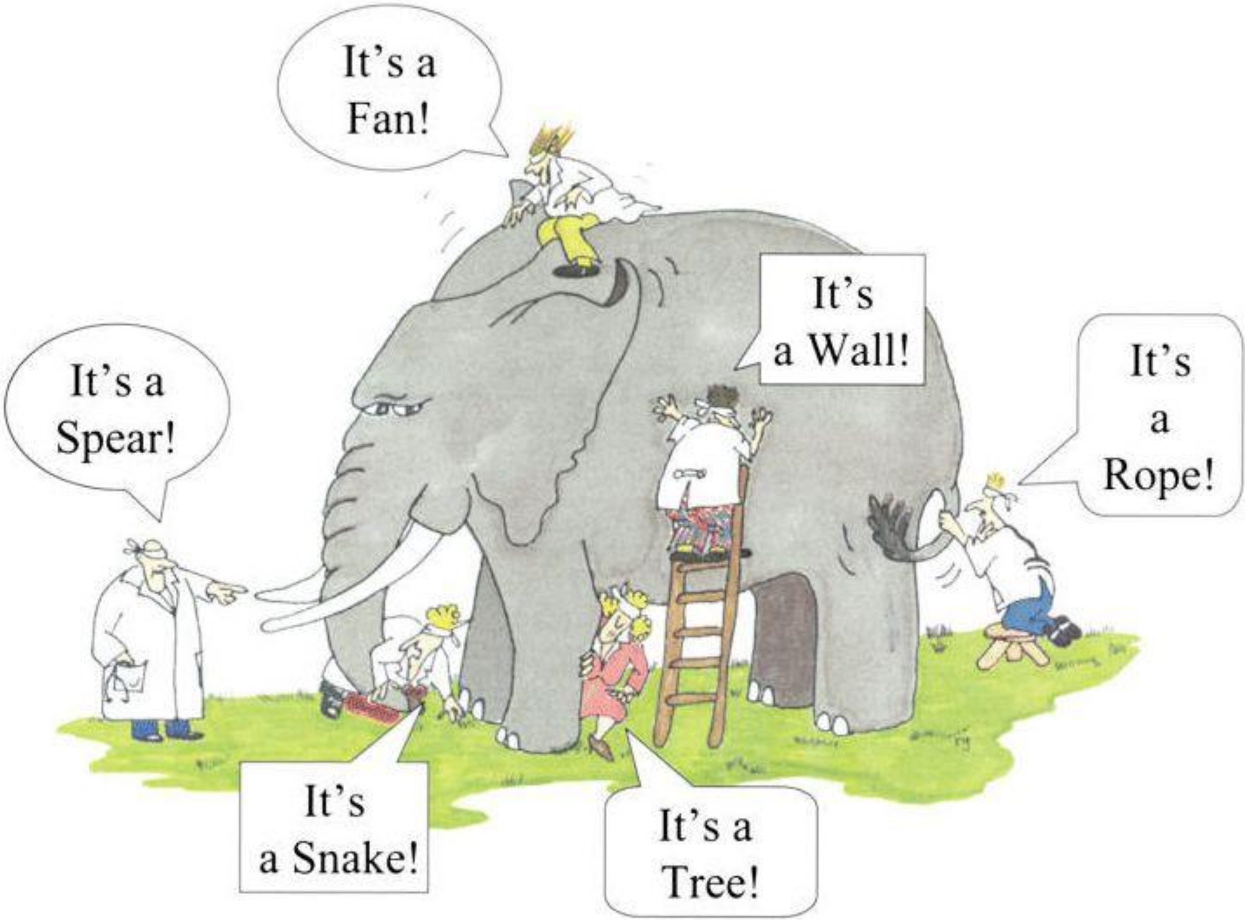


**Figure 1.2** Pictorial representation of an old Indian storyline in which six blindfolded men tries to understand the elephant's structure by touching various parts of an elephant. The storyline carries the concept of understanding the true picture of reality by the unification of all different viewpoints Image adopted from Wikipedia.

In the story, six men who are blindfolded tried to understand the structure of the elephant by touching it. Each one of them described it as a different shape as each man was feeling a different part of the elephant's body. An intelligent man made them understand that all of them were right and the elephant possesses all those shapes. Thus, the story reflects the idea that the complete view of the truth can be acquired only by combining knowledge from different levels of analysis. In addition, this process of systemic integration will also guide us to design biological parts and whole systems with desirable functionality. A scientific discipline mentioned to as systems biology aims at attaining this goal.

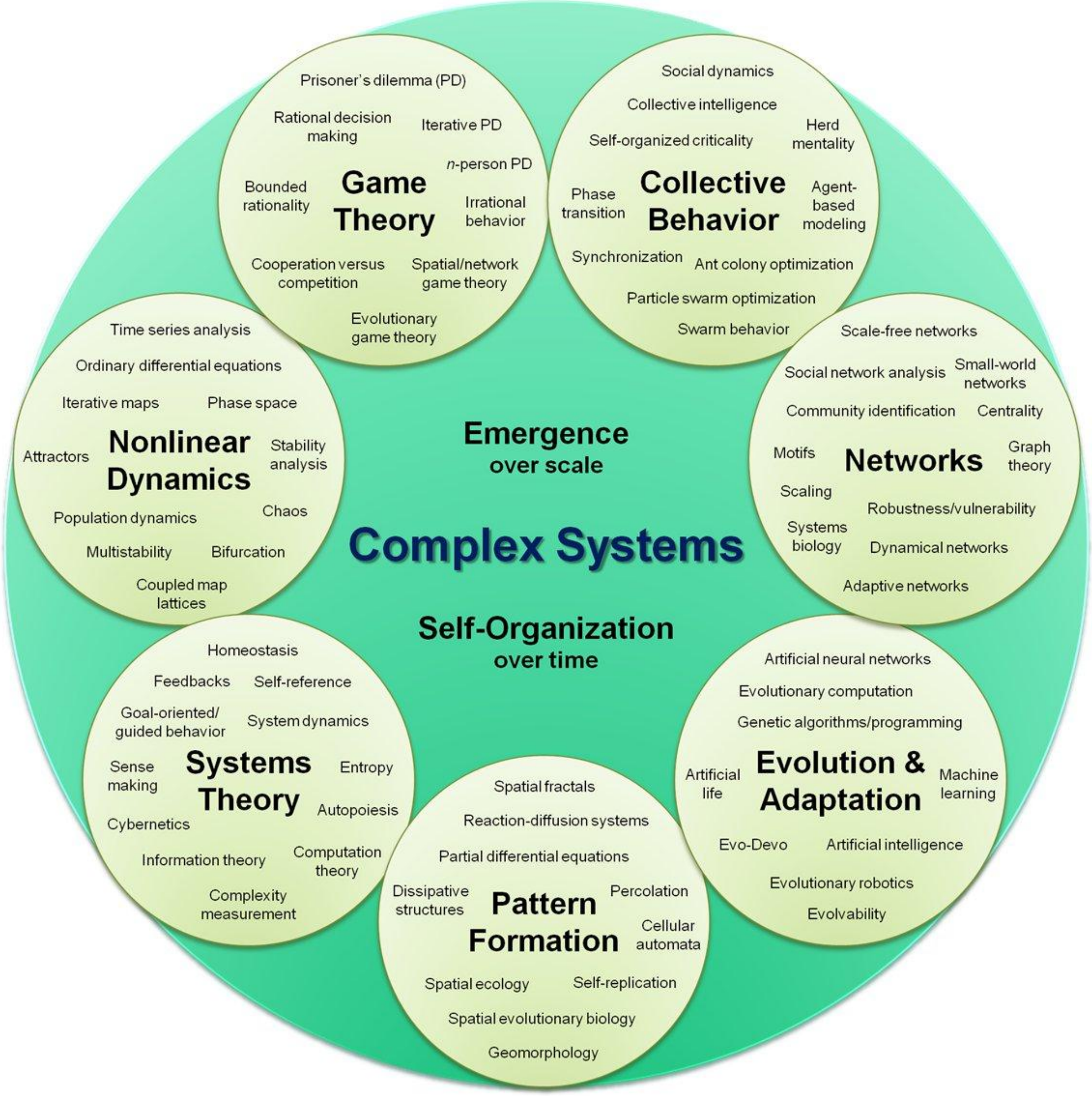


**Figure 1.3** The emergence of complex systems over time in various domains of interdisciplinary fields.

## 1.2 Networks

Understanding complex systems usually need a detailed approach, which splits the system into tiny and primary components and plots the interactions between these components. In most studies, the numerous components and interactions are remarkably identified as networks. For instance, the society comprises of a network of people bridged by various connections, such as relationships, partnerships or scientific co-authorships. Electronic communications mainly depends on two extremely different networks: such as the web of homepages joined by Uniform Resource Locator (URLs), World Wide Web (WWW) and the network wiring to the routers to-gather (Internet) and cellular-phone, airlines, power-grid or business data networks denotes other instances of complex data structures of scientific-technological or economic engrossment.

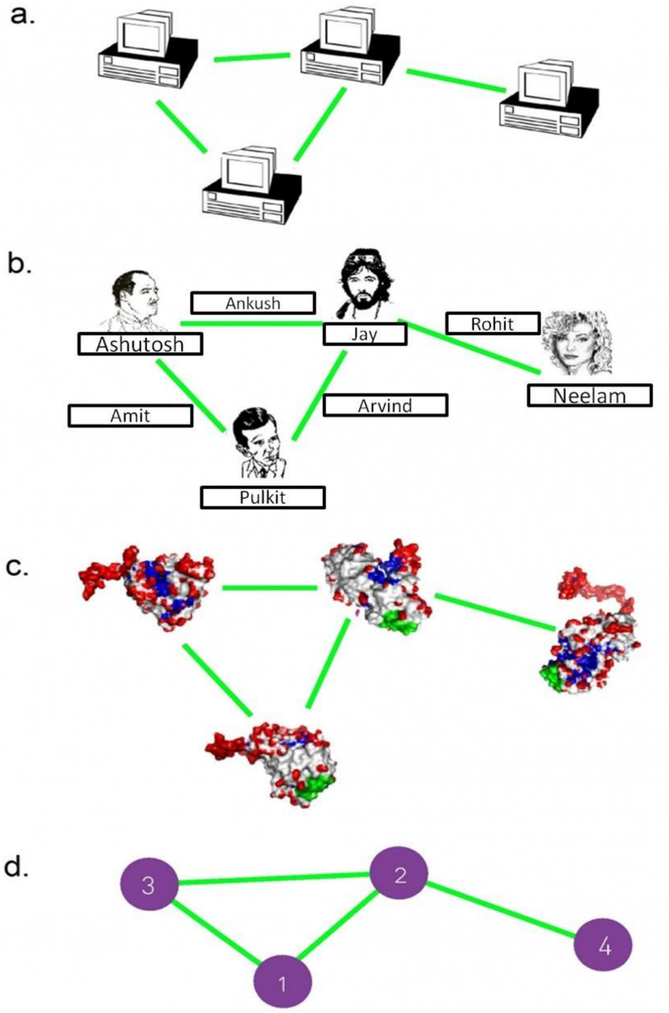


**Figure 1.4** Figure comprises a basic representation of a variety of network types a) computer networks; b) community networks; c) biological networks; d) mathematical networks

### 1.2.1 Network Structure

Network structure is associated with nodes and their interactions. Node's degree represents the number of links $k$ associated with node under supervision. Degree of node is considered as the utmost basic parameter of network analysis. For instance, in Fig. 1 nodes $i$ and $j$ exhibit three links ($k = 3$). The overall graph, although, is denoted by the average degree $\langle k \rangle$, which exhibit the value $\langle k \rangle = 2.6$ in this example. Although, the average degree does not represent the potential degree variations in the network. This is better denoted by the degree distribution $P(k)$, which provides the number of nodes with $k$ links.

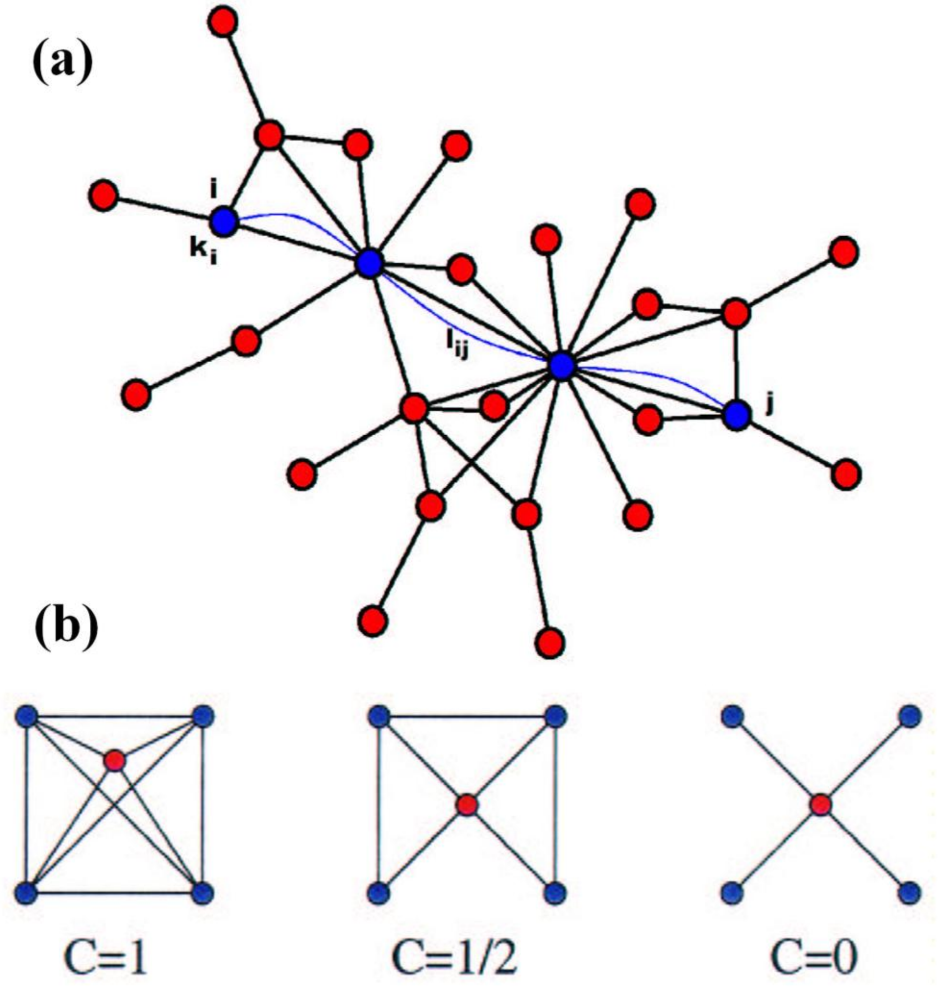


**Figure 1.5** Structure of network a) red color represents overall network and blue color represents the selected path on the basis of topological parameters b) C represents the clustering coefficient of network under investigation

Mostly all networks have various paths in between any two nodes $i$ and $j$. The main contributing distance measure is shortest path length i.e., $l_{ij}$ (see Fig. 1.5).

The mean path length is given by

$$\langle l \rangle = \frac{2}{N(N-1)} \sum_{i=1}^{n} l_{ij} \qquad (1)$$

The small world property was often displayed when a network is intersected by the numerous small steps, initially demonstrated on social network structures stating that two chosen individuals can only be connected by six connections [1].

Various authentic systems have the nodes which exhibit a susceptibility to cluster for which clustering coefficient is used to quantify them [2]. Generally, the clustering coefficient of a node is written as

$$C_i = \frac{2n_i}{k_i(k_i - 1)} \tag{2}$$

where $n_i$ represents the total count of links joining the $k_i$ neighbors of node $i$ to one another. Consequently, we can define an average clustering coefficient as

$$\langle C \rangle = \frac{1}{N}\sum_{i=1}^{N} C_i \tag{3}$$

As stated, the degree distribution $P(k)$ and the $k$ dependence of $C(k)$ can have generic features, which allow us to differentiate various network. Parameters such as the average degree$\langle k \rangle$, average path length $\langle l \rangle$and average clustering coefficient $\langle C \rangle$ distinguish the unique properties of the specific network under contemplation, and thus are less generic.

### 1.2.2 Network Models

The key role of the network models is to define the emergence and behavior of some of the most relevant network properties. As they play an important role in framing our perception of complex networks, we have to pay attention to some of the efficient models.

### 1.2.3 Random Networks

As graph theory initially represents regular graphs, there was no inference of large networks as random onces before 1950’s [3]. As per the Erdos & Renyi (ER) model of random graphs

(ErdÄos & Renyi, 1960), we begin with $N$ nodes and join each couple of nodes with probability $p$, designing a graph with approximately randomly divided links. The ER graph has an exponential degree distribution and exhibits the small-world property as shown in Fig 1.6. Indeed, in the ER network, most nodes have approximately the same number of links, and the mean path length is proportional to the network size.

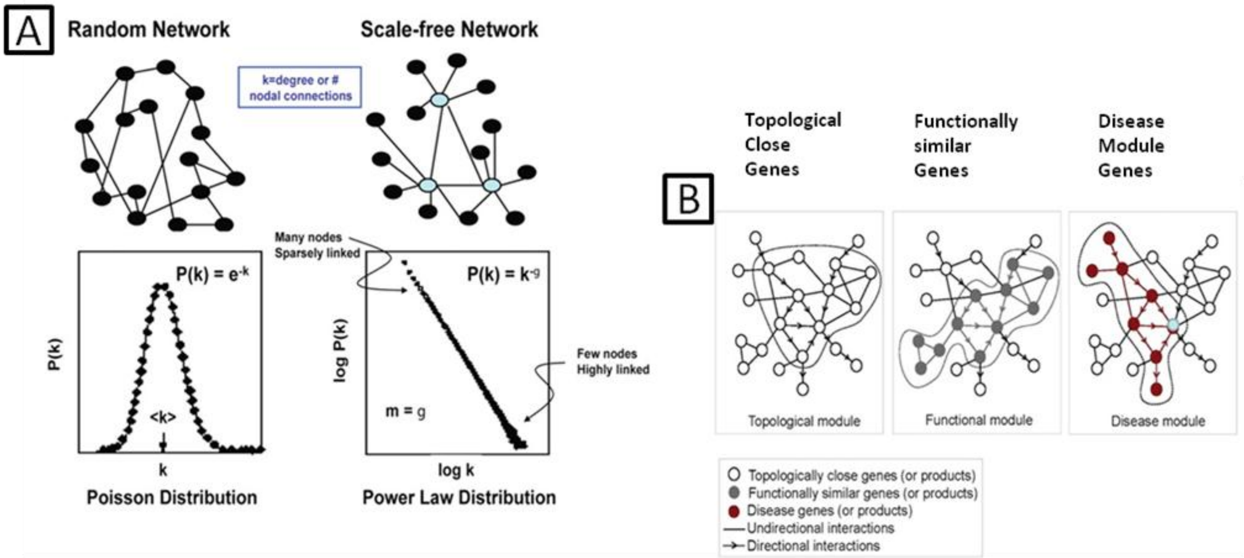

A
Random Network
Scale-free Network
k=degree or # nodal connections
P(k) = e^-k
P(k)
<k>
k
Poisson Distribution
Many nodes Sparsely linked
P(k) = k^-g
log P(k)
m = g
Few nodes Highly linked
log k
Power Law Distribution
B
Topological Close Genes
Functionally similar Genes
Disease Module Genes
Topological module
Functional module
Disease module
○ Topologically close genes (or products)
● Functionally similar genes (or products)
● Disease genes (or products)
— Undirectional interactions
→ Directional interactions

**Figure 1.6** Random and Scale Free Networks a) Random networks follow Poisson distribution while Scale free networks follow Power Law distribution b) Variation in network under differential investigation like topological module, function module, and disease module

### 1.2.4 Scale-free Networks

The significant growth in our learning of complex networks was due to the various wide matrix. The degree division follows a power-law which involves the metabolic and protein interaction networks [4], [5]. These grids are known as scale-free, as a law of power does not follow the continuance of a particular scale. For the creation of this power-law degree distribution, two mechanisms missing from the classical random network model, are accountable [6]. Initially, many web grids develop by the involvement of newer nodes and they connect to nodes already existing in the network system. There is maximum probability of connecting to a node with numerous links, which is a characteristic named as a preferential attachment in most of the true networks. Barabasi and Albert (BA) instigated this model (scale-free) which comprises these features.

A significant result of the hubs in that networks (scale-free) have high sufferance to unspecified perturbations, but these are susceptible to target on the joined nodes [7]. Consistently failure of selected nodes may not destroy the network's nobility. Although, planned eradication of hubs will lastly break the network. In biological systems, this feature has marked significance because it reveals the biochemical network's flexibility against mutations. This is because biochemical networks have highly connected nodes which may be strong candidates for targets of drugs.

In a scale-free biological network, the existence of hubs has a fundamental effect on the virus. Epidemiological models reveal that infectious diseases will inescapably die out which are having transmission probability under an epidemic threshold. Nevertheless, the epidemic threshold is minimized to zero in scale-free networks [8]. Even highly weakly infectious viruses can expand and prevail, making random immunization unproductive, as some social and sexual networks are known to possess a scale-free topology [9].

### 1.2.5 Hierarchical Networks

The network can be seamlessly divided into a collection of modules as most of the true networks (grid) are assumed to be fundamentally modular. Every module performs a perceptible task, which differs from the function of other modules [10]–[13]. Therefore, with the network's potential modularity we should restore the scale-free characteristic. A random or the scale-free network model is not modular as it is shown by numerical simulations [12].

For accounting the coexistence of modularity, scale-free topology and local clustering in systems, we supposed that clusters join in an interactive manner, forming a stratified network [14], [15]. These networks appear from frequentative integration and duplication of grouped nodes; it is a procedure which in principle occurred again perpetually. A beginning point is a tiny group of four thickly connected junctions.

Standard not always guarantee clear-cut subnetworks which are joined in clear-cut ways. The module's boundaries are often significantly blurred which are activated by joined nodes having interconnected modules.

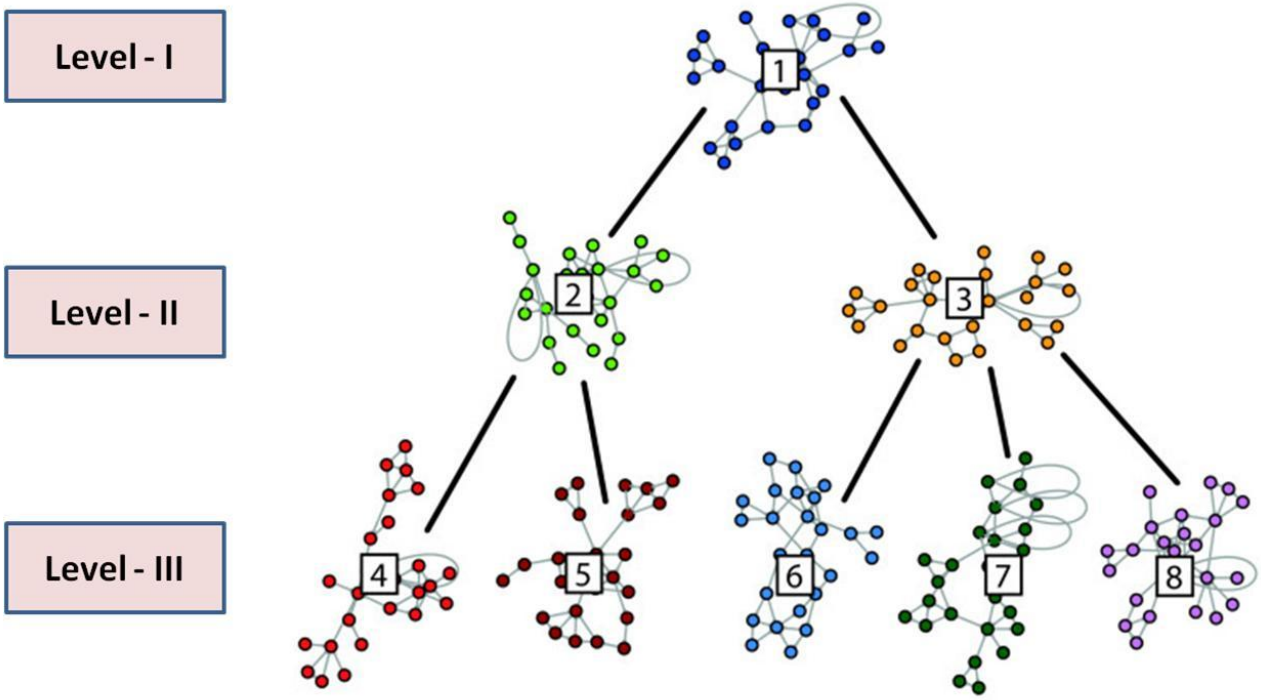


**Figure 1.7** Hierarchal network represents three level where 1module is Level-I; II, III modules are at level-II; and 4,5,6,7,8 modules are at Level-III.

## 1.3 Biological Networks

### 1.3.1 Transcriptional Regulatory Networks

Expression at the genetic level can be estimated and regulated at transcriptional and translational level through various regulatory elements. Moreover, mRNA decay, RNA cleavage, translational repression plays a major role in transcriptional regulation. There are various regulatory elements like miRNA, Transcription Factors (TFs) and cis-trans regulatory elements. TFs and associated target genes form a multi-partite network where one TF is regulating multiple genes in a network under supervision. IN transcriptional regulatory networks (TRNs) nodes represent TFs and target genes while edges represent a direct association between interacting entities. TRNs show variation in physiological and phenotypic levels. TRNs understanding can lead to precise detection for cellular and molecular changes which can be used for targeted therapies. Dysfunction in such regulatory mechanisms leads to deterioration at interactome levels.

Computational modeling of TRNs can help in a much sophisticated manner to understand complex diseases like cancer, neurodegenerative disorders, etc.

### 1.3.2 Metabolic Networks

There are various studies in the literature which have described metabolic network in many ways. By using stoichiometric equations, building blocks of metabolic networks and metabolic energy distribution was performed in *E.coli* model [16], [17]. If two substrates are falling in the same reactions, they can be connected and substrate graph can be formed. *E.coli* model depicted network as scale-free and specified that few nodes showed the highest degree compared to others. Nodes are showing highest degree form cliques and supposed to remain conserved during evolution. All biological systems follow the behavior of power-law. Variation in the substrate doesn’t influence the behavior of the systems and its mean deviation. Therefore, in most model construction, various researchers found a similar number of substrates whether vary in organism type. In various network biology studies, it has been found that structural and spectral properties remain conserved in various organisms which might be due to convergence at evolutionary level [11], [18], [19].

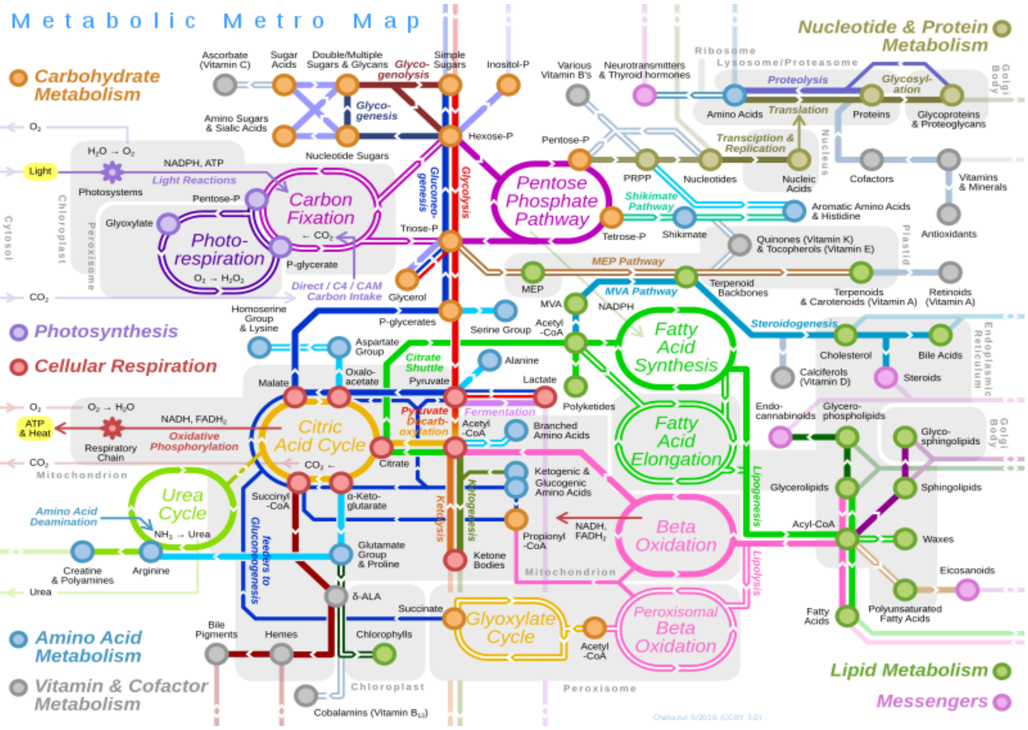


**Figure 1.8** Metabolic Networks representing various routes and intersection among routes

### 1.3.3 Protein Interaction Networks

Protein interactions play a major role in providing insights about network regulation where proteins play the role of junction while physical interactions are termed as binding. From literature studies, it is evident that S. cerevisiae and H.pylori networks show scale-free behavior (Jeong et al. 2001; Wagner and Fell 2001). Interactome data at various protein level indicate the robustness and consistent behavior of scale-free networks. Scale-free networks were not found targeted for any target based studies as networks are unprotected. Due to high degree of joined nodes in a network it is very difficult to find out mutations in interacting proteins. It has been shown in the studies that hub protein keep networks intact and sparse nodes can be seen together in Yeast interactome dataset (Farkas et al. 2003).

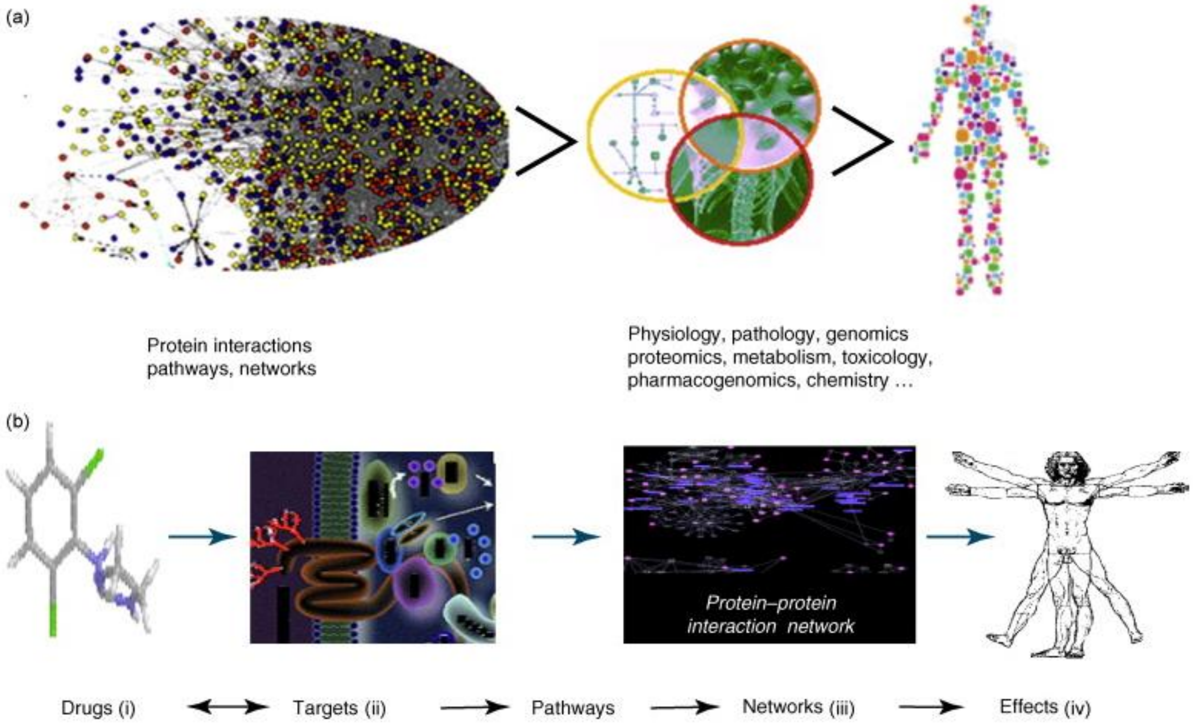


**Figure 1.9** Protein Interaction Networks a) Protein-Protein interaction networks show the impact on various levels including physiology, pathology, and omics level b) Drug-target interaction and association of pathway at the network level to trace down the effects at the phenotypic level

### 1.3.4 Protein Domain Networks

Protein domains usually exist in the same patterns. Therefore, studies have proved domains as a cluster in various protein domain networks which lies under the scale-free network behavior

[20]–[23]. Domain origins showed dynamic behavior which was consistent to scale-free networks. Domain functions reserve in various multi-cellular organisms while their positions vary in many organisms: kinases, ribosomal machinery, mitochondrial genes found to play a crucial role in their domain-specific activity. Moreover, there is variation in the degree distribution behavior in large complex domain networks. Highly joined domains form clusters regarding the family while conservation has been seen at sequence as well as structural level [23], [24].

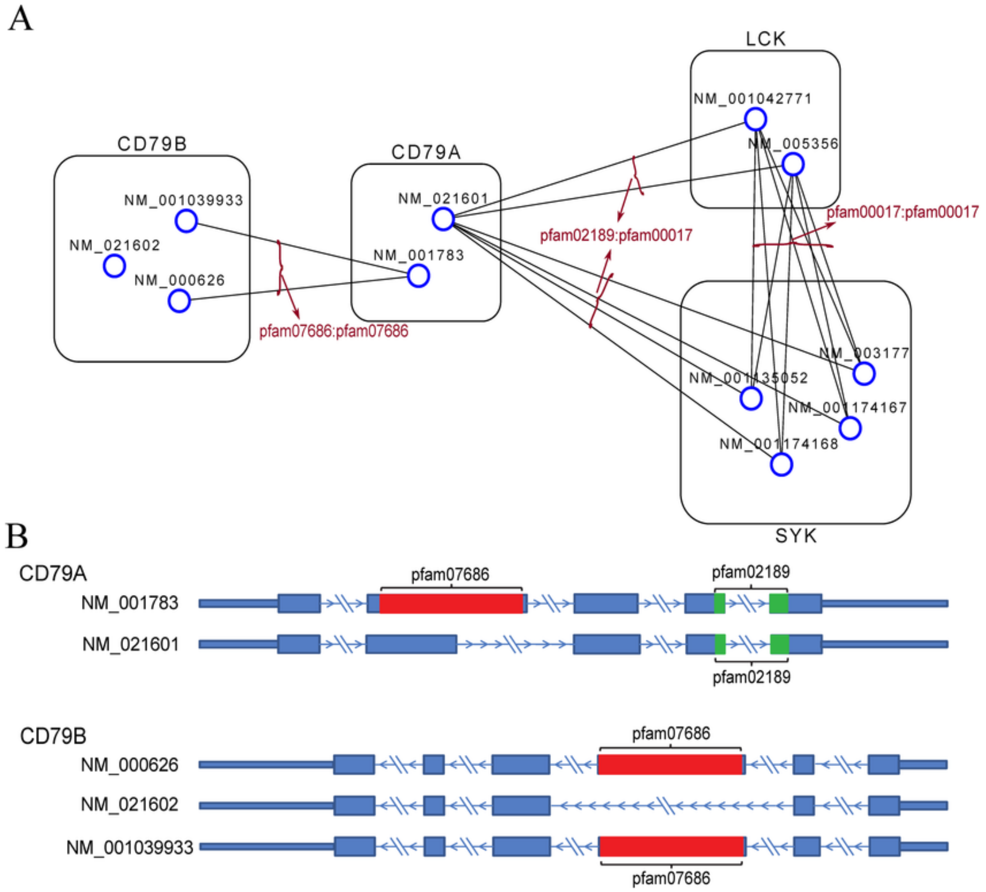


**Figure 1.10** Protein Domain Networks a) PFAM domain exploration in between CD79A, CD79B, LCK, SYK and finding communicating common network b) Linear representation of protein domain network shown in a).

### 1.3.5 Hierarchies in Biological Networks

Basic scale of biological networks has a great impact on the regulatory understanding of nodes regarding biological entities. Here, biological entities represent genes, proteins, metabolites, etc. [25]–[28]. These entities further form clusters and variation in clusters and association in clusters forms networks. Further, there is an integration of similar modules, and dissimilar modules need to be differentiated using sub-network construction [29]. Next stage needs regulation of these network modules using transcriptional activation or transcriptional repression. Alternate splicing and variation in protein form further results in differential gene expressions which can be estimated regarding FPKM (Fragments Per kilobase Per Million Reads) values [30], [31].

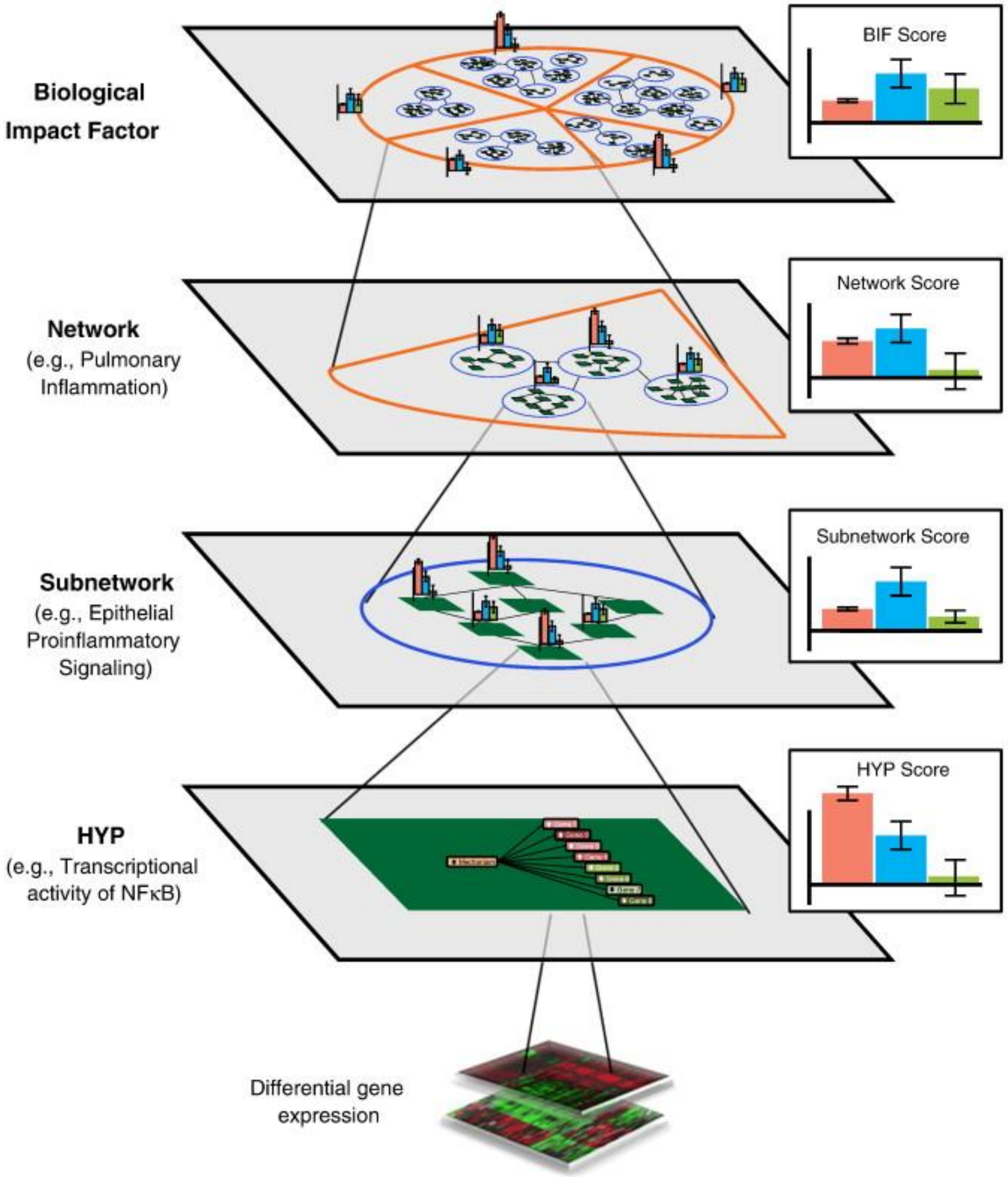

Biological Impact Factor
BIF Score
Network (e.g., Pulmonary Inflammation)
Network Score
Subnetwork (e.g., Epithelial Proinflammatory Signaling)
Subnetwork Score
HYP (e.g., Transcriptional activity of NFκB)
HYP Score
Differential gene expression

**Figure 1.11** Network hierarchy represents the biological impact on network and sub-network construction with the inference of regulatory elements and gene expression analysis

## 1.4 Network Construction

### 1.4.1 Molecular Mechanisms

Modern cellular biology is generally understood and studied in the form of flow diagram (based on central dogma) [32] showing how information encoded in genes (genotype) is considered at the level of cellular function and state (phenotype). Data stored in the genesis in the form of sequence order consists of four nucleotides (Adenine, Cytosine Guanine and Thymine). Out of 20 amino acids a triplet (e.g., ATG) codes for one amino acids. Many of the amino acids combined to form a protein. Hence, gene codes for a particular protein via transcription (DNA to mRNA) and translation (mRNA to protein).

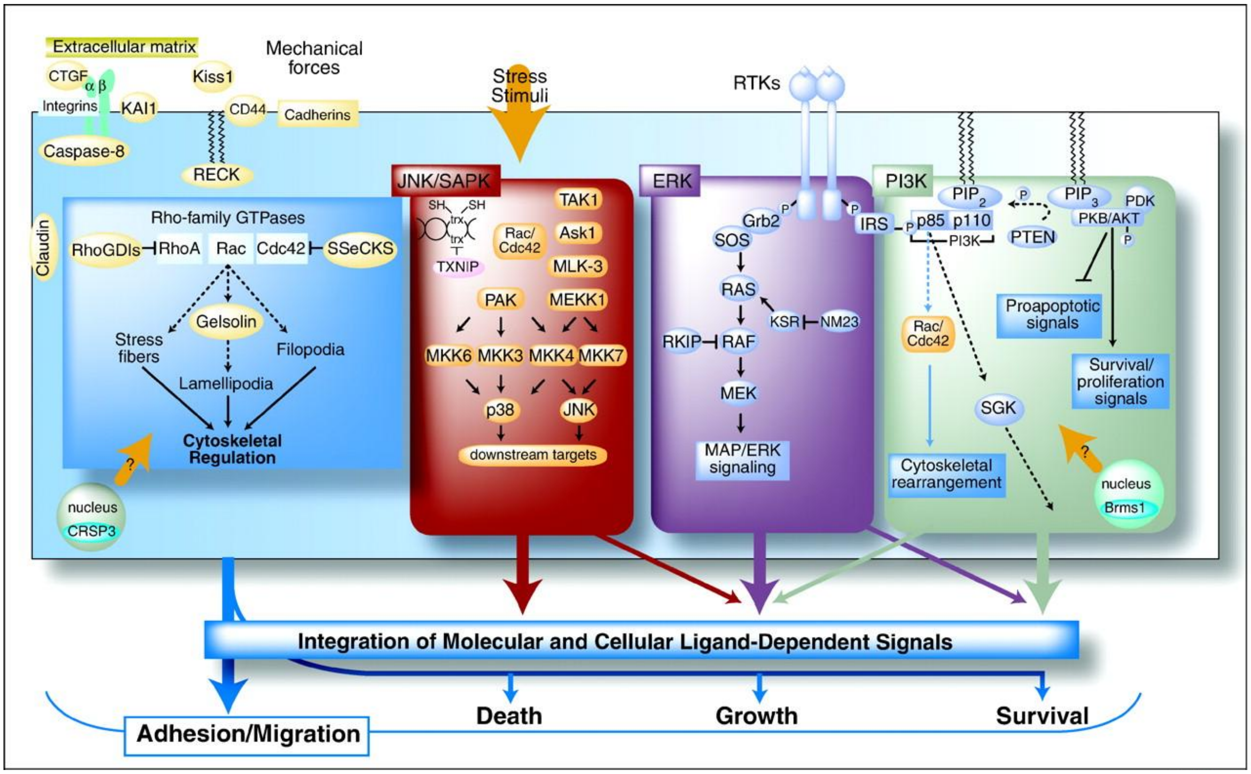


**Figure 1.12** Differential metabolic networks analysis leads to differential effects on cellular phenotype like Adhesion/Migration, Death, Growth, and survival

Proteins play a significant role as a catalyst for chemical and physical transformation of various chemical substances. These proteins are commonly called as enzymes. Together, these enzymes frame a network of reactions where substrates (food) present in the environment is crushed down to form energy and building block molecules. Generated energy is then utilized to assemble these building blocks towards framing new cells and for retaining the existing cells. The whole process is called metabolism which works through a metabolic network. The term metabolite is normally used to refer to only specifically “low” molecular weight compounds, and it eliminates all cellular substances that are genetically encoded (e.g., RNA and proteins) [33].

### 1.4.2 Network Inference and Implementation

Exceptionally, the primary architecture of metabolic networks is widely preserved across many divergent species ranging from microscopic bacteria to plants and humans [34]. Hence, the cellular machinery fueling well-defined functionality and phenotypes are founded on similar metabolic processes. To understand the graph-theoretical representation of metabolic networks can do these simple organizational principles of metabolic networks. Enzymes and metabolites can be seen as nodes in this network, whereas interactions between them from edges. The number of neighbors for a node is called its degree. The topological analysis is the study of network characteristics based on its connectivity. Metabolic networks normally form a complete combined network, i.e., it is feasible to go from one node to other nodes. The total count of edges that crossed such path is called the distance between two nodes. Metabolic networks in various species reveal same scale-free topology [35] where some of the metabolites are involved in a large number of reactions (also called as hubs), whereas most of the metabolites participate in a lesser number of reactions [5]. Metabolic hubs also confer riveting small-world property to these networks [16]; which meant no two nodes (enzymes or metabolites) in a metabolic network are too away from one another. For instance, any two nodes enzymes/metabolites in yeast metabolic network are on averagely five edges far away from one another. However, the high connectivity in metabolic networks should regularly be viewed only in the ambiance of stoichiometric restrictions on the flow of materials. In this regard, metabolic networks vary from other networks like protein-protein interaction networks, electrical grids, and the internet.

Metabolic networks over variant species are, although, divergent in many respects. Variations are more prominent at the level sequences/structures of particular enzymes and the modulation of enzymes in response to environmental/genetic challenges. Modulation of enzyme production is mandatory for an organism to i) To assign resources optimally so as to produce only enzymes that are required under given context and only in desired amounts, and ii) avoid surplus (or too less) amounts of enzymes which may result in unstable dispensation of substrates that are invading the cell. To obtain further details on the notion of metabolic modulation, it is essential to describe the term “flux”. The number of substrates processed (or products produced) per unit time are the flux for a particular reaction, in a metabolic context. The integrated effect of fluxes by variant reactions in a network can be seen in the phenotype of a cell (e.g., amounts of substrate used and final products and by-products produced, the growth rate of an organism, etc.). Because of the highly interconnected character of metabolic networks and stoichiometric constraints, fluxes are to a greater extent inter-dependent. Moreover, inter-dependencies between fluxes can be properly plotted in a flux-coupling graph [36] where an edge joins two flux nodes effecting each other. Surprisingly, however not unexpectedly, the flux-coupling graph also depicts a scale-free topology marked by some hub fluxes [36].

Operation and organization of metabolic networks are typically seen and understood as the ensemble of pathways. Pathways consist of groups of enzymes acting upon production/breakdown of some of the metabolites (group of metabolites). Glycolysis and TCA cycle are common instances of pathways. However, the idea of a pathway is extensively used and very helpful for pictorial representation, the meaning of a pathway is very indefinite from the stoichiometric point of view. A different and more detailed way of understanding the operation of metabolism is by enumeration of all possible reactions combinations that can hold an equitable flow from substrates to final products. Every combination (termed Elementary Flux Mode) can then be viewed as a meaning of a stoichiometrically “complete” pathway. A significant factor is that the number of such feasible routes is of the order of millions, even in a normal microbial metabolic network [37]. Any vital metabolism at a stable state can be shown as a linear combination of these elementary flux modes. What are the factors/mechanisms accountable for specific phenotype under provided conditions? It seems that this choice is attained through coordinated modulation of enzymes. Hence, not even a single enzyme is modulated for its optimal operation [38], but also the whole network is subjected to modulation for optimal network functionality [39]. Modulation of enzymes can take place at the level of transcription, translation or post-translational alterations (e.g., phosphorylation). Moreover, enzyme activity may be modulated by small effector molecules. The flux by a reaction is based not only on the activity and availability of the enzyme but also on the concentrations of substrates, effectors, and products [40]. These relationships are commonly non-linear. Also, because of the interconnected nature of the metabolic network, all steps in the metabolism can influence every step. Accordingly, simulation, understanding, and prediction of both dynamic and steady-state operations of metabolic network are demanding tasks.

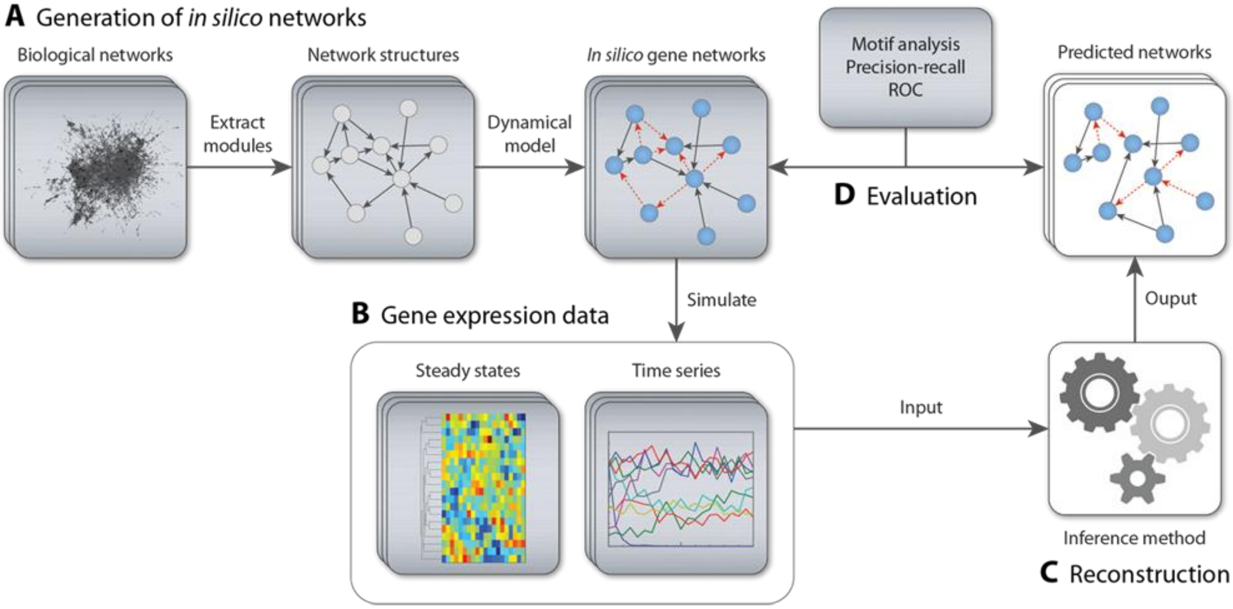


**Figure 1.13** Global network reconstruction steps a) Generation for *in-silico* networks b) Gene expression data c) reconstruction d) evaluation

## 1.5 Overview of thesis

Networks are the most basic ways to understand and depict various perspectives of life. Many such network examples can easily be seen and inferred to solve the biotechnology engineering problems. Understanding the general principles governing the functioning of biological networks is a major objective of the work presented in this thesis. The functionality of biological networks can be viewed from two aspects, *viz.*, differential and co-expression network. The differential expression analysis confines all possible modes of operations of biological networks. The regulation then imposes additional constraints that determine which of the numerous possible phenotypes is observed under a given condition. It is relatively easy and intuitive to understand and interpret differential aspects of biological function in light of information flow regarding gene expression. In contrast, the operation of regulatory circuits and their effects on cellular operations has been too large extent a descriptive science. Nonetheless, new opportunities to uncover and test principles of regulation are being opened through the availability of a large amount of molecular abundance data on a genome-wide scale. Several of the existing approaches in this direction, however, are data-driven and thus lack potential to be generalized and extrapolated to different species. Thus new pipelines and platform need to be built on hypotheses and data rather than data-only are necessary for integration of omics data and discovery of biologically meaningful patterns.

The cellular and molecular response in any organism is often traced from expression analysis. Expression variation often occurs due to changes in regulatory mechanisms regulated by transcriptional activation of repression. Regulation is complex mechanisms where a number of pathways collaborates to contribute the expression at a genetic level. It is very difficult to understand the global interactions through traditional approaches where targeted genes can be considered for supervision. Therefore, I have focused on the development of pipelines and framework we understand the topological properties of the network and referring useful information concerning meaningful patterns. Major focus in this thesis focuses on the regulatory network construction from regulatory elements, constructing the global network and narrow down to the pathway of interest. Also, we have developed a pipeline to look into drug-target.

Another outcome of the work presented in this thesis is the demonstration of the importance of highly connected genes in the regulation and functionality of the regulatory network. Many of the highly connected genes (such as redox and energy co-factors) are usually omitted *a priori* from the analysis of transcriptome and other omics data. Here it is shown that not only such omission is unnecessary, but it may also critically affect the results obtained. Since highly connected metabolites glue the network together, their role regarding transcriptional regulation is significant for understanding and interpreting global changes in the network.

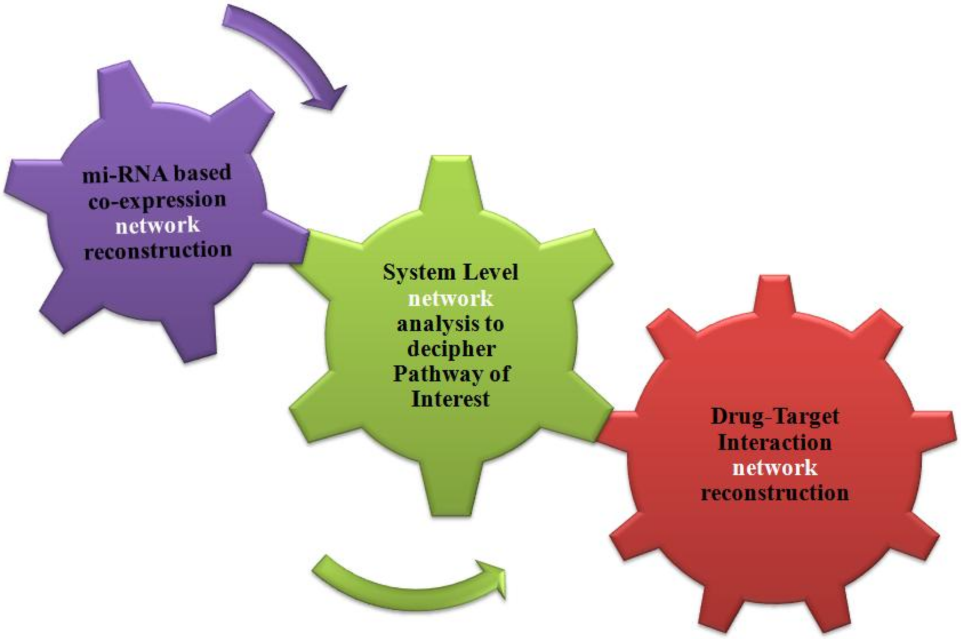

mi-RNA based co-expression network reconstruction
System Level network analysis to decipher Pathway of Interest
Drug-Target Interaction network reconstruction

**Figure 1.14** Objectives underlying in this thesis

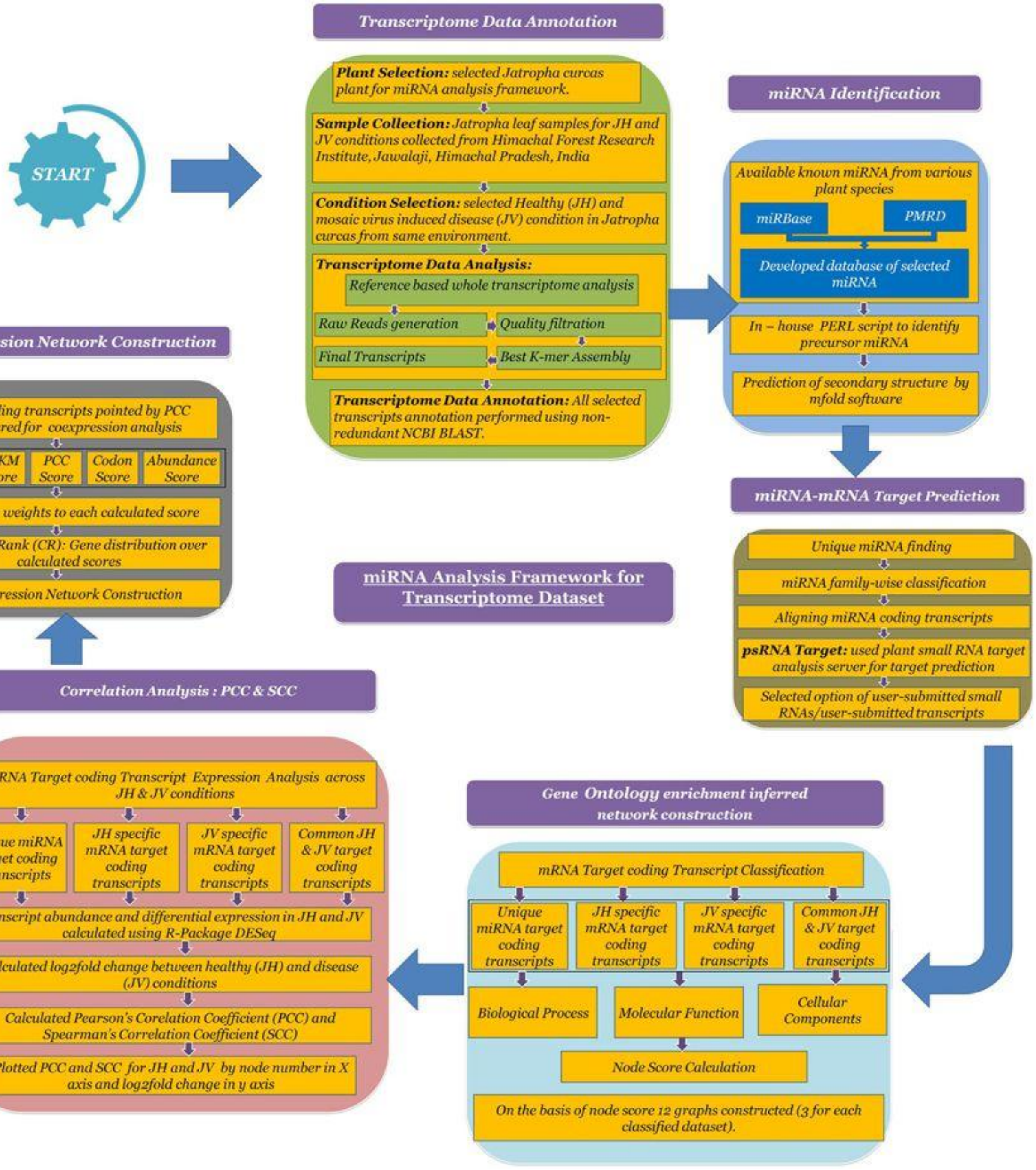
Objective – I: Development of novel micro-RNA analysis and co-expression network framework to analyze differential biological networks.

Objective – II: A system level network analysis framework to snare transition from metabolic flux.

Objective – III: A computational pipeline for drug-target interaction network analysis.

# CHAPTER 2

## A novel miRNA analysis framework to analyze differential biological networks


START
Transcriptome Data Annotation
Plant Selection: selected Jatropha curcas plant for miRNA analysis framework.
Sample Collection: Jatropha leaf samples for JH and JV conditions collected from Himachal Forest Research Institute, Jawalaji, Himachal Pradesh, India
Condition Selection: selected Healthy (JH) and mosaic virus induced disease (JV) condition in Jatropha curcas from same environment.
Transcriptome Data Analysis:
Reference based whole transcriptome analysis
Raw Reads generation
Quality filtration
Final Transcripts
Best K-mer Assembly
Transcriptome Data Annotation: All selected transcripts annotation performed using non-redundant NCBI BLAST.
miRNA Identification
Available known miRNA from various plant species
miRBase
PMRD
Developed database of selected miRNA
In – house PERL script to identify precursor miRNA
Prediction of secondary structure by mfold software
miRNA-mRNA Target Prediction
Unique miRNA finding
miRNA family-wise classification
Aligning miRNA coding transcripts
psRNA Target: used plant small RNA target analysis server for target prediction
Selected option of user-submitted small RNAs/user-submitted transcripts
Gene Ontology enrichment inferred network construction
mRNA Target coding Transcript Classification
Unique miRNA target coding transcripts
JH specific mRNA target coding transcripts
JV specific mRNA target coding transcripts
Common JH & JV target coding transcripts
Biological Process
Molecular Function
Cellular Components
Node Score Calculation
On the basis of node score 12 graphs constructed (3 for each classified dataset).
Correlation Analysis : PCC & SCC
mRNA Target coding Transcript Expression Analysis across JH & JV conditions
Unique miRNA target coding transcripts
JH specific mRNA target coding transcripts
JV specific mRNA target coding transcripts
Common JH & JV target coding transcripts
Transcript abundance and differential expression in JH and JV calculated using R-Package DESeq
Calculated log2fold change between healthy (JH) and disease (JV) conditions
Calculated Pearson's Corelation Coefficient (PCC) and Spearman's Correlation Coefficient (SCC)
Plotted PCC and SCC for JH and JV by node number in X axis and log2fold change in y axis
Coexpression Network Construction
Gene coding transcripts pointed by PCC considered for coexpression analysis
GO Score
FPKM Score
PCC Score
Codon Score
Abundance Score
Assigning weights to each calculated score
Combined Rank (CR): Gene distribution over calculated scores
Co-expression Network Construction
miRNA Analysis Framework for Transcriptome Dataset

## 2.1 Abstract

A system biology approach is providing an edge to deal with complex networks. Various approaches have integrated the nodes and edges by graph theory and topological parameters. As large high throughput data always face the problem of global network visualization, there is need a framework of regulatory network understanding. In this chapter, we have developed a unique miRNA analysis framework by considering *Jatropha curcas* as query dataset — two differential conditions, i.e., healthy and jatropha mosaic virus diseased condition. To resolve the network complexity, we have used gene ontology and correlation analysis. By gene ontology networks and correlation analysis, we have constructed the co-expression modules. These co-expression modules were further integrated with pathways to get an overall understanding of pathway regulations. The defined framework will help experimental biologist to screen out candidate genes from next-generation sequencing data.

## 2.2 Introduction

Complex networks have resolved various complexities in computer science, physics, and electronics. Complex networks properties in biological sciences have been highlighted in last decade. Network-based drug-target understanding become a trend to determine gene function and potential inhibition of disease. Various networks have been studied in interactome analysis like transcriptional regulatory networks, protein-protein interaction networks, and protein-drug networks [41]–[43].

Network biology deals with deciphering the molecular interactions from large biological datasets. All biological networks have interactions within different types of networks to decipher regulatory mechanisms [44], [45]. Integration of theoretical and experimental knowledge always been a crucial task as an estimation of network parameters do not remain constant over the different data types or differential conditions.

miRNA are a small nucleotide sequence of range 18-22 nucleotides which do not code for protein. miRNA are known to regulate or control the expression at protein level through various mechanisms like transcriptional repression and mRNA decay. These small regulatory elements found to be crucial in controlling metabolic processes and associated pathways [46]. miRNAs have a great impact on the various biological process (BPs), and alterations in miRNA targets affect the pathway regulation. miRNAs leads to over-expression and down-regulation of its targets and affect the phenotype like host–gut microbiota interactions [47], host-parasite interactions [48], and transgenerational epigenetic inheritance[49]. Various studies have indicated that homology-based correlation in differential conditions resulted in variation at phenotypic levels [50], but the mechanism of controlling mechanism remains unclear.

Major identification is a major step in understanding the function of miRNAs [51], [52]. Computational tools and approaches have paved the path towards understanding miRNAs role, but false positive rates are very high in established methods [53]. From many years next generation sequencing has been used to identify novel genes in various plant species [54], [55]. Prioritization of genes for silencing and inhibiting the specific pathway usually carried out by RNAseq or microarray-based expression profiling in differential conditions. In various studies, these high throughput methods have reduced the efforts of biotechnologists by getting a screenshot of expression profiles in given conditions [56]. Role of miRNA in defense and virus-induced response reported in various species [57]. The high throughput sequencing results in better quality and quantity of miRNA target screening.

There are various online tools, and web-serves developed to identify miRNAs from various plant species [58][59] [60] [61]. In the last decade, various miRNA based databases have been constructed to annotate the information at various disease and regulatory mechanism understanding levels [62][63][64][65][66]. These web servers, tools, and databases provide rich information, but there is a need for exploration of functional regulation. Understanding of targets at functional level can be done using c-expression network reconstruction and absolute pathway mapping of gene regulatory modules.

After identification of miRNA and its associated targets, determining the function of target remains as an essential task which needs to be addressed for a better understanding of regulatory mechanisms. miRNAs have the potential to target multiple genes, so it is very difficult to prioritize the targets affecting various BPs. Scientists look into enrichment analysis of targets to resolve such issues. Enrichment analysis also gives insight about randomly occurred targets and comparison with conserved patterns in miRNA targets by function and pathway association. But, at the same time, these techniques are not universal for various conditions hence parameter based optimization, and correlation analysis are required to confirm the targets.

Comparison of protein-coding genes with miRNA targets usually results in variation in results at the phenotypic level. [67]. It might be due to miRNAs regulation occurs through inhibition at various targets instead of single target. Henceforth, it is required to see the combined effect of miRNAs targets and associated functions in subsequent pathways, and maybe it is resulting in synergistic effect in mechanistic regulation. Network biology can help to understand many such miRNAs to many targets association using holistic network visualization techniques. Moreover, correlation and co-expression analysis using various enrichment analysis may aid in sophisticated outcome regarding regulation. Identifying genes which are targeted by various miRNA may present a novel way to control gene of interest through in-house mechanisms. Such kind of analytical studies is not deciphered from available tools and web-servers.

In this chapter, we are presenting a miRNA analysis framework which can be used for any omics data analysis by providing datasets as sample input. Our framework works in the form of a module driven pipeline which used transcriptome analysis, miRNA identification and target prediction. Target prioritization was done using gene ontology parameters, correlation score by drawing a co-expression network reconstruction. Further, mapping of co-expression module may provide insight about pathway regulation.

## 2.3 Methods

### 2.3.1 Data Assortment

Transcriptomic data of healthy and virus infected diseased condition was collected from http://14.139.240.55/download.php [68].As reference genome of the Jatropha was available the We have performed comparison of transcriptomes against known genome for possible gene coding transcripts. SAMtools was used for alignment and mapping.

### 2.3.2 miRNA Identification

High quality reads from both healthy and diseased transcriptome datasets were retrieved. We have downloaded non-redundant (nr) databases to compare high quality reads for further identification. In-house perl script has been used for miRNA identification [69]. By sequence data availability in miRBase [70]and plant microRNA database (PMRD) [71] we have constructed an in-house database and performed the homology-based comparison with given transcriptomic condition. There was a need to remove duplicate entries to screen out unique miRNAs. To look into the secondary structure, we have used mFold software[72]. By miRNA precursor, stem-loop, mature miRNA sequence, and loop break secondary structure of lowest energy were determined.

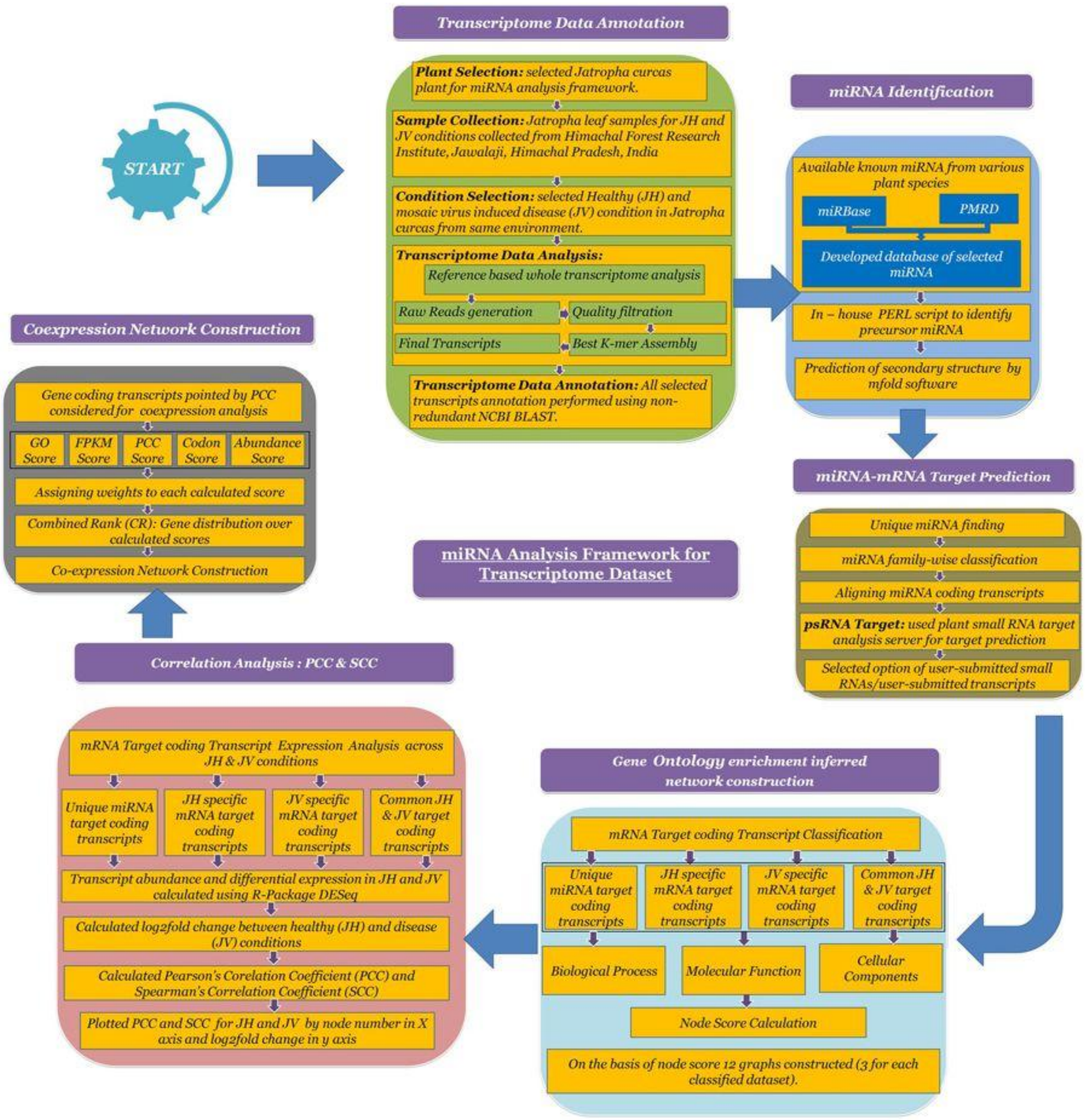

START
Transcriptome Data Annotation
Plant Selection: selected Jatropha curcas plant for miRNA analysis framework.
Sample Collection: Jatropha leaf samples for JH and JV conditions collected from Himachal Forest Research Institute, Jawalaji, Himachal Pradesh, India
Condition Selection: selected Healthy (JH) and mosaic virus induced disease (JV) condition in Jatropha curcas from same environment.
Transcriptome Data Analysis:
Reference based whole transcriptome analysis
Raw Reads generation
Quality filtration
Final Transcripts
Best K-mer Assembly
Transcriptome Data Annotation: All selected transcripts annotation performed using non-redundant NCBI BLAST.
miRNA Identification
Available known miRNA from various plant species
miRBase
PMRD
Developed database of selected miRNA
In – house PERL script to identify precursor miRNA
Prediction of secondary structure by mfold software
miRNA-mRNA Target Prediction
Unique miRNA finding
miRNA family-wise classification
Aligning miRNA coding transcripts
psRNA Target: used plant small RNA target analysis server for target prediction
Selected option of user-submitted small RNAs/user-submitted transcripts
Gene Ontology enrichment inferred network construction
mRNA Target coding Transcript Classification
Unique miRNA target coding transcripts
JH specific mRNA target coding transcripts
JV specific mRNA target coding transcripts
Common JH & JV target coding transcripts
Biological Process
Molecular Function
Cellular Components
Node Score Calculation
On the basis of node score 12 graphs constructed (3 for each classified dataset).
mRNA Target coding Transcript Expression Analysis across JH & JV conditions
Unique miRNA target coding transcripts
JH specific mRNA target coding transcripts
JV specific mRNA target coding transcripts
Common JH & JV target coding transcripts
Transcript abundance and differential expression in JH and JV calculated using R-Package DESeq
Calculated log2fold change between healthy (JH) and disease (JV) conditions
Calculated Pearson's Corelation Coefficient (PCC) and Spearman's Correlation Coefficient (SCC)
Plotted PCC and SCC for JH and JV by node number in X axis and log2fold change in y axis
Correlation Analysis : PCC & SCC
Coexpression Network Construction
Gene coding transcripts pointed by PCC considered for coexpression analysis
GO Score
FPKM Score
PCC Score
Codon Score
Abundance Score
Assigning weights to each calculated score
Combined Rank (CR): Gene distribution over calculated scores
Co-expression Network Construction
miRNA Analysis Framework for Transcriptome Dataset

**Figure 2. 1** Methodology for proposed miRNA analysis framework

### 2.3.3 miRNA Target Prediction

For prediction of miRNA targets, we have used The plant small RNA (psRNA) target web-server. For target identification, we have used default parameters.[73]. Parameters considered for target prediction includes the length of complementary scoring, maximum expectation value, number of top target genes, flanking length, target accessibility maximum energy. By stated parameters, the selected targets were classified to respective families. There is a need to cross-check the psRNA target results with other tools. Hence we have made a comparison with

TargetFinder and in-house shell scripts to remove false positive hits. Results of psRNA target remain same compared with other methods.

### 2.3.4 miRNA–mRNA Interaction Network Analysis

Systems biology states that cellular networks follow most of the principles of nature and network biology oriented framework can be used to understand disease pathology and other regulatory mechanisms [74]. The miRNA analysis framework is shown in Fig 1.

### 2.3.5 Confusion Matrix Construction

miRNA and mRNA target can be represented using confusion matrix:

$$j = \begin{cases} 1\ if i \sim j \\ 0\ otherwise \end{cases} \qquad (1)$$

Where *i* and *j* depicts the miRNA and mRNA target for interactome analysis.

### 2.3.6 Bipartite Network Construction

Network is Bi-partite if we are able to split in two parts and the elements of one part are related to other but there should not be any connection in the elements of one part. Directed bipartite network represents the control behavior of the systems where elements of one side control the elements of other side. Fig 2.4 represents the graphical view of miRNA and its target as regulating elements. In our analysis, we have constructed three networks, healthy, diseased and network with differential expression of the targets. Moreover, we have performed the GO analysis to fiind out the association between healthy and diseased condition.

### 2.3.7 GO enrichment inferred network

Gene ontology infers involvement of transcripts into biological processes, molecular functions, and cellular functions. We have used BLAST2G)[75] to annotate transcripts at given molecular molecular levels using scoring functions which may be referred as:

$$Scoring\ (g) = \sum_{g_\alpha \in desc(g)} gp(g_\alpha).\alpha^{dist(g,g_\alpha)} \qquad (2)$$

By scoring functions, same transcripts clustered together. Node score highlights the regulation of genes in same or different pathways. Same score genes processed in the same rectangle while variation in the score, leads to a different rectangle. Nearest scoring rectangles were connected as it considered being a part of the same regulation mechanism.

### 2.3.8 Degree and Correlation Analysis

Degree of a node can be represented by network, network may be directed or undirected. If network is undirected then we might have unknown behavior of system or it may be evident from random behavior of the system. In case, network is directed we will be having the chance to look at in degree and out degree of a given network. Further, correlation analysis in terms of PCC also referred as was performed to understand relation and variation of same nodes in differential conditions.

$$r = \frac{\left[M^{-1}\sum_{i=1}^{M} j_i k_i\right] - [M^{-1}\sum_{i=1}^{M} \frac{1}{2}(j_i + k_i)^2]}{[M^{-1}\sum_{i=1}^{M} \frac{1}{2}(j_i^{\,2} + k_i^{\,2})] - [M^{-1}\sum_{i=1}^{M} \frac{1}{2}(j_i + k_i)^2]} \qquad (3)$$

where *ji* and *ij* represents the degree of considered mRNA target transcripts while *M shows* connections in total.

We have used in-house script to calculate Pearson’s correlation and Spearman's correlation. From FPKM values, Pearson’s correlation found to be more consistent with our results. Unique candidates in JH and JV conditions were examined separately while differentially expressed targets were considered for further co-expression network construction.

### 2.3.9 Co-expression Network Reconstruction

A co-expression network represents an undirected graph where nodes of same scoring functions are supposed to have common regulation and interactions to contribute to the pathway of interest. We have developed our scoring function in this work by including codon score, transcript abundance, gene ontology score, and correlation score. By combined score of these stated parameters, we have constructed a co-expression network. This network was further investigated after mapping with other available pathway databases. Co-expression module was limited to 20 top genes as we were not receiving any addition in literature-based support. Fig 2.2 represents an algorithm for co-expression network construction.

## 2.4 Results and Discussion

Proposed miRNA analysis framework will help in miRNA identification and target prioritization by constructing miRNA specific co-expression networks. There are various advantages of regulatory networks over other random networks as they follow the behavior of scale-free networks. Alternations in miRNA level are very sensitive as they result in protein expression and ultimately affect phenotype[76]. In a given study, we have used two different conditions, i.e. JH and JV. We have tried to screen out the cause of variation in differential conditions. From miRNA regulatory networks, sub-networks and co-expression networks, we have prioritized the candidate genes for further experimental validation. The overlapping genes in both differential conditions are considered as key points to variation in cellular reprogramming in different conditions.

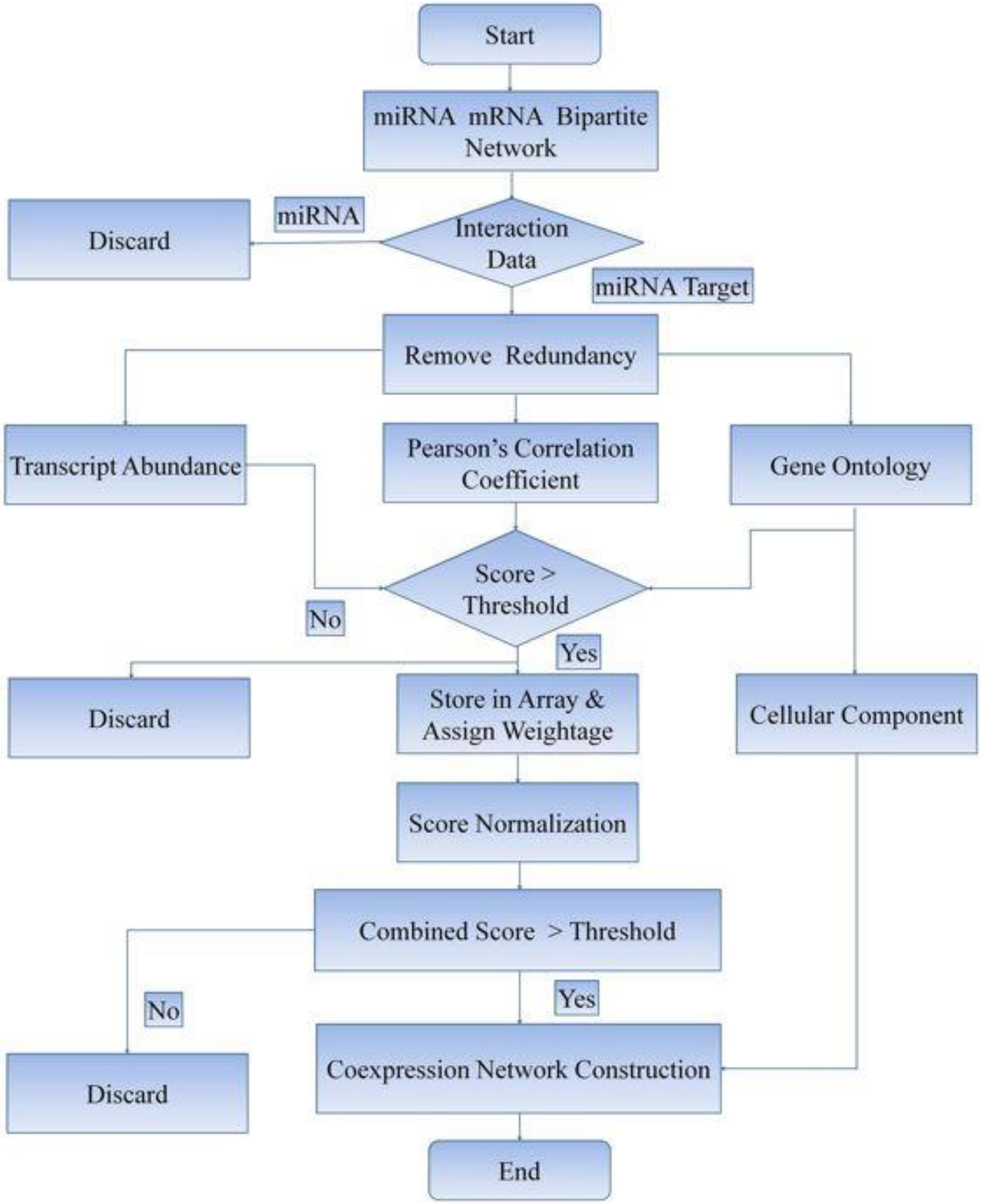


**Figure 2. 2** Methodology adopted for co-expression network construction

### 2.4.1 miRNA Target Classification

miRNA controls its targets and respective effect on gene or protein level. Various post-transcriptional modifications alter the effect of miRNA, yet miRNAs found to play a crucial role in the mechanistic regulation of various cellular and molecular processes.[77]. A total of 11 and

13 miRNAs are identified in JH and JV, respectively, of which eight are common in both, namely miR-156, miR-157, miR-159, miR-319, miR-4995, miR-5021, miR-5658, and miR-f11908 (Table 1). To gain more insight about the cellular process and biological functions we have screened the associated targets of miRNAs. Hence, we identified unique miRNAs in JH as miR-172, miR-414, and miR-529. Besides, the miR-2910, miR-2914, miR-477, miR-f11953, and miR-f12158 are identified uniquely in JV (Table 2). JH oriented targets can further be explored for resistant mechanisms while JV specific targets can be seen for genes which are supposed to involve in virus pathogenesis. miR-f11908, miR-f11953, and miR-f12158 are novel miRNAs identified by the proposed miRNA analysis framework in *J. curcas*, and these miRNAs are also not experimentally validated in other plant species. This revelation was performed by target screening from transcriptome and in total 39 targets were predicted in JH while 61 are identified in JV condition.

**Table 2. 1 miRNA targets common in healthy and diseased transcriptomic conditions**

| S.No | miRNA | JH Target Name | JV Target Name | JH FPKM | JV FPKM | Regulation (up/down) |
|---|---|---|---|---|---|---|
| 1.1 | miR-156 | choline monooxygenase [EC:1.14.15.7] | choline monooxygenase [EC:1.14.15.7] | 2092.64 | 6282.87 | ↑ |
| 1.2 | miR-156 | - | ferrochelatase [EC:4.99.1.1] | - | 114.97 | |
| 1.3 | miR-156 | - | histone H3 | - | 292.65 | |
| 1.4 | miR-156 | - | ketol acid reductoisomerase [EC:1.1.1.86] | - | 261.3 | |
| 2.1 | miR-157 | choline monooxygenase [EC:1.14.15.7] | choline monooxygenase [EC:1.14.15.7] | 2092.64 | 6282.87 | ↑ |

| | | | | | | |
|---|---|---|---|---|---|---|
| 2.2 | miR-157 | - | ferrochelatase [EC:4.99.1.1] | - | 114.97 | |
| 2.3 | miR-157 | - | ketol acid reductoisomerase [EC:1.1.1.86] | - | 261.3 | |
| 3.1 | miR-159 | acetyl-CoA carboxylase, biotin carboxylase subunit [EC:6.4.1.2 6.3.4.14] | acetyl-CoA carboxylase, biotin carboxylase subunit [EC:6.4.1.2 6.3.4.14] | 205.13 | 287.43 | ↑ |
| 3.2 | miR-159 | ubiquitin-conjugating enzyme E2 W [EC:2.3.2.25] | ubiquitin-conjugating enzyme E2 W [EC:2.3.2.25] | 253.35 | 296.57 | ↑ |
| 4 | miR-319 | ubiquitin-conjugating enzyme E2 W [EC:2.3.2.25] | ubiquitin-conjugating enzyme E2 W [EC:2.3.2.25] | 253.35 | 296.57 | ↑ |
| 5.1 | miR-4995 | RecQ-mediated genome instability protein 2 | - | 25.26 | - | |
| 5.2 | miR-4995 | small subunit ribosomal protein S5 | small subunit ribosomal protein S5 | 131.65 | 64.02 | ↓ |
| 6.1 | miR-5021 | acetyl-CoA C-acetyltransferase [EC:2.3.1.9] | acetyl CoA C acetyltransferase [EC:2.3.1.9] | 100.27 | 74.47 | ↓ |
| 6.2 | miR-5021 | alanine-glyoxylate transaminase / (R)-3-amino-2- | alanine glyoxylate transaminase / (R) 3 amino 2 | 101.03 | 155.47 | ↑ |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | methylpropionate-pyruvate transaminase [EC:2.6.1.44 2.6.1.40] | methylpropionate pyruvate transaminase [EC:2.6.1.44 2.6.1.40] | | | |
| 6.3 | miR-5021 | bud site selection protein 31 | bud site selection protein 31 | 39.04 | 33.97 | ↓ |
| 6.4 | miR-5021 | DNA polymerase epsilon subunit 2 [EC:2.7.7.7] | DNA polymerase epsilon subunit 2 [EC:2.7.7.7] | 32.91 | 35.28 | ↑ |
| 6.5 | miR-5021 | fanconi anemia group M protein | fanconi anemia group M protein | 382.71 | 475.56 | ↓ |
| 6.6 | miR-5021 | ferulate-5-hydroxylase | ferulate-5-hydroxylase | 123.23 | 189.44 | ↑ |
| 6.7 | miR-5021 | hydroxymethylglutaryl-CoA synthase [EC:2.3.3.10] | hydroxymethylglutaryl CoA synthase [EC:2.3.3.10] | 100.27 | 355.36 | ↑ |
| 6.8 | miR-5021 | mRNA export factor | mRNA export factor | 265.6 | 233.86 | ↓ |
| 6.9 | miR-5021 | nucleolar protein 58 | nucleolar protein 58 | 88.02 | 151.55 | ↑ |
| 6.10 | miR-5021 | protein disulfide-isomerase A6 [EC:5.3.4.1] | protein disulfide isomerase A6 [EC:5.3.4.1] | 141.6 | 220.8 | ↑ |
| 6.11 | miR-5021 | translation initiation factor 5B | translation initiation factor 5B | 991.98 | 2164.84 | ↑ |
| 6.12 | miR-5021 | - | (+) abscisic acid 8' hydroxylase [EC:1.14.13.93] | - | 84.92 | |
| 6.13 | miR- | - | 1 deoxy D xylulose 5 | - | 154.16 | |

|  |  |  |  |  |  |  |
|---|---|---|---|---|---|---|
|  | 5021 |  | phosphate synthase [EC:2.2.1.7] |  |  |  |
| 6.14 | miR-5021 | - | beta fructofuranosidase [EC:3.2.1.26] | - | 148.94 |  |
| 6.15 | miR-5021 | - | crossover junction endonuclease EME1 | - | 220.8 |  |
| 6.16 | miR-5021 | - | glutathione reductase (NADPH) [EC:1.8.1.7] | - | 265.22 |  |
| 6.17 | miR-5021 | - | large subunit ribosomal protein L17 | - | 33.97 |  |
| 6.18 | miR-5021 | - | peroxidase [EC:1.11.1.7] | - | 90.15 |  |
| 6.19 | miR-5021 | - | phosphoenolpyruvate carboxylase [EC:4.1.1.31] | - | 1085.69 |  |
| 6.20 | miR-5021 | - | photosystem I subunit X | - | 57.49 |  |
| 6.21 | miR-5021 | - | Ras GTPase activating protein 4 | - | 151.55 |  |
| 6.22 | miR-5021 | - | signal recognition particle subunit SRP14 | - | 52.26 |  |
| 6.23 | miR-5021 | - | small ubiquitin related modifier | - | 37.89 |  |
| 6.24 | miR-5021 | - | STIP1 homology and U box containing protein 1 | - | 53.57 |  |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | [EC:2.3.2.27] | | | |
| 6.25 | miR-5021 | - | tRNA specific 2 thiouridylase | - | 163.31 | |
| 6.26 | miR-5021 | - | ubiquinone biosynthesis monooxygenase Coq6 | - | 145.02 | |
| 7.1 | miR-5658 | 1-phosphatidylinositol-3-phosphate 5-kinase [EC:2.7.1.150] | 1-phosphatidylinositol-3-phosphate 5-kinase [EC:2.7.1.150] | 269.43 | 283.51 | ↑ |
| 7.2 | miR-5658 | bloom syndrome protein [EC:3.6.4.12] | - | 114.05 | - | |
| 7.3 | miR-5658 | diacylglycerol kinase (ATP) [EC:2.7.1.107] | diacylglycerol kinase (ATP) [EC:2.7.1.107] | 41.33 | 32.66 | ↓ |
| 7.4 | miR-5658 | DNA (cytosine-5)-methyltransferase 1 [EC:2.1.1.37] | DNA (cytosine-5)-methyltransferase 1 [EC:2.1.1.37] | 229.62 | 320.09 | ↑ |
| 7.5 | miR-5658 | large subunit ribosomal protein L9 | - | 45.16 | - | |
| 7.6 | miR-5658 | serine/threonine-protein kinase CTR1 [EC:2.7.11.1] | serine/threonine-protein kinase CTR1 [EC:2.7.11.1] | 166.09 | 271.75 | ↑ |
| 7.7 | miR-5658 | small subunit ribosomal protein S6 | small subunit ribosomal protein S6 | 143.13 | 142.41 | ↓ |

| | | | | | | |
|---|---|---|---|---|---|---|
| 7.8 | miR-5658 | transcription initiation factor TFIIF subunit alpha | transcription initiation factor TFIIF subunit alpha | 244.17 | 282.2 | ↑ |
| 7.9 | miR-5658 | transcription-repair coupling factor (superfamily II helicase) | transcription-repair coupling factor (superfamily II helicase) | 433.22 | 257.38 | ↓ |
| 7.10 | miR-5658 | translation initiation factor 5B | translation initiation factor 5B | 991.98 | 2164.84 | ↑ |
| 7.11 | miR-5658 | U4/U6.U5 tri-snRNP-associated protein 2 | U4/U6.U5 tri-snRNP-associated protein 2 | 241.11 | 244.31 | ↑ |
| 7.12 | miR-5658 | - | non lysosomal glucosylceramidase [EC:3.2.1.45] | - | 310.94 | |
| 7.13 | miR-5658 | - | peptidyl prolyl cis trans isomerase like 2 [EC:5.2.1.8] | - | 121.5 | |
| 7.14 | miR-5658 | - | RIO kinase 1 [EC:2.7.11.1] | - | 299.18 | |
| 7.15 | miR-5658 | - | serine/threonine protein phosphatase PP1 catalytic subunit [EC:3.1.3.16] | - | 84.92 | |
| 7.16 | miR-5658 | - | translation initiation factor eIF 2B subunit beta | - | 180.29 | |
| 7.17 | miR-5658 | - | xanthine dehydrogenase/oxidase [EC:1.17.1.4 | - | 326.62 | |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | 1.17.3.2] | | | |
| 8.1 | miR-f11908 | lupus La protein | lupus La protein | 120.17 | 94.07 | ↓ |
| 8.2 | miR-f11908 | splicing factor, arginine/serine-rich 4/5/6 | splicing factor, arginine/serine rich 4/5/6 | 96.44 | 91.45 | ↑ |
| 8.3 | miR-f11908 | - | hydroxymethylpyrimidine kinase / phosphomethylpyrimidine kinase / thiamine phosphate diphosphorylase [EC:2.7.1.49 2.7.4.7 - 2.5.1.3] | - | 134.57 | |

### 2.4.2 Bipartite Network Inference

To understand the JH and JV directed regulation selected miRNA and associated target transcripts were considered as a bipartite network as shown in Fig 2.3. Only 50 and 74 interacting nodes from JH and JV, respectively, were considered for network construction. Some nodes found more dominant while others do not. Bipartite network is representing the properties of the network where there will be no interaction of miRNA to miRNA regulation and also not target to target regulation. Regulation always corresponds to miRNA to target and hence depicted as directed bipartite network.

### 2.4.3 Pathway Analysis

The virus has various pathway regulation which results in disease-related phenotype. There is various molecular pathway which is specifically known to play a crucial role in virus-induced disease pathology [68]. miR-5021 and miR-5658 have major regulation in both JH and JV condition as its targets have differential and varied expression in given conditions. Some nodes, such as choline monooxygenase, histone H3, and ferrochelatase, showed an association with more than one miRNA. To understand the BPs, molecular functions (MFs), and cellular components (CCs) were further investigated.

### 2.4.4 Gene Ontology Inference

GO defines the terms in the form of BPs, MFs, and CCs for the transcripts under investigation. By node score, selected transcripts were evaluated and presented in the form of a cluster. We have used BLAST2Go, INTERPRO and PANTHER servers for crosschecking the role of miRNA targets and we have not found biased results in our analysis. The transcripts involved in BPs in JH, such as the phosphatidylinositol metabolic process, transcription from RNA polymerase II promoter, isoprenoid biosynthesis process, oxidation-reduction process, and positive regulation of the transcription elongation factor RNA polymerase II promoter, were filtered by a high GO node score (Fig 2.5). In JV, we observed number of transcripts involved in the terpenoid biosynthesis process, aerobic respiration, tricarboxylic acid cycle, an oxidative reduction process, citrate metabolic processes, organic substance biosynthetic process, and macromolecule biosynthetic processes (Fig 2.6). Literature mining was performed to identify the role of BPs and found the JV oriented pathways are basic ones while [78]–[84] while JH specific processes are involved in stress tolerant mechanisms [85]–[87].

To understand the molecular function activity in JH and JV condition, we have prioritized transcripts by node score. Catalytic activity, transferase activity, transferase acyl activity, organic cyclic compound binding, heterocyclic binding, and nucleic acid binding were observed in JH (Fig 2.5). Response to oxidative stress, primary metabolic process, organic substance metabolic process, protein ubiquitination, transcription biosynthetic process, carbon fixation, oxidation-reduction process, and tricarboxylic acid cycle was observed in JV (Fig 2.5). What MFs found in JH were indicative towards normal condition while JV specific MFs shows association with host-pathogen response[57], [68], [88]. CCs remain conserved in both the conditions but intrinsic and integral components of the membrane are uniquely found in JV condition which might be due to their role in membrane association regulation (Fig 2.6-2.7 [89].

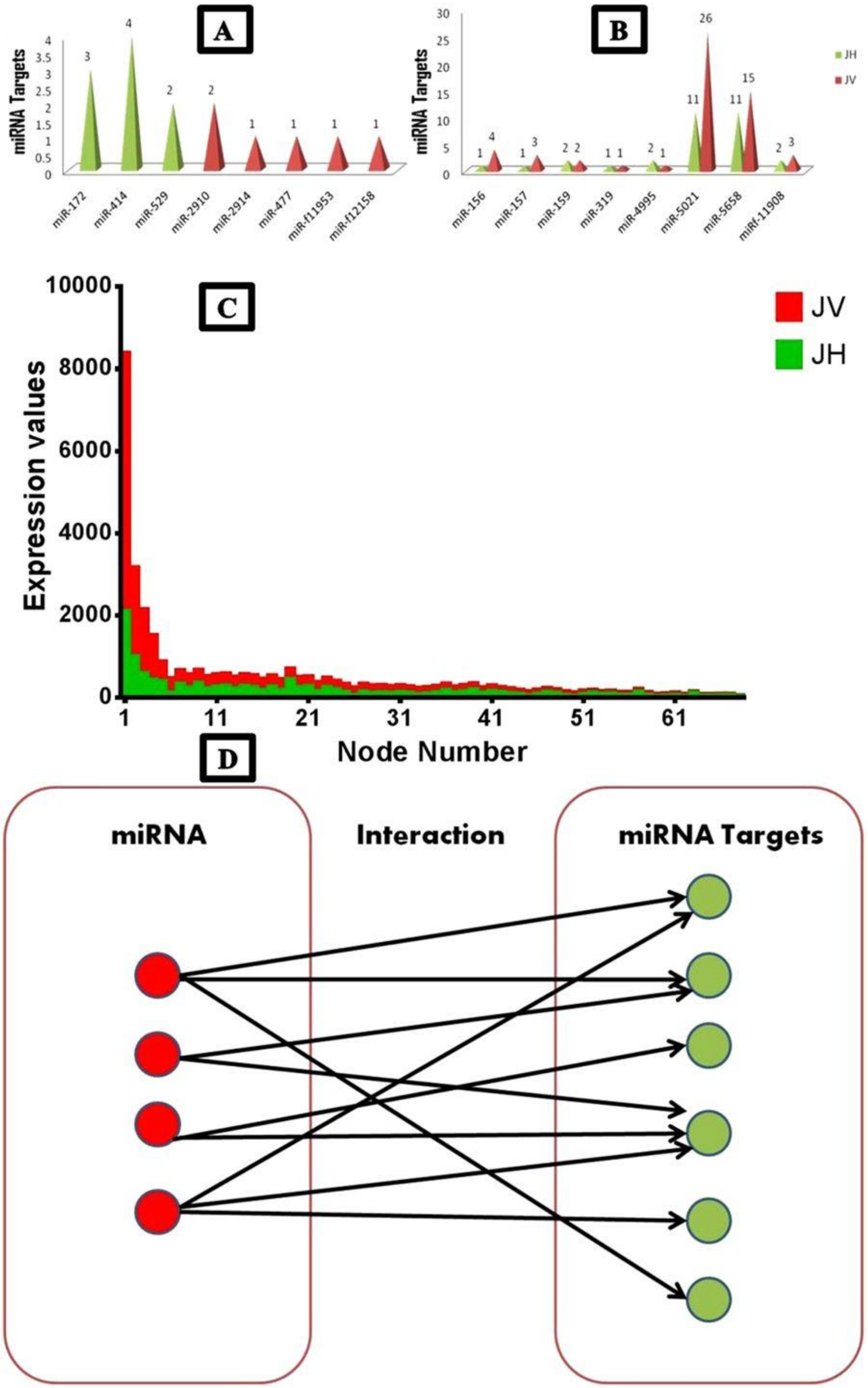


**Figure 2. 3** miRNA-mRNA interaction based target distribution (A) unique target distrubtion (B) differentially expressed targets (C) PCC analysis (D) Directed Bi-partite network pictorial representation

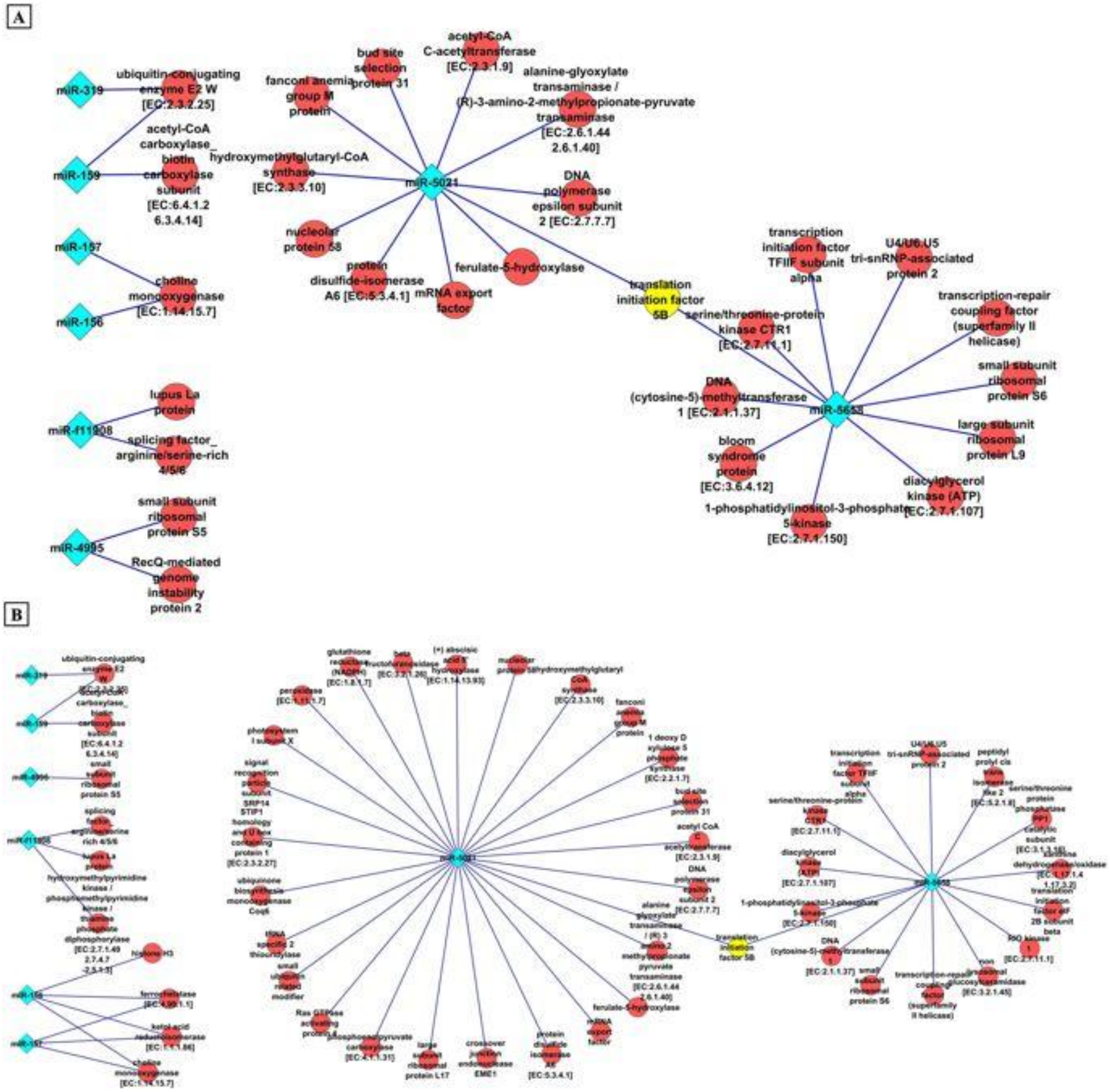

A
miR-319
ubiquitin-conjugating enzyme E2 W [EC:2.3.2.25]
miR-159
acetyl-CoA carboxylase_ biotin carboxylase subunit [EC:6.4.1.2 6.3.4.14]
miR-157
miR-156
choline monooxygenase [EC:1.14.15.7]
miR-f11908
lupus La protein
splicing factor_ arginine/serine-rich 4/5/6
miR-4995
small subunit ribosomal protein S5
RecQ-mediated genome instability protein 2
bud site selection protein 31
acetyl-CoA C-acetyltransferase [EC:2.3.1.9]
fanconi anemia group M protein
alanine-glyoxylate transaminase / (R)-3-amino-2-methylpropionate-pyruvate transaminase [EC:2.6.1.44 2.6.1.40]
hydroxymethylglutaryl-CoA synthase [EC:2.3.3.10]
miR-5021
DNA polymerase epsilon subunit 2 [EC:2.7.7.7]
nucleolar protein 58
protein disulfide-isomerase A6 [EC:5.3.4.1]
mRNA export factor
ferulate-5-hydroxylase
translation initiation factor 5B
transcription initiation factor TFIIF subunit alpha
U4/U6.U5 tri-snRNP-associated protein 2
serine/threonine-protein kinase CTR1 [EC:2.7.11.1]
transcription-repair coupling factor (superfamily II helicase)
small subunit ribosomal protein S6
DNA (cytosine-5)-methyltransferase 1 [EC:2.1.1.37]
miR-5658
large subunit ribosomal protein L9
bloom syndrome protein [EC:3.6.4.12]
diacylglycerol kinase (ATP) [EC:2.7.1.107]
1-phosphatidylinositol-3-phosphate 5-kinase [EC:2.7.1.150]
B
miR-319
ubiquitin-conjugating enzyme E2 W [EC:2.3.2.25]
miR-159
acetyl-CoA carboxylase_ biotin carboxylase subunit [EC:6.4.1.2 6.3.4.14]
miR-4995
small subunit ribosomal protein S5
miR-f11908
splicing factor arginine/serine rich 4/5/6
lupus La protein
hydroxymethylpyrimidine kinase / phosphomethylpyrimidine kinase / thiamine phosphate diphosphorylase [EC:2.7.1.49 2.7.4.7 2.5.1.3]
histone H3
miR-156
ferrochelatase [EC:4.99.1.1]
miR-157
ketol-acid reductoisomerase [EC:1.1.1.86]
choline monooxygenase [EC:1.14.15.7]
glutathione reductase (NADPH) [EC:1.8.1.7]
beta fructofuranosidase [EC:3.2.1.26]
(+) abscisic acid 8' hydroxylase [EC:1.14.13.93]
nucleolar protein 58
hydroxymethylglutaryl CoA synthase [EC:2.3.3.10]
peroxidase [EC:1.11.1.7]
fanconi anemia group M protein
photosystem I subunit X
1 deoxy D xylulose 5 phosphate synthase [EC:2.2.1.7]
signal recognition particle subunit SRP14
bud site selection protein 31
STIP1 homology and U box containing protein 1 [EC:2.3.2.27]
acetyl CoA C acetyltransferase [EC:2.3.1.9]
miR-5021
DNA polymerase epsilon subunit 2 [EC:2.7.7.7]
ubiquinone biosynthesis monooxygenase Coq6
alanine glyoxylate transaminase / (R) 3 amino 2 methylpropionate pyruvate transaminase [EC:2.6.1.44 2.6.1.40]
tRNA specific 2 thiouridylase
small ubiquitin related modifier
ferulate-5-hydroxylase
Ras GTPase activating protein 4
mRNA export factor
phosphoenolpyruvate carboxylase [EC:4.1.1.31]
large subunit ribosomal protein L17
crossover junction endonuclease EME1
protein disulfide isomerase A6 [EC:5.3.4.1]
translation initiation factor 5B
transcription initiation factor TFIIF subunit alpha
U4/U6.U5 tri-snRNP-associated protein 2
peptidyl prolyl cis trans isomerase like 2 [EC:5.2.1.8]
serine/threonine-protein kinase CTR1 [EC:2.7.11.1]
serine/threonine protein phosphatase PP1 catalytic subunit [EC:3.1.3.16]
diacylglycerol kinase (ATP) [EC:2.7.1.107]
miR-5658
xanthine dehydrogenase/oxidase [EC:1.17.1.4 1.17.3.2]
1-phosphatidylinositol-3-phosphate 5-kinase [EC:2.7.1.150]
translation initiation factor eIF 2B subunit beta
DNA (cytosine-5)-methyltransferase 1 [EC:2.1.1.37]
PIG kinase 1 [EC:2.7.11.1]
small subunit ribosomal protein S6
transcription-repair coupling factor (superfamily II helicase)
non lysosomal glucosylceramidase [EC:3.2.1.45]

**Figure 2. 4** Bi-partite network representing (A) healthy target condition in and (B) diseased target conditions

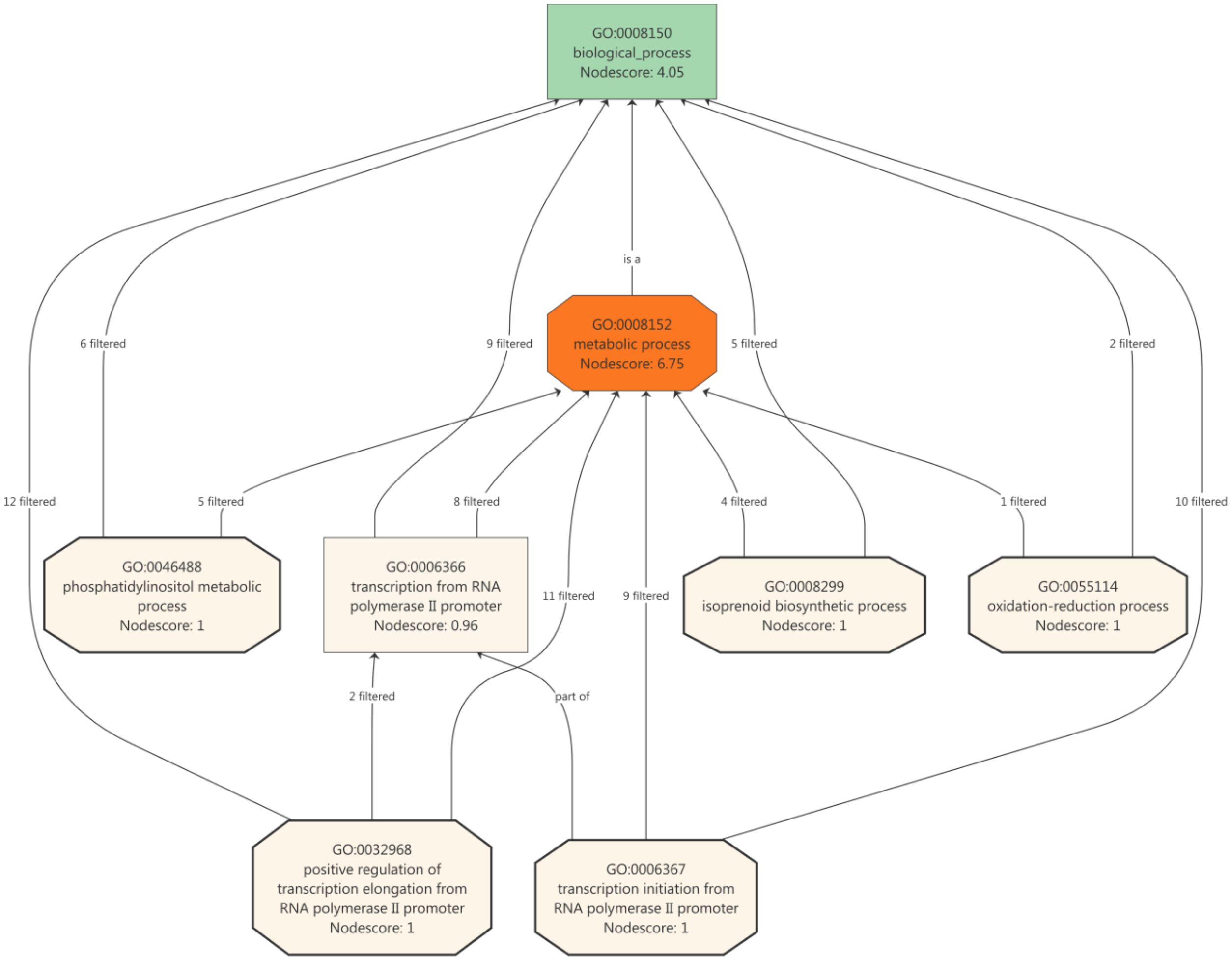


**Figure 2. 5** Biological processes involved in healthy conditions

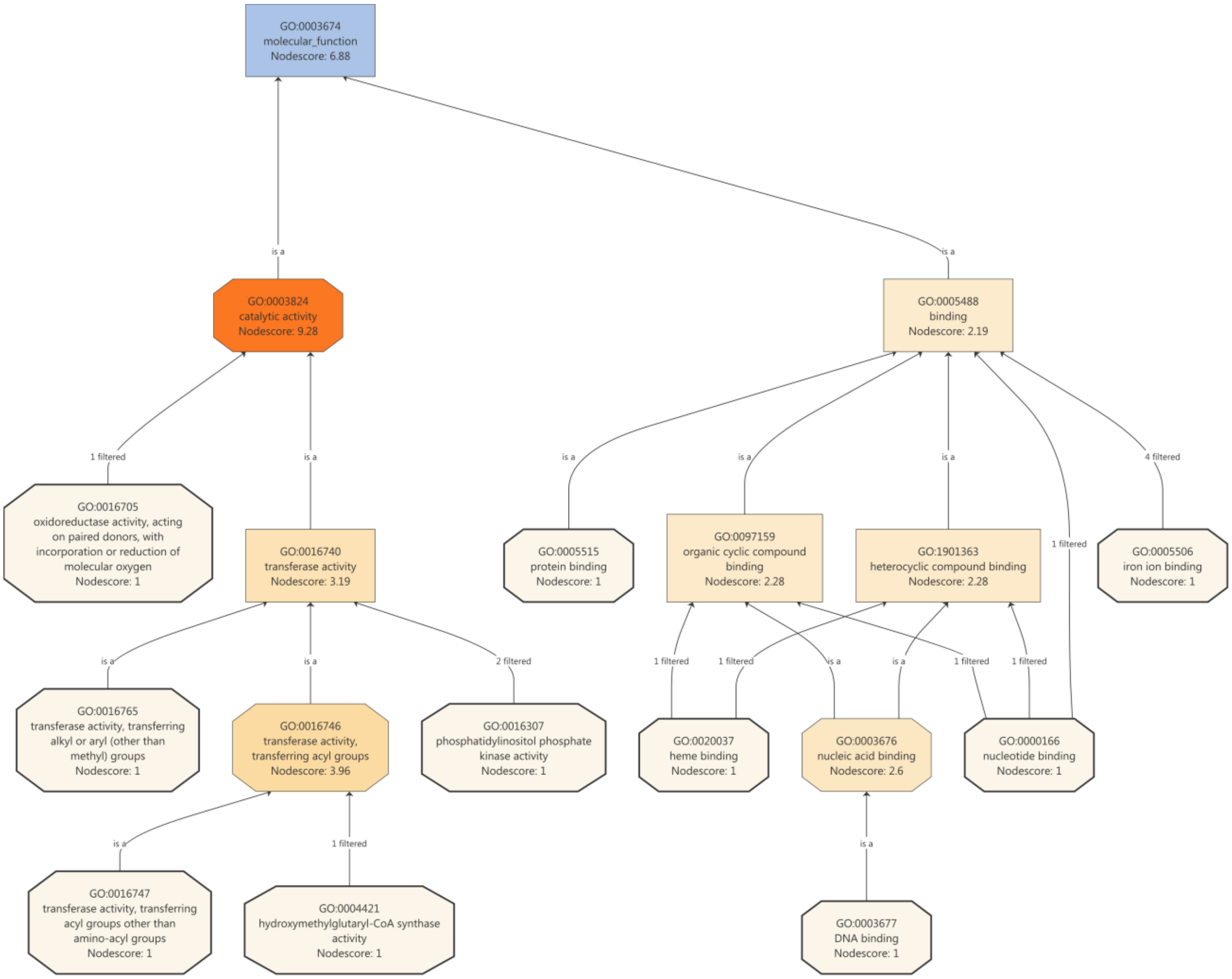


**Figure 2. 6** Molecular functions involved in healthy condition

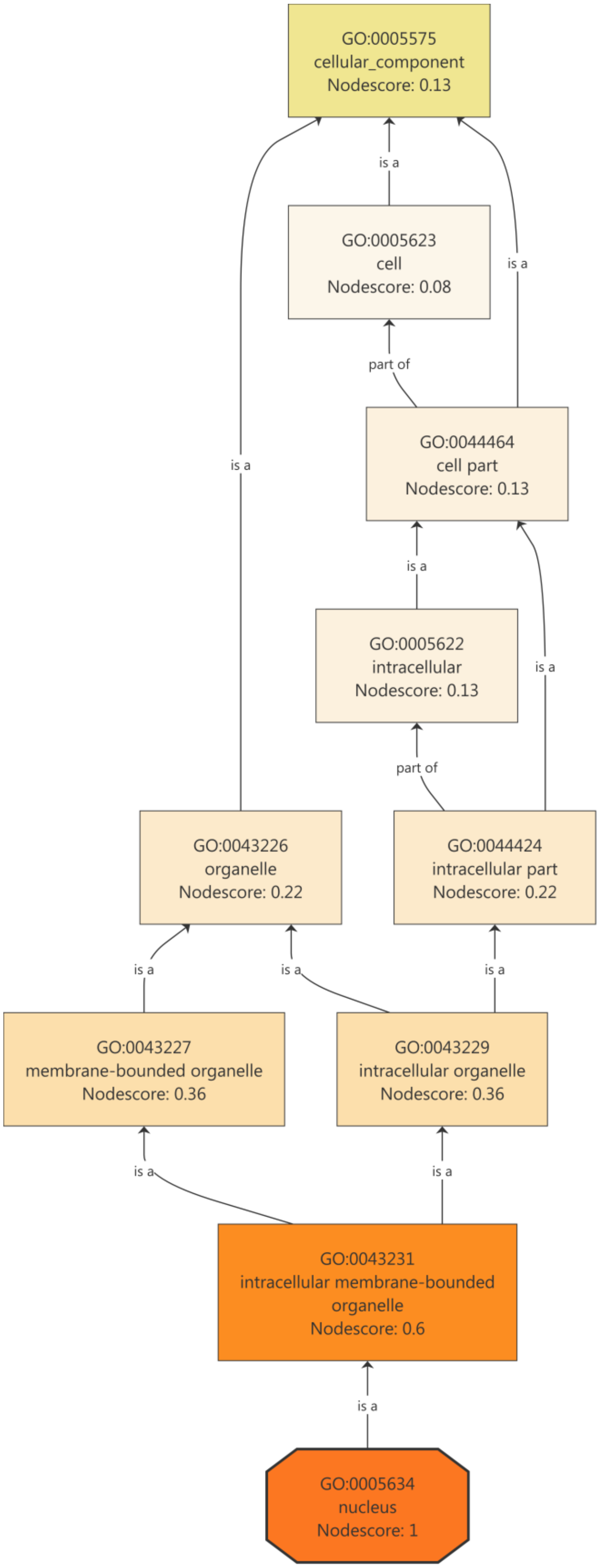


**Figure 2. 7** Cellular components involved in heathy conditions

*For global network of healthy and virus condition please check at http://bioinfoindia.org/miRAF/gene_ontology.html

**Table 2. 2** miRNAs, target genes, co-expressed genes and associated pathway (For more details see Supplementary File)

| S. No. | miRNAs | Target Genes | Top Co-expressed genes | Co-expressed genes contributing Pathways |
|---|---|---|---|---|
| 1 | miR-156 | CMO | RCC1,DUF1624,NF-YC12,MIR5344,3767731,hydrolase, kinase,tudor-like, CAMP,CHT-type C,SCAMP,MIR-834a,MBD7,TRAF-Like,NHX5,BET10,TAF6B4,DUF2358,PLDGAMMA2,TIR-NBS-LR,MRS2-7,Phosphoester | Biosynthesis of Secondary Metabolites |
| 2 | miR-157 | | | Carbon Metabolism |
| 3 | miR-4995 | RPS5 | TIR-NBS-LR,ENTH,NLP7,ARM repeat, CC-NBS-LRR,PP2-A7,PP2-A6,KINASE,RFL1,F-box,Calmodulin,RPP4,,RPP5,SNC1,LRR,RLM3 | Disease Resistance Response |
| 4 | miR-5658 | RPL9 | Emb1473,RPL15,PSRP5,EMB3105,S10p/S20e,PRPL11,L19,RPS17,L28,ribosome,EMB3113, Heavy metal ion,L5P,TWN3,GHS1,ROC4,S20,emb2394,NDPK2,RPL21C | Ribosomal Machinery |
| | | EIF5 | Hydrolase,inhibitor,CYN,CUTA,NAT,819216,UBC30,GB2,G18 | Protein Processing in Endoplasmic Reticulum |

| | | | | |
|---|---|---|---|---|
| 5 | miR-5021 | | a,GRXC2,OB-Fold Ligand,TRXH3,UBQ7,ADF6,UBC3,CHMP1A,W1H1,G8B,UBC11 | Ubiquitin |
| | | MVA1 | ACP1,MOD1,PLE2,KASI,Thioesterase,BIOB,840894,CAC2,FPS1,MVD1,EMB1276,URH2,hydrolase,mutase,FaTA,NagB,UPF0041,ZHD13,CAC1-B | Biosynthesis of Secondary Metabolites |
| | | | | Fatty Acid Biosynthesis |
| | | | | Fatty Acid Metabolism |
| | | | | Carbon Metabolism |
| | | PPC | SOS1,MMT,alpha/beta subunit,kinase,PFK7,iPGAM2,PGM3,MDAR1,PGM2,Galactose,UGP1,mMDH2,HXK1,RR10,GLU2,JAR1,PGDH,ACO3,EMB1467,Kinase | Biosynthesis of Secondary Metabolites |
| | | | | Carbon Metabolism |
| | | | | Glycolysis/ Gluconeogenesis |
| | | | | Galactose Metabolism |
| | | PSAX | PDAD-2,PSII,Photosynthesis,NdhS,LHCA1,PSAH-1,PSAL,LHCA3,PSAG,YCF32,PSAF,PSBW,LHCB5,PSBX,PSAN,PSAE-1,838749,PSAD-1,PSII-Q | Photosynthesis |
| | | | | Photosynthesis-Antenna Proteins |
| | | SUMO | HMGA,HTB1,HTB2,819315,TYRDC1,alpha/beta ligand,PEL3,CYP7731,LTP6,PIP2D,hydrolase,kinase,inhibitor,DRG,PME5,PDCB2,Putative mutase, | Ribosomal Machinery |
| | | | | Sliceosome |
| | | CYP707A1 | AFP1,AFP3,SAG113,AB12,RAB18-PUB19,PP2-B11,SPSA2,BETAVPE,LEA7,L | Plant Hormone Signal Transduction |

| | | | | |
|---|---|---|---|---|
| | | | EA,LEA4-5,TSPO,transporter,RD29B,phosphotriase,XERO2,FMO-GS-OX-4,ESL1 | |
| | | DXS | GUN5,OSA1,JAC1,821278,CH1,SIGB,PSY,CP5,815980,DUF2358,COLA4,PSII,PSAD2,LHCB6,CRD1,oxidoreductase,HEMA1,818819,SLP1,rosamann | Biosynthesis of Secondary Metabolites |
| | | | | Photosynthesis-antenna Proteins |
| | | | | Porphrin and Chlorophyll Metabolism |
| | | | | Proteosome |
| | | | | mRNA Surveillance Pathway |

### 2.4.5 Correlation Analysis

We have used parametric and non parametric test to understand the correlation of targets in differential JH and JV conditions. PCC and Spearman's correlation were used to get an idea about standard deviation from the central genetic machinery. But, only PCC showed relevant results with our analysis. Only common differentially expressed targets were selected for this analysis to trace out the variations. As shown in Fig 10, miRNA target transcripts showed higher expression profiles in JV compared to JH transcriptome derived targets.

The analysis of PCC results revealed ten genes, namely *CMO* (miR-156 and miR-157), *RPS5* (miR-4995), *RPL9* (miR-5658), *EIF5* (miR-5658 andmiR-5021), *MVA1*, *PPC*, *PSAX*, *SUMO*, *CYP707A1*, and *DXS* (miR-5021), which were further considered in co-expression network construction. To further analyze the role of co-expressed genes, they were mapped with KEGG to identify pathways in which all these genes were intolved **[90]**. *CMO*-associated co-expressed genes were found to be involved in the biosynthesis of secondary metabolites and carbon metabolism; *RPS5*-related co-expressed genes in disease resistance response; *RPL9*-related co-expressed genes in ribosomal machinery, *EIF5*-related co-expressed genes in protein processing in the endoplasmic reticulum and ubiquitin-mediated proteolysis; *MVA1*-related co-expressed genes in the biosynthesis of secondary metabolites, fatty acid biosynthesis, carbon metabolism, and fatty acid metabolism; *PPC*-related co-expressed genes in the biosynthesis of secondary metabolites, carbon metabolism, glycolysis/gluconeogenesis, and galactose metabolism; *PSAX*-related co-expressed genes in photosynthesis and photosynthesis-antenna proteins; *SUMO*-related co-expressed genes in ribosome machinery and spliceosome; *CYP707A1*-related co-expressed genes in plant hormone signal transduction; and *DXS*-related co-expressed genes in the biosynthesis of secondary metabolites, photosynthesis-antenna proteins, prohrine and chlorophyll metabolism, proteosome, and the mRNA surveillance pathway. From the results of the co-expression network construction, it can be deduced that co-expressed genes are involved in various subcellular processes, which may be controlled or altered by miRNA regulation.

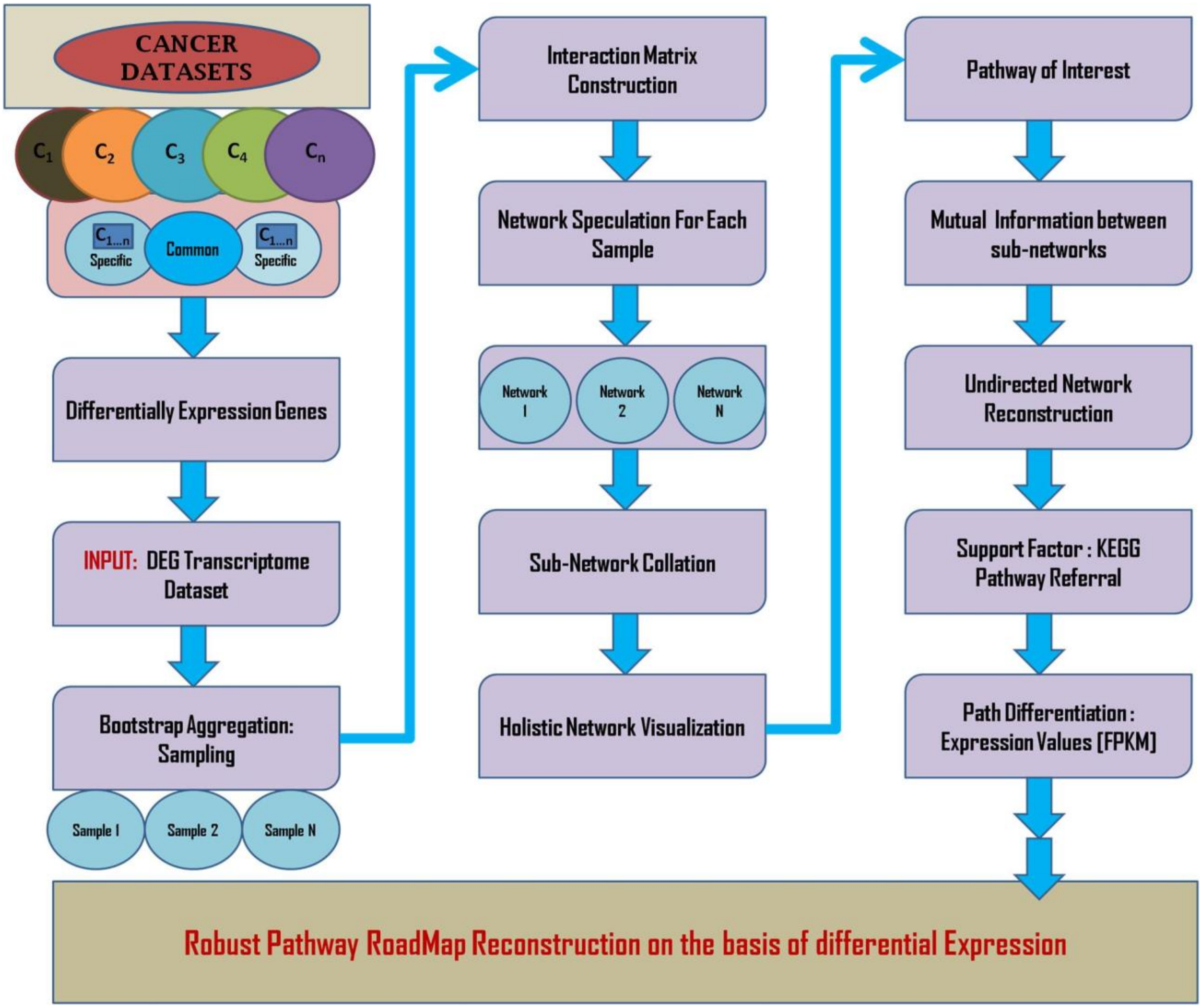
The proposed framework depicts a novel analysis to discern complex data regarding omics domain. Moreover, this framework can be used to trace regulation mechanisms derived from miRNA, transcription factors and cis or trans-regulatory elements. This framework can also provide insight into comparative and differential analysis for n number of datasets. To understand the implication of the stated framework, we have successfully demonstrated the JH and JV condition based miRNA analysis using network biology. This framework will help experimental biologists to trace down candidate biomarkers in their study of interest.

## 2.5 Conclusion

Use of miRNA identification and target prediction techniques gives input to construct biological networks. Moreover, the function level annotation of selected targets can be inferred under network biology framework. Correlation between genotype and phenotype can be established by using various enrichment analysis. With the availability of large-scale omics data, identification of candidate genes for future interventions becomes comparatively easy. Predicted targets through our miRNA analysis framework can further experimentally be validated. This framework can aid in the comparative transcriptomic analysis to understand key regulatory mechanisms and identifying regulatory nodes.

# CHAPTER 3

## A system level network analysis framework to snare transitions in co-expression networks


CANCER DATASETS
C1
C2
C3
C4
Cn
C1...n Specific
Common
C1...n Specific
Differentially Expression Genes
INPUT: DEG Transcriptome Dataset
Bootstrap Aggregation: Sampling
Sample I
Sample 2
Sample N
Interaction Matrix Construction
Network Speculation For Each Sample
Network I
Network 2
Network N
Sub-Network Collation
Holistic Network Visualization
Pathway of Interest
Mutual Information between sub-networks
Undirected Network Reconstruction
Support Factor : KEGG Pathway Referral
Path Differentiation : Expression Values [FPKM]
Robust Pathway RoadMap Reconstruction on the basis of differential Expression

## 3.1 Abstract

Molecular biology aim at stalking cellular and molecular reasons of biological functions. The identification of the relation between molecular variations and their respective phenotypes is a difficult work for genetic engineering. A huge amount of data is getting generated from the past decade, but an understanding of molecular mechanism underlying signaling cascade remains a mystery. Generating condition-wise or variation-wise big data undoubtedly answered various questions related to differential expression. But, still, there is a need to exploit available data in such fashion so that we can maximize the level of information retrieval from various data layers.

Interactions at molecular and cellular levels can be represented using transcriptional and gene regulatory networks. Henceforth, various regulatory network based analysis can suggest potential interaction and patterns which might be useful in biological context. In this chapter, we are proposing a method which is based on ensemble methods. We have used caner metastasis gens for biological network reconstruction. Random sampling has been performed for network speculation using interaction matrix. This method will be add-on to the previously existing approaches to global network construction and then moving forward to pathway of interest by node prioritization.

## 3.2 Introduction

Incredible interdisciplinary research in the last few decades has aimed to interpret the biological entities of diseases since elucidation of the molecular regulatory mechanisms provides great assistance in pharmaceutical and medical fields [91]. But, differential regulatory functions in disorders like cancer, diabetes, myocardial infarction makes the deciphering of mechanisms involved in gene dysregulation remains remotely unrevealed. Network biology has paved the path for the exploration of genome-wide gene expression characteristics to further transform basic genomic information into valuable knowledge in medical and biological research fields [74].

On a theoretical note, researches in past have indicated that networks from a broad range of application areas share common structural properties offering hope that such tools may reflect useful applications in a wide variety of disciplines [41]. Generally speaking, three classes of bio-molecular networks have fascinated most interest till date: protein-protein interaction (PPI) networks by the physical interactions; the transcriptional regulatory networks which depict the regulatory interactions between a set of genes; and metabolic networks of biochemical reactions to understand flux diversion [34].

Recent development in the high-throughput RNA-Seq techniques in molecular biology has shed light to an unprecedented quantity of data available on key cellular and molecular networks in a diversity of simple organisms [56]. However, swift development of transcriptomic technologies, as compared to other systematic platforms, has supported a variety of studies on environmental and genetic perturbations at the transcriptomic level in several organisms [92]. In this perspective, networks have swiftly become an eye-catching approach to handle, show and contextualize these bulky data sets with the intention of obtaining a molecular and system level understanding of molecular mechanisms and key biological processes [93]. As a result, it is becoming difficult to understand the mechanistic role of the regulator concerning big data. Hence, there is a need to decipher regulatory interactions with an alternate perspective which involves analysis of sub-networks resulting in a holistic view of the biological system. This can be achieved through construction and interpretation of sub-networks involving regulation across cellular, molecular and biological levels with their key genes that can be resolved using genes regulatory networks.

Gene regulatory networks are usually modeled regarding directed graphs where a node represents the genes and edges represent the interaction among genes [94]. For instance, node A regulates node B and stating direct regulation concerning directed graphs which starts from node A and ends at node B with or without any additional factor. However, from the implication point of view, it becomes unviable to understand regulatory interactions of missing links present in biological data. To decipher the missing links and undirected association between scattered dataset, it becomes indispensable to consider a viewpoint of the undirected graph [92]. In addition to directed and undirected graphs, one must deduce statistical inference regarding the network as affirmed in BC3NET module derivative of ensemble C3NET program. Apart from BC3NET, there are various other methods developed such as Boolean method, GENIE3, ARACNe and other bayesian methods. But, these methods face problem while reconstruction of network roadmaps in interest-specific pathways by linking the transcriptomic gene expression data with correlation analysis henceforth making the statistical study more complex [36].

The following paper brings in a novel approach to construe network for gene regulatory data and unwire the genomic regulatory interaction regarding information more than following a traditional protocol. To serve the purpose, a sampling of data using bootstrap aggregation followed by graph reconstruction and null hypothesis testing of sub-networks based on mutual information sharing was performed on the input dataset. Moreover, intervention done manually in threshold with a change in sub-network while reconstructing the complete map was achieved via employment of parametric and non-parametric tests. To display the effectiveness of the above stated framework, we performed cancer metastasis network construction to deduce pathway of interest.

## 3.3 Methods

### 3.3.1 Data Assortment

To understand the network reconstruction and diversion at specific pathway of interest, we have considered cancer metastasis network reconstruction. Metastasis is a complex and deadly process in which primary tumors circulate or distribute to secondary organs. It is very difficult to classify the genes under metastasis category due to various complications at the signaling level. The complete understanding of this process needs information of activation of metastasis promoter and suppressor genes in tumors. Metastasis suppressor genes affect various cellular mechanisms and control the regulations of various processes which are significant for clinical studies. Henceforth, we have collected data from Metastasis Suppressor Gene Database (MSGene). MSGene is a knowledge base containing all the gene level information associated with metastasis suppressor genes. We have considered another database Human Cancer Metastasis Database (HCMDB) for cross-validation and association studies results with literature support.

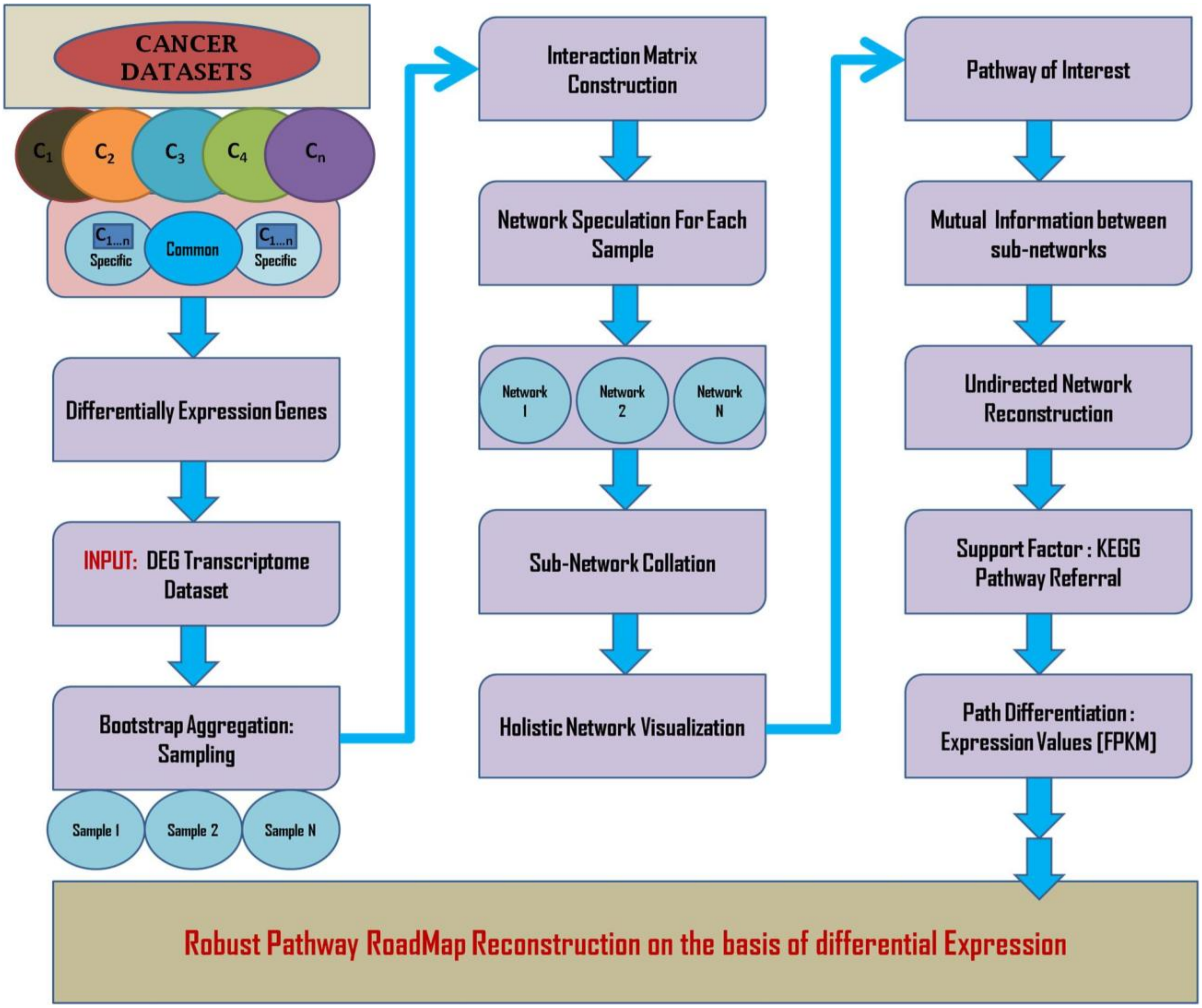

CANCER DATASETS
C1
C2
C3
C4
Cn
C1…n Specific
Common
C1…n Specific
Differentially Expression Genes
INPUT: DEG Transcriptome Dataset
Bootstrap Aggregation: Sampling
Sample 1
Sample 2
Sample N
Interaction Matrix Construction
Network Speculation For Each Sample
Network 1
Network 2
Network N
Sub-Network Collation
Holistic Network Visualization
Pathway of Interest
Mutual Information between sub-networks
Undirected Network Reconstruction
Support Factor : KEGG Pathway Referral
Path Differentiation : Expression Values [FPKM]
Robust Pathway RoadMap Reconstruction on the basis of differential Expression

**Figure 3. 1** Robust pathway roadmap reconstruction on the basis of differential expression analysis

### 3.3.2 Data Normalization and Co-expression Inference

MSGene data were classified into different cancer type and five cancer types along with a number of metastasis suppressor genes viz. Breast Cancer (77), Lung Cancer (43), Colorectal Cancer (43), Prostate Cancer (34), and Liver Cancer (26) were considered for module level analysis by availability of data using threshold value 25. Using these five different gene datasets, we have downloaded co-expressed genes from STRING database. Further, we have performed the comparison in between selected cancer types to identify common and unique genes.

### 3.3.3 Network Construction

Network-based representation and investigation of information are progressively being utilized in both depictions and recognize the segments from high-throughput exploratory innovations and their communications engaged with a given cell framework — overall methodology using mathematical and graphical inference shown in Figure 2.

Generally, common data based gene regulatory network inference strategies comprises three foremost steps. In the initial step, a common data matrix is achieved in light of common data estimates for all conceivable gene pairs in a gene articulation data set. In the second step, a speculation test is performed for each common information value estimate. At long last, in the third step, a gene regulatory network is construed from the noteworthy common data esteems, as indicated by a technique particular method.

The basic initiative of our method is to create an assembly of $A$ autonomous bootstrap datasets $\left\{D_k^b\right\}_{k=1}^{A}$ from one dataset $D(t)$, comprising of $t$ samples, by examining from/with substitution by utilizing a non-parametric bootstrap [36] with $A = 1000$. At that point, for each produced informational collection $D_k^b$ in the group, a network $N_k^b$ is derived by utilizing other methods.

From the assembly of networks, we develop one weighted network

$$N_k^b \overset{aggregate}{\leftarrow} \left\{N_k^b\right\}_{k=1}^{A} \tag{1}$$

which is utilized to decide the factual centrality of the association between gene pairs.

In brief, the method comprises three primary steps. To begin with, common data esteem among all gene pairs is evaluated. In the second step, an extreme selection strategy is connected to permit every one of the $P$ genes in an offered dataset to contribute at most one edge to the inferred

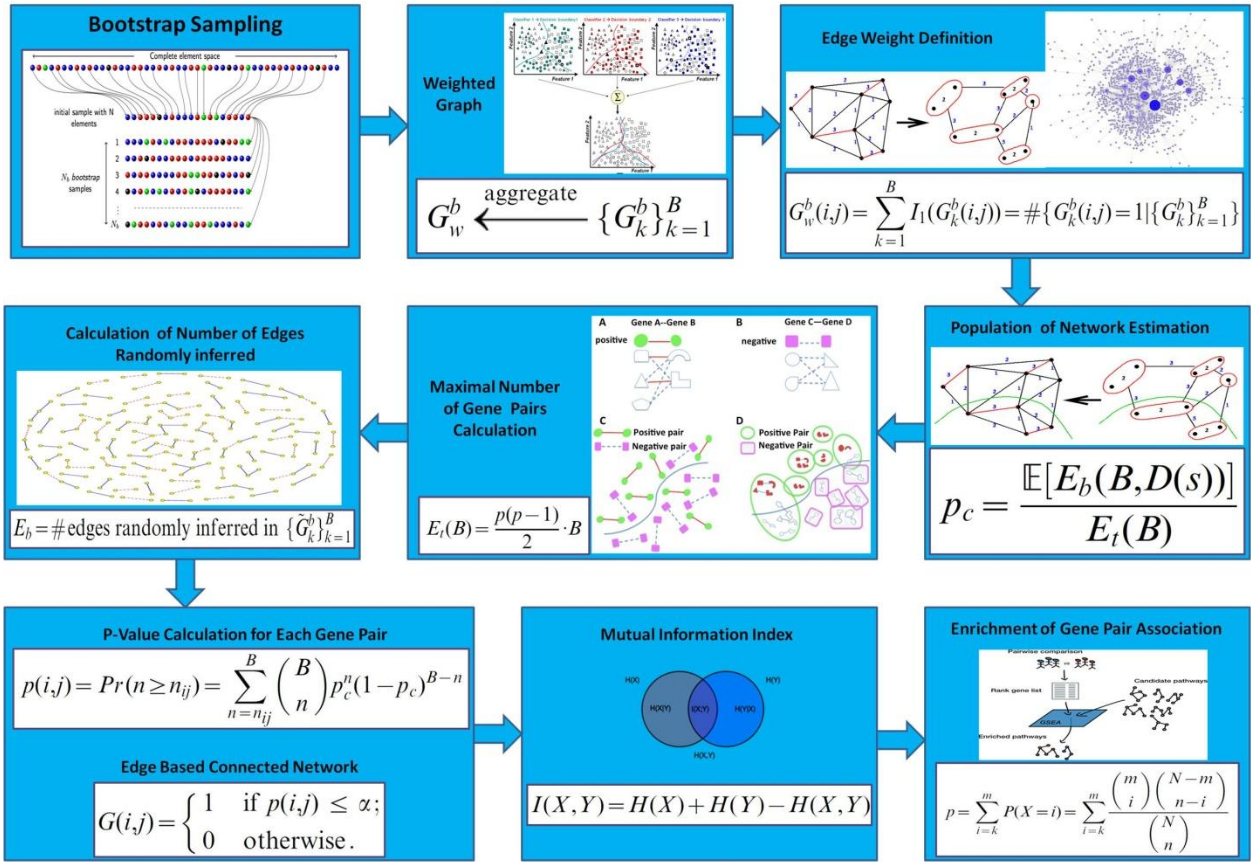

Bootstrap Sampling
Complete element space
initial sample with N elements
1
2
3
4
$N_b$
$N_b$ bootstrap samples
Weighted Graph
Classifier 1 → Decision boundary1
Classifier 2 → Decision boundary 2
Classifier 3 → Decision boundary 3
Feature 1
Feature 2
$\Sigma$
$G_w^b \xleftarrow{\text{aggregate}} \{G_k^b\}_{k=1}^B$
Edge Weight Definition
$G_w^b(i,j) = \sum_{k=1}^{B} I_1(G_k^b(i,j)) = \#\{G_k^b(i,j)=1|\{G_k^b\}_{k=1}^B\}$
Population of Network Estimation
$p_c = \frac{\mathbb{E}[E_b(B,D(s))]}{E_t(B)}$
Maximal Number of Gene Pairs Calculation
A
Gene A--Gene B
positive
B
Gene C—Gene D
negative
C
Positive pair
Negative pair
D
Positive Pair
Negative Pair
$E_t(B) = \frac{p(p-1)}{2} \cdot B$
Calculation of Number of Edges Randomly inferred
$E_b = \#\text{edges randomly inferred in } \{\tilde{G}_k^b\}_{k=1}^B$
P-Value Calculation for Each Gene Pair
$p(i,j) = Pr(n \geq n_{ij}) = \sum_{n=n_{ij}}^{B} \binom{B}{n} p_c^n (1-p_c)^{B-n}$
Edge Based Connected Network
$G(i,j) = \begin{cases} 1 & \text{if } p(i,j) \leq \alpha; \\ 0 & \text{otherwise}. \end{cases}$
Mutual Information Index
H(X)
H(Y)
H(X|Y)
I(X;Y)
H(Y|X)
H(X,Y)
$I(X,Y) = H(X) + H(Y) - H(X,Y)$
Enrichment of Gene Pair Association
Pairwise comparison
Rank gene list
Candidate pathways
GSEA
Enriched pathways
$p = \sum_{i=k}^{m} P(X=i) = \sum_{i=k}^{m} \frac{\binom{m}{i}\binom{N-m}{n-i}}{\binom{N}{n}}$

**Figure 3. 2** Graphical and mathematical representation of proposed method

network. That implies we have to test just/extraordinary hypotheses and not $\frac{p(p - 1)}{2}$. This potential edge relates to the hypothesis test that should be directed for every one of the $P$ genes. In the last step, different testing methodologies are useful to control the type one error. In the above-portrayed setting, this outcome in a system In the above-described context, this outcome in a network $N_k^b$.

To test the factual centrality of the association between gene pairs method uses the edge weights of the aggregated network $N_w^b$ as test statistics. The edge weights of $N_w^b$ are component-wise characterized by

$$N_w^b(i,j) = \sum_{k=1}^{A} I_1(N_w^b(i,j)) = \#\{N_w^b(i,j) = 1|\{N_k^b\}_{k=1}^{A}\} \qquad (2)$$

Here $I()$ the indicator function which is given by

$$I = \begin{cases} 1 & if \;\; N_k^b(i,j) = 1 \\ 0 & otherwise \end{cases}$$

This articulation relates to the number of networks in $\{N_k^b\}_{k=1}^{A}$ which have an edge amongst gene $i$ and $j$. For quickness, we write in the accompanying$n_{ij} = N_w^b(i,j)$. From eq.(2), it takes after that $n_{ij}$ assumes integer values in $\{0,1,2,...,A\}$. Given the test measurement$n_{ij}$, we formulate the following null hypothesis which we test for each gene pair$(i,j)$.

$H_0^{n_{ij}}$ : The number of networks $n_{ij}$ in the ensemble $\{N_k^b\}_{k=1}^{A}$ with an edge between gene $i$ and $j$ is less than $n_0(\alpha)$.

Here the cut-off esteem $n_0$ relies upon the criticalness level α. Because of the freedom of the bootstrap datasets, we expect the null distribution of $n_{ij}$ to take after a binomially distributed$Bin(A,p_c)$, while $A$ compares to the span of the bootstrap group and $p_c$ is the likelihood that two genes are associated by a shot. The parameter $p_c$ a identifies with a populace of systems, assessed from randomized information by utilizing the method, and compares to the portion of arbitrarily gathered edges in the bootstrap population $(E[E_b(A,D(t))])$ partitioned by the aggregate number of conceivable edges in this population $(E_t(A))$ that implies

$$p_c = \frac{E[E_b(A,D(t))]}{E_t(A)} \tag{3}$$

The maximal number of gene pairs that can be framed from $p$ genes in $A$ bootstrap datasets is given by

$$E_t(A) = \frac{p(p-1)}{2}B \tag{4}$$

This esteem is autonomous of the example estimate. $E\left[E_b(A,D(t))\right]$ Compares to the desire estimation of the quantity of arbitrarily deduced edges for a populace of an outfit of bootstrap datasets of size $A$. Since $E_b(A,D(t))$ is an irregular variable it is important to normal overall conceivable bootstrap datasets of size $A$ with test measure $t$. On a hypothetical note, we comment that these bootstrap datasets constitute a populace that indicates a likelihood mass capacity (pmf) for which the desire for $E_b(A,D(t))$ should be assessed. Because of the way that this pmf is obscure the estimation of $E\left[E_b(A,D(t))\right]$ should be evaluated.

Keeping in mind the end goal to evaluate $E\left[E_b(A,D(t))\right]$ we randomize the information to gauge the number of edges haphazardly deduced in a bootstrap group of size $A$,

$$\left\{\breve{G}_k^b\right\}_{k=1}^{A} E_b = \# \text{ Edges randomly inferred in } \left\{\breve{G}_k^b\right\}_{k=1}^{A} \tag{5}$$

Using $E_b \approx E\left[E_b(A,D(t))\right]$ as a module estimator for Eqn. Three we get a gauge for $p_c$. This enables us to ascertain a p-esteem for every gene pair $(i,j)$ and a given test measurement $n_{ij}$, given by Eqn. 2, from the null distribution of $n_{ij}$ by

$$p(i,j) = \Pr\left(n \geq n_{ij}\right) = \sum_{n=n_{ij}}^{A} \binom{A}{n} p_c^n (1 - p_c)^{A-n}$$ 6)

Here $p(i,j)$ is the likelihood to watch $p(i,j)$ or more edges by chance in a bootstrap group of size $A$ and test measure $t$.

Since we have to test $\frac{p(p-1)}{2}$ speculations at the same time (one for every gene pair), we have to apply numerous testing remedies (MTC). For our examination, we are utilizing a Bonferroni strategy for a solid control of the family-wise blunder rate (FWER). Regularly, techniques controlling the FWER are more traditionalist than systems controlling, e.g., the false revelation rate (FDR) by making just gentle suspicions about the basic information.

In light of these theories tests, the final network $N$ is componentwise characterized by

$$N(i,j) = \begin{cases} 1 & if\ p(i,j) \leq \alpha \\ 0 & otherwise \end{cases} \tag{7}$$

That implies if the association between a gene combine is measurably critical they are associated by and edge, generally there is no association. Null-distribution of shared data esteems. Keeping in mind the end goal to decide the factual centrality of the shared data esteems between genes, we test for each match of genes the accompanying null hypothesis.

$H_0^I$: The shared data amongst genes $i$ and $j$ is zero.

Since we are utilizing a nonparametric test, we have to get the relating null distribution for/from randomization of the information. Essentially, there are a few different ways to perform such randomization which fit in with the detailed null hypothesis. Consequently, we perform diverse randomizations and contrast the acquired outcomes with reference to the execution of the surmising strategy to choose the most proper one. Two randomization plans (RM1 and RM2) permute the articulation profiles for every gene pair independently. RM1 permutes just the example marks, and RM2 permutes the example and the gene names.

Conversely, the randomization conspires RM3 permutes the example and genes marks for all genes of the whole articulation grid on the double.

### 3.3.4 miRNA: a controlling network system

Once the network is reconstructed, the major task is to understand its regulation. There are various inbuilt mechanisms in a cellular system which controls the biological processes. Sometimes, these regulations are not taking place precisely due to various genetic alterations. miRNA control many genes at a time, which makes it an interest of systems biologists. We have referred our miRNA analysis framework to redraw the mechanistic control for metastasis network.

## 3.4 Results and Discussion

To get insight into the holistic view through scattered and unaligned data, literature mining and database derived data was considered to identify known metastasis suppressor genes.

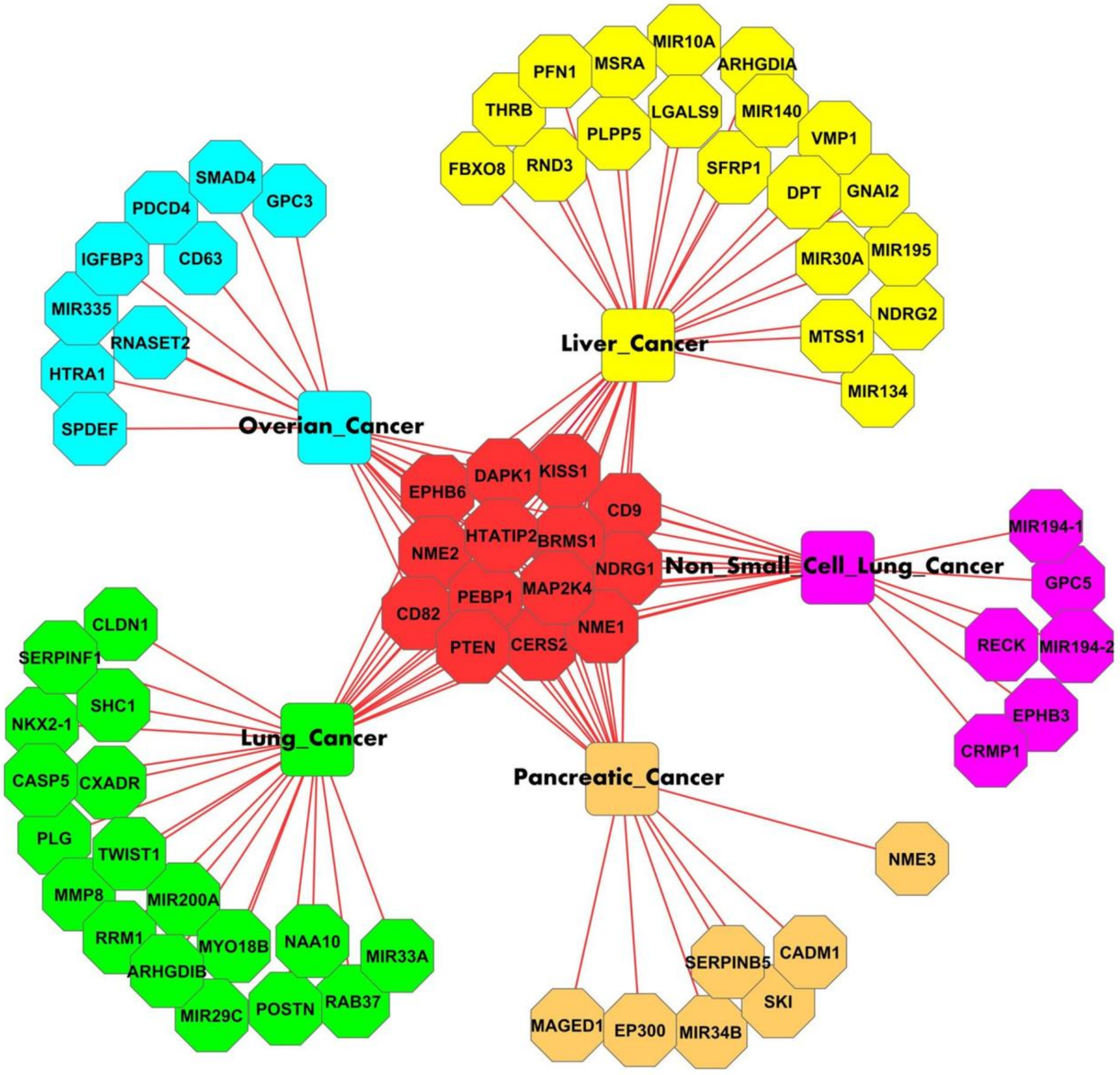


**Figure 3. 3** Module wise analysis of cancer metastasis sub-networks. Square Box is representing Cancer Type and Octagonal Box is highlighting the associated genes to specific type of cancer. Red color Octagonal Box in centre are representing the genes involved metastasis of all cancers

limited genes (~195 in numbers) were experimentally validated to involve in different kinds of cancer as stated in. While mapping these genes with pathway databases like KEGG [90], Reactome [95], and Pathway Interaction Database (PID), no clear understanding of pathway regulation is observed. Further, to uncover missing links in scattered pathway we derived co-expressed genes associated with selected five cancers. Co-expression modules genes were further classified by topological analysis. Degree distribution and betweenness centrality measures were correlated with five different types of cancers as per our previously published co-expression modules in drug-target interaction pipeline [92]. Further, we have converged commonly contributing nodes as central regulatory module and rest of the gene modules are considered as pathway-specific modules as shown in Figure 3.3.

To get overall understanding of genes involvement in various pathways can be done using KEGG database. But, there are number of biological processes which remain hidden due to proper annotation of every dataset. To avoid such kind of discrepancies in our data we used BLAST2GO and GORILLA to understand various anthologies[93]. Prioritization of this genes done on the basis of p-values.

### 3.4.1 The inference from Reconstructed Network: Common genes in various cancers

Through gene pair association score reveals that *CD9* associated with non-small cell lung and (NSCLC), small-cell lung cancer (SCLC), *DAPK1* associated with ovary and lung cancer, *EPHB6* associated NSCLC, *CD82* associated with ovary, pancreas, lung, liver and NSCLC, *KISS1* associated with ovary, pancreas, liver, and NSCLC, NME1 associated with ovary, lung, liver, NSCLC, *NME2* associated with ovary, lung, NSCLC, PEBP1 associated with ovary, pancreas, PTEN associated with lung, NSCLC, MAP2K4 associated with ovary, pancreas, NDRG1 associated with pancreas, lung, liver, HTATIP2 associated with lung, liver, SCLC, BRMS1 associated with ovary, lung, liver, NSCLC, CERS2 associated with pancreas, liver cancer. Combine these genes are contributing to different routes leads to different pathways as shown in Figure 4.3.

### 3.4.2 Lung Cancer Specific genes and pathway association

There are fifteen genes viz. *ARHGDIB, CASP5, CXADR*, *MMP8, SERPINF1*, *PLG, RRM1*, *SHC1*, *NKX2-1, TWIST1, NAA10, CLDN1, POSTN*, *MYO18B, RAB37* involved specifically in lung cancer response. Genes associated with different pathways shown in Figure 4.3.

### 3.4.3 Ovarian Cancer specific genes and pathway association

Eight genes viz. *CD63*, *GPC3*, *IGFBP3, SMAD4, HTRA1*, *RNASET2, SPDEF*, *PDCD4* involved in specifically in ovarian cancer. Genes associated with different pathways shown in Figure 4.3.

### 3.4.4 Pancreatic Cancer-specific genes and pathway association

Six genes viz. *EP300, NME3, SERPINB5*, *SKI*, *MAGED1*, *CADM1* involved in specifically in pancreatic cancer. Genes associated with different pathways shown in Figure 4.3.

### 3.4.5 Liver Cancer specific genes and pathway association

Fourteen genes viz. *RND3, ARHGDIA*, *DPT*, *GNAI2*, *LGALS9*, *MSRA*, *PFN1*, *SFRP1*, *THRB*, *MTSS1*, *FBXO8*, *NDRG2*, *VMP1*, *PLPP5* involved in specifically in liver cancer. Genes associated with different pathways shown in Figure 4.3.

### 3.4.6 Non-small cell lung Cancer specific genes and pathway association

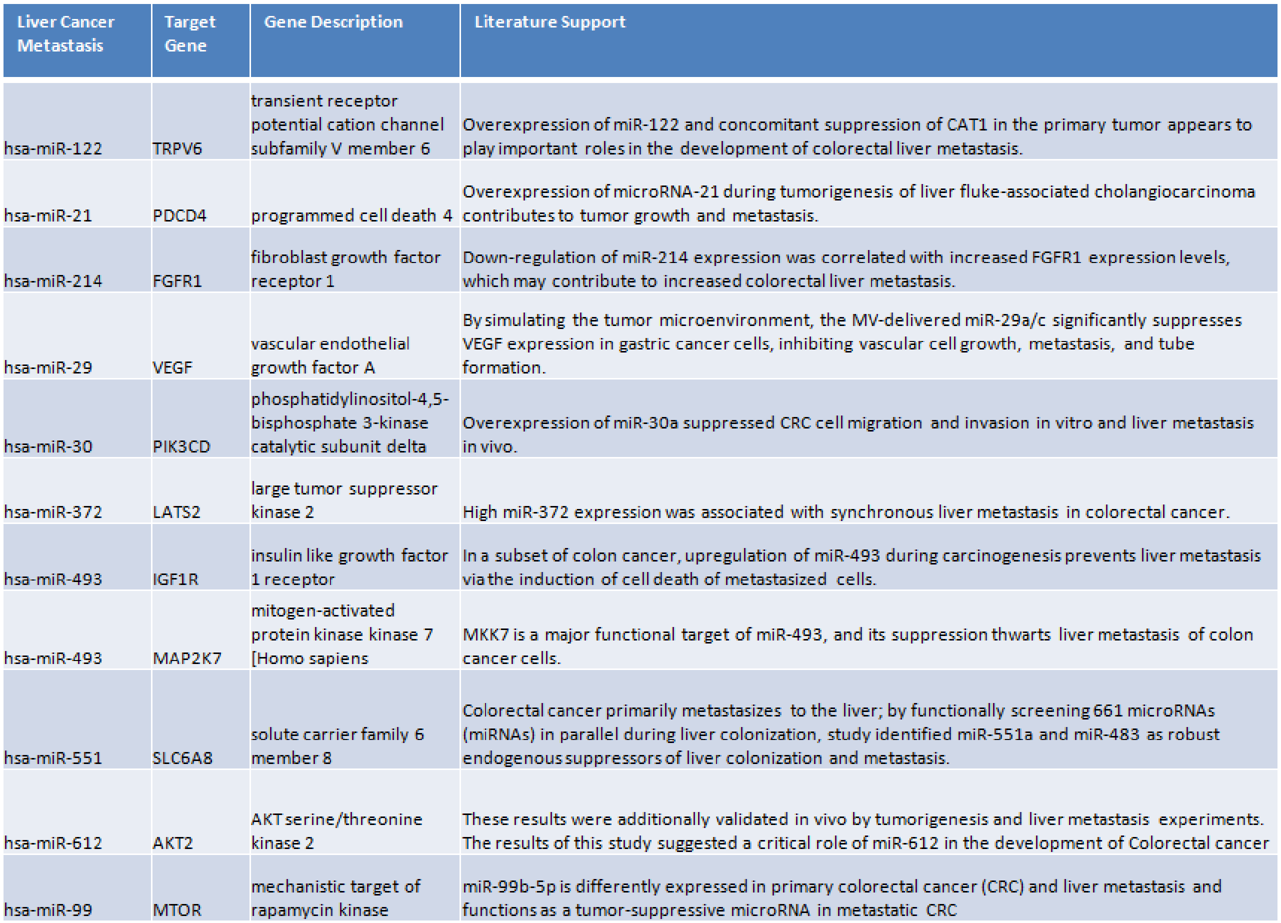
Four genes viz. *CRMP1*, *EPHB3*, *GPC5*, *RECK* involved in specifically in NSCLC cancer. Genes associated with different pathways shown in Figure 4.3.

### 3.4.7 miRNA involved in various cancers

MIR200A, MIR29C, MIR33A especially found in lung cancer, MIR335 in ovarian cancer, MIR34B in pancreatic cancer, MIR10A, MIR134, MIR140, MIR195, MIR30A in liver cancer, MIR194-1, MIR194-2 in NSCLC. Overall miRNA based regulation on different cancer type pathways shown in Figure 4.3.

There was a need to check the robustness of our proposed methods, so we have used another metastasis databases. Results seems consistant over comparison but for more sophisticated analysis patient profiles can be estimated using differential expression data referal on the basis of our method. Subsequently, there is a need of miRNA target identification which may help in possible regulation at transcriptional level of metastasis genes.

**Table 3. 1** Role of miRNAs and its targets in cancer metastasis



| Liver Cancer Metastasis | Target Gene | Gene Description | Literature Support |
|---|---|---|---|
| hsa-miR-122 | TRPV6 | transient receptor potential cation channel subfamily V member 6 | Overexpression of miR-122 and concomitant suppression of CAT1 in the primary tumor appears to play important roles in the development of colorectal liver metastasis. |
| hsa-miR-21 | PDCD4 | programmed cell death 4 | Overexpression of microRNA-21 during tumorigenesis of liver fluke-associated cholangiocarcinoma contributes to tumor growth and metastasis. |
| hsa-miR-214 | FGFR1 | fibroblast growth factor receptor 1 | Down-regulation of miR-214 expression was correlated with increased FGFR1 expression levels, which may contribute to increased colorectal liver metastasis. |
| hsa-miR-29 | VEGF | vascular endothelial growth factor A | By simulating the tumor microenvironment, the MV-delivered miR-29a/c significantly suppresses VEGF expression in gastric cancer cells, inhibiting vascular cell growth, metastasis, and tube formation. |
| hsa-miR-30 | PIK3CD | phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit delta | Overexpression of miR-30a suppressed CRC cell migration and invasion in vitro and liver metastasis in vivo. |
| hsa-miR-372 | LATS2 | large tumor suppressor kinase 2 | High miR-372 expression was associated with synchronous liver metastasis in colorectal cancer. |
| hsa-miR-493 | IGF1R | insulin like growth factor 1 receptor | In a subset of colon cancer, upregulation of miR-493 during carcinogenesis prevents liver metastasis via the induction of cell death of metastasized cells. |
| hsa-miR-493 | MAP2K7 | mitogen-activated protein kinase kinase 7 [Homo sapiens | MKK7 is a major functional target of miR-493, and its suppression thwarts liver metastasis of colon cancer cells. |
| hsa-miR-551 | SLC6A8 | solute carrier family 6 member 8 | Colorectal cancer primarily metastasizes to the liver; by functionally screening 661 microRNAs (miRNAs) in parallel during liver colonization, study identified miR-551a and miR-483 as robust endogenous suppressors of liver colonization and metastasis. |
| hsa-miR-612 | AKT2 | AKT serine/threonine kinase 2 | These results were additionally validated in vivo by tumorigenesis and liver metastasis experiments. The results of this study suggested a critical role of miR-612 in the development of Colorectal cancer |
| hsa-miR-99 | MTOR | mechanistic target of rapamycin kinase | miR-99b-5p is differently expressed in primary colorectal cancer (CRC) and liver metastasis and functions as a tumor-suppressive microRNA in metastatic CRC |

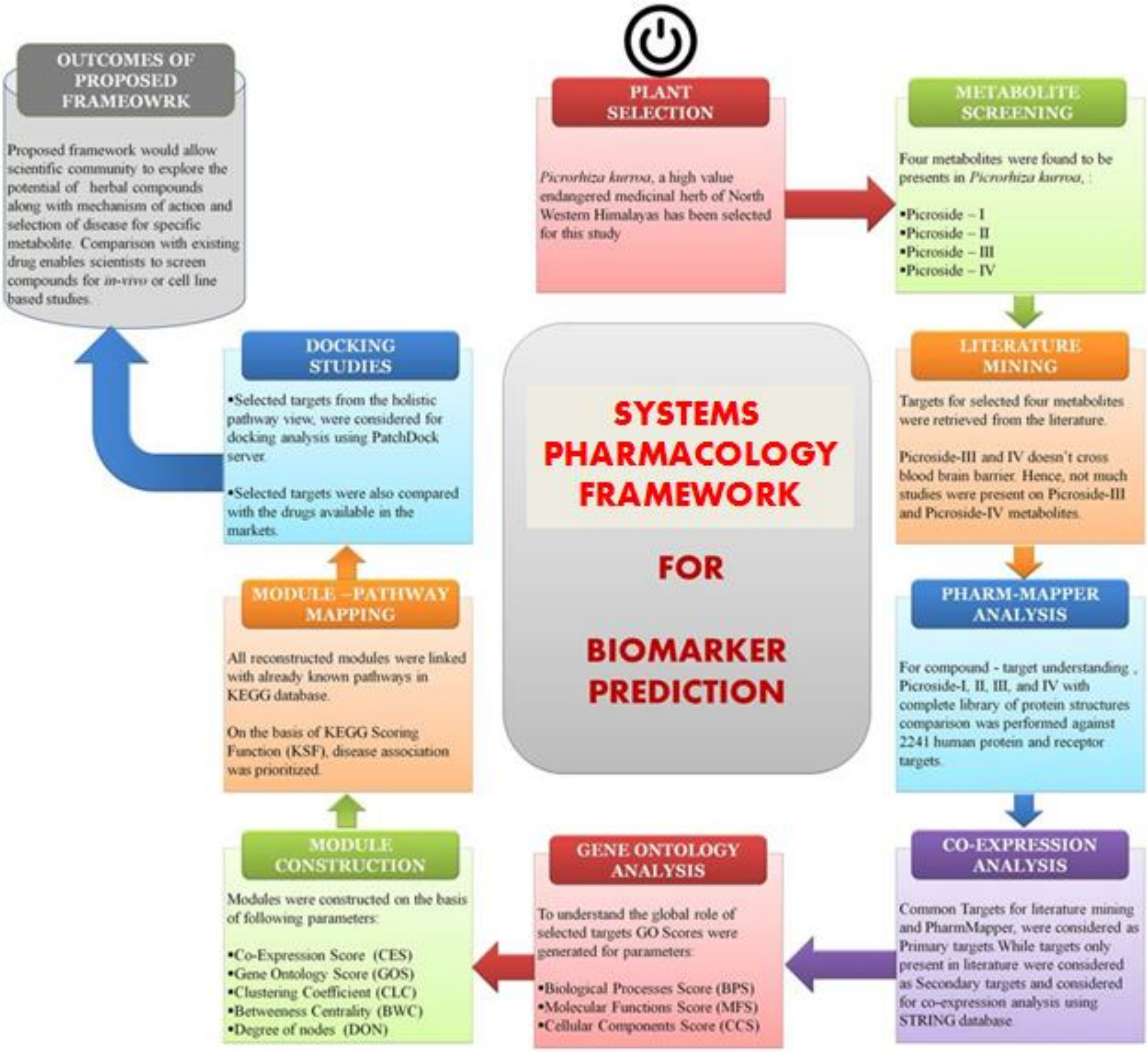
Overall this study paves a path towards understanding complex diseases by constructing a global network on the basis of available information and navigating to pathway of interest on the basis of available information in databases. Plausible targets can be predicted for experimental validation with much sophisticated computational validations. Unique patterns can be noted down to make method more reliable so that homology based inference can be considered and machine learning methods can be implemented for therapies.

## 3.5 Conclusion

Understanding of pathway of interest and control over the network cannot be done through a module level study. Hence, there is a need to combine the available information in the form of network and deduce relationship to infer the pathway of interest. With the help of the presented network reconstruction framework, biologists can identify key regulations in pathway if interest. For instance, cancer metastasis network of five major cancer types was reconstructed, and common and specific network routes were deciphered. Our study provides a miRNA and network based inference for potential target screening. Proposed candidates can be used for further experimental validation.

# CHAPTER 4

## A computational pipeline for drug-target interaction network construction



**SYSTEMS PHARMACOLOGY FRAMEWORK**

**FOR**

**BIOMARKER PREDICTION**

**PLANT SELECTION**

*Picrorhiza kurroa*, a high value endangered medicinal herb of North Western Himalayas has been selected for this study

**METABOLITE SCREENING**

Four metabolites were found to be presents in *Picrorhiza kurroa*, :

- Picroside – I
- Picroside – II
- Picroside – III
- Picroside – IV

**LITERATURE MINING**

Targets for selected four metabolites were retrieved from the literature.

Picroside-III and IV doesn't cross blood brain barrier. Hence, not much studies were present on Picroside-III and Picroside-IV metabolites.

**PHARM-MAPPER ANALYSIS**

For compound - target understanding , Picroside-I, II, III, and IV with complete library of protein structures comparison was performed against 2241 human protein and receptor targets.

**CO-EXPRESSION ANALYSIS**

Common Targets for literature mining and PharmMapper, were considered as Primary targets.While targets only present in literature were considered as Secondary targets and considered for co-expression analysis using STRING database.

**GENE ONTOLOGY ANALYSIS**

To understand the global role of selected targets GO Scores were generated for parameters:

- Biological Processes Score (BPS)
- Molecular Functions Score (MFS)
- Cellular Components Score (CCS)

**MODULE CONSTRUCTION**

Modules were constructed on the basis of following parameters:

- Co-Expression Score (CES)
- Gene Ontology Score (GOS)
- Clustering Coefficient (CLC)
- Betweeness Centrality (BWC)
- Degree of nodes (DON)

**MODULE –PATHWAY MAPPING**

All reconstructed modules were linked with already known pathways in KEGG database.

On the basis of KEGG Scoring Function (KSF), disease association was prioritized.

**DOCKING STUDIES**

- Selected targets from the holistic pathway view, were considered for docking analysis using PatchDock server.
- Selected targets were also compared with the drugs available in the markets.

**OUTCOMES OF PROPOSED FRAMEOWRK**

Proposed framework would allow scientific community to explore the potential of herbal compounds along with mechanism of action and selection of disease for specific metabolite. Comparison with existing drug enables scientists to screen compounds for *in-vivo* or cell line based studies.

## 4.1 Abstract

It is very difficult to understand the universal laws controlling functionality of biological networks. Functionality measures can accurately be determined regarding drug-target interaction network analysis. Network biology implicates the effects of drug targeting as the single gene to single drug hypothesis doesn’t work in today’s scenario. Therefore, there is a need for system biology derived framework which can be used for efficient diagnosis by understanding the global perspective of pathway regulation. Previously defined approaches in this domain are data dependent only, but there is a need for hypothesis validation along with data incorporation. Henceforth, we have used gene ontology, correlation analysis, and docking studies to prioritize candidate biomarkers by module construction by topological properties of the network. The proposed pipeline can be used for inferring the information of any drug target understanding which is still not revealed by other scientists or researchers.

## 4.2 Introduction

Data present in various online databases like OMIM or FDA suggests that there are many targets for drugs yet selectivity of the drug remains a matter of concern. There is a need for genetic interventions to deal with the drug target hypothesis [96]. A lot of information is not explored about protein-protein interactions, so identified targets by various drugs do not follow the absolute path which later results in various side effects by altering subsequent cellular mechanisms [92], [97].

There is a need of a multi-disciplinary approach to study various dimensions of data as done by *Albert-László Barabási Group.* Their group works on social networks, physical networks, and network medicine concept by taking inference of system biology [93], [98]–[101]. It has been found that genes which are known to associate with the disease usually form a cluster or sub-network. Interactome analysis of such sub-networks may reveal the association of targeted gene derived signaling cascade [102]–[104]. There are various online tools, web-servers, and databases which catalogues various drug associated information along with interacting partners and structure level information [105], homogeneous and heterogeneous network construction [106], miRNA and other non coding RNA based drug designing [107], machine learning based drug models [108], [109], forest prediction methods [110], [111], deep learning based methods [112]. Such methods made us think that there is a need to target protein in efficient manners to inhibit the effects of mutations or deregulation at pathway level [113], [114]. Henceforth, for a better understanding of therapies, we have developed unsupervised and network-based drug disease understanding to provide potent molecules for disease blockage.

Picroliv was considered for drug-target interaction study to treat diseases in the human body. Picroliv is a product of *Picrorhiza kurroa* which grows at hight of 3,000-5,000 meters. It is usually is usually a mixture of kutkoside and picroside-I in 1:1.5 ratio [115]. Other products of the plant are picroside-I and picroside-II [42].

The component of picrorhiza has been used from so many decades to treat liver related disorders and chronic pain. It has been used for the treatment of malaria in rats [115], [116]. Picrosides also shows positive results on anti-lipid peroxidative effect[117]. Picroside-II involved in the cure of I/R injury[118], [119]. The exact mechanism of action of picrosides remains unknown. Hence there is need for further exploration of this medicinally important species[120].

To attain the overall understanding of regulatory mechanisms, there is need to identify neighboring genes associated with the targets. Target based regulation can be studied by the complete mechanism of action understanding [121]. Combined networks of associated proteins can suggest a broader way to deal with the disease by blocking all possible routes [103]. By these parameters, specific disease-associated targets can be filtered using enrichment analysis of patient profiles.

By information available in the literature, picrorhiza based drug target network has been reconstructed to understand the exact mechanism of action. Construction of biological networks of medicinal herb doesn't only require with respect to network biology by also give insight about phenotypic changes in healthy and diseased conditions. Apart from direct target inhibition, there is various signaling cascade which gets affected by the inhibition of upstream. Henceforth, need of a holistic view of network construction and narrow down to the target of interaction is much needed for drug targeting.

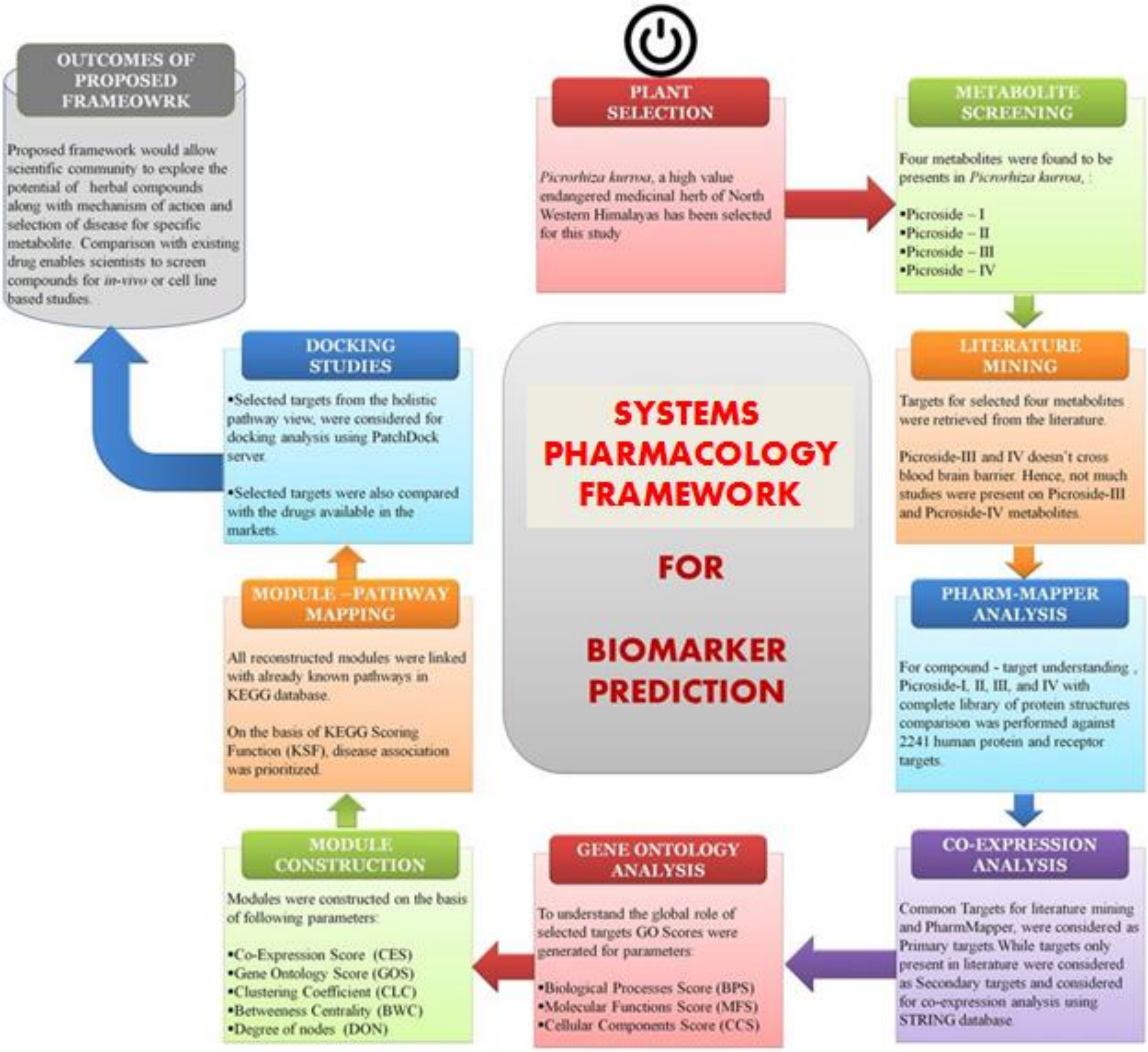
By combining the concepts of interdisciplinary fields may result in an optimal solution to understand drug-target regulation. Therefore, we have integrated the literature support based experimental data along with computational structure level predictions under systems pharmacology framework. Adverse effects of the drugs can be subsequently observed by inhibition of multi-modules in the drug targeted therapies. In the proposed framework, we are presenting pipeline incorporation with gene ontology, co-expression, topological parameters, pathway mapping and docking studies for understanding and discerning the properties of drug target interaction network.

## 4.3 Methods

Medicinal compounds can be used for understanding the complex therapeutic mechanisms by performing the various experimental analysis. There is no doubt that many compounds of medicinal herbs which are being used for symptomatic effects instead of identifying the exact mechanism of action. This understanding has led us to think that there is a need of a systematic way of precise identification of therapeutic targets. We have performed literature based screening of metabolic for target understanding and predicting novel drug targets for picroside specific targeting. Figure 4.1 shows a systematic representation of our framework.

### 4.3.1 Literature Mining

A number of articles is being published in a variety of journals at a very fast rate which makes it difficult for researchers to catalog all related information at one repository. For our study, we have used many keywords associated with associated with picroside and associated therapeutic actions to screen out the candidate drug targets. For this purpose, we have used FACTA+, PUBMED, and other search engines.



OUTCOMES OF PROPOSED FRAMEOWRK
Proposed framework would allow scientific community to explore the potential of herbal compounds along with mechanism of action and selection of disease for specific metabolite. Comparison with existing drug enables scientists to screen compounds for *in-vivo* or cell line based studies.
PLANT SELECTION
*Picrorhiza kurroa*, a high value endangered medicinal herb of North Western Himalayas has been selected for this study
METABOLITE SCREENING
Four metabolites were found to be presents in *Picrorhiza kurroa*, :
•Picroside – I
•Picroside – II
•Picroside – III
•Picroside – IV
LITERATURE MINING
Targets for selected four metabolites were retrieved from the literature.
Picroside-III and IV doesn't cross blood brain barrier. Hence, not much studies were present on Picroside-III and Picroside-IV metabolites.
PHARM-MAPPER ANALYSIS
For compound - target understanding , Picroside-I, II, III, and IV with complete library of protein structures comparison was performed against 2241 human protein and receptor targets.
CO-EXPRESSION ANALYSIS
Common Targets for literature mining and PharmMapper, were considered as Primary targets. While targets only present in literature were considered as Secondary targets and considered for co-expression analysis using STRING database.
GENE ONTOLOGY ANALYSIS
To understand the global role of selected targets GO Scores were generated for parameters:
•Biological Processes Score (BPS)
•Molecular Functions Score (MFS)
•Cellular Components Score (CCS)
MODULE CONSTRUCTION
Modules were constructed on the basis of following parameters:
•Co-Expression Score (CES)
•Gene Ontology Score (GOS)
•Clustering Coefficient (CLC)
•Betweeness Centrality (BWC)
•Degree of nodes (DON)
MODULE –PATHWAY MAPPING
All reconstructed modules were linked with already known pathways in KEGG database.
On the basis of KEGG Scoring Function (KSF), disease association was prioritized.
DOCKING STUDIES
•Selected targets from the holistic pathway view, were considered for docking analysis using PatchDock server.
•Selected targets were also compared with the drugs available in the markets.
SYSTEMS PHARMACOLOGY FRAMEWORK
FOR
BIOMARKER PREDICTION

**Figure 4. 1** Systems level analysis framework for drug target screening

### 4.3.2 Target Prediction

We were interested to screen out the targets which are supposed to play a role in humans. These metabolites can be compared by performing a drug-target interaction studies using complementary shape principles. For this study, 2D structures were downloaded for these picroside derivatives from the PubChem database. Downloaded structures were supplied as input for in PharmMapper server for therapeutic target prediction. This server has around 2241 human proteins as target dataset for comparison with supplied 2D structures. We have limited targets up to 300 for our study by atomic contact energy (ACE) and shape complementary score.

### 4.3.3 Common Target Identification

We have performed a comparative analysis by comparing literature based targets and PharmMapper targets to understand the regulatory mechanisms. Common targets in both the analysis were considered as the primary targets as these targets have direct interaction with selected compounds and already validated by researchers. While there were some targets which were present in literature but PharmMapper has not resulted same in their 300 target list. So, we have termed these targets as secondary targets. By primary and secondary targets, we have narrow down our analysis to further co-expression and gene ontology studies.

### 4.3.4 Gene Ontology and Co-Expression Network

To understand the collaborative effect of primary and secondary targets we have focused on gene ontology analysis. Gene ontology describes its vocabulary in the form of biological processes, molecular functions, and cellular components. Biological processes give inference of various regulation at the molecular level; molecular function describes the role of the genes in various pathways and cellular component leads to a variation of localization of gene at various compartment in a cellular system. These three parameters were combining used to define a node score to prioritize the correlation between a set of genes. Same type of genes clustered together while other was differentiated from the given clusters. We were interested in elaborating our network using correlation analysis that is why we have used parametric and non parametric test which we have already standardized in our previous pipelines. Pearson and spearmen correlation coefficients were used to prioritize candidate genes. Moreover, we were interested to construct co-expression networks. For this purpose, we have used a combination of gene ontology and correlation scoring functions. By selected candidates, we have STRING inferred co-expression network which was studied under the network biology parameter estimation. Centrality, shortest path and degree distribution based inference were major criteria for selection of nodes into hierarchal co-expression network construction.

### 4.3.5 Pathway Mapping of Co-Expressed Modules

Once we have constructed the co-expression modules, there was a need to investigate the involvement of these genes in pathways. For identifying gene association with various pathways, we took reference of standard pathway databases. KEGG, Reactome, Pathway Interaction Database and other databases. Apart from these standard databases, we have taken inference literature to understand the regulatory mechanism of thesis targets. The constructed network is static, and hence we did not have any dynamic behavior to show the absolute regulation. But, pathway differentiation values regarding transcript abundance can be considered for further analysis[90], [95], [122].

### 4.3.6 Patch Dock Analysis

After pathway mapping which was derived from co-expression module construction, we have performed docking studies to understand absolute interaction between selected picroside derivatives and targets. We were interested in understanding global machinery which was regulating the various mechanisms [123]. After having a clear understanding of pathway regulation and differentiation points along with merging key nodes, we have selected few biomarkers for our study and proposed that as plausible targets for experimental validation.

## 4.4 Results and Discussion

To understand the pathway regulation, we have used traditional way to screen out data from the literature for P-I, II, III and IV. We have considered P-I and P-II for further computational studies as we had enough literature to validate our analysis. P-III and P-IV were not further considered as we did not have enough literature support to cross-validate our targets. Major reason for dropping this molecule in our study is that it doesn’t cross blood-brain barrier[124]. Structure-based comparison of PharmMapper was limited 30 targets while checking against receptor or protein database[125]. Combined primary targets were considered as inhibitory molecules while secondary were referred as downstream inhibitory proteins for confirmation of complete blockage of disease-oriented pathways.

Moreover, to understand the role of secondary targets and their validation remains crucial, so we have selected top co-expressed partners by primary target based confidence score. Node selection was prioritized by betweenness centrality which will state the genes which are centrally having a controlling effect on associated genes. Further, we have degree distribution of the selected targets. The inference has been taken by $k$ and $B_c$ graphs. In figure 4.2, bigger the node size states more important of the genes compared to others.

Genes highlighted through co-expression analysis were collated together with targets suggested biological processes, molecular functions and cellular components under gene ontology analysis using BLAST2GO and GORILLA servers [126]. From *p*-value, major target was considered to have strong association while others one considered to having a weak association. To move forward towards validation of our results, we have used PatchDock servers to select the best conformation of primary targets. Results of PatchDock are shown in (Table 4.1). Moreover, we have compared our drug targets with FDA approved drugs to understand the best regulation of P-I, P-II compared to other drugs[123]. Candidates selected from our analysis were considered as main contributing elements in carcinogenesis and hence their understanding treated as crucial for drug target understanding. Primary targets were further referred by module construction by co-expression score. Different colors have been used along with different size by topological parameters as shown in Figure 4.2. Further signaling and their regulation is shown through pathway reconstruction (Figure 4.3).

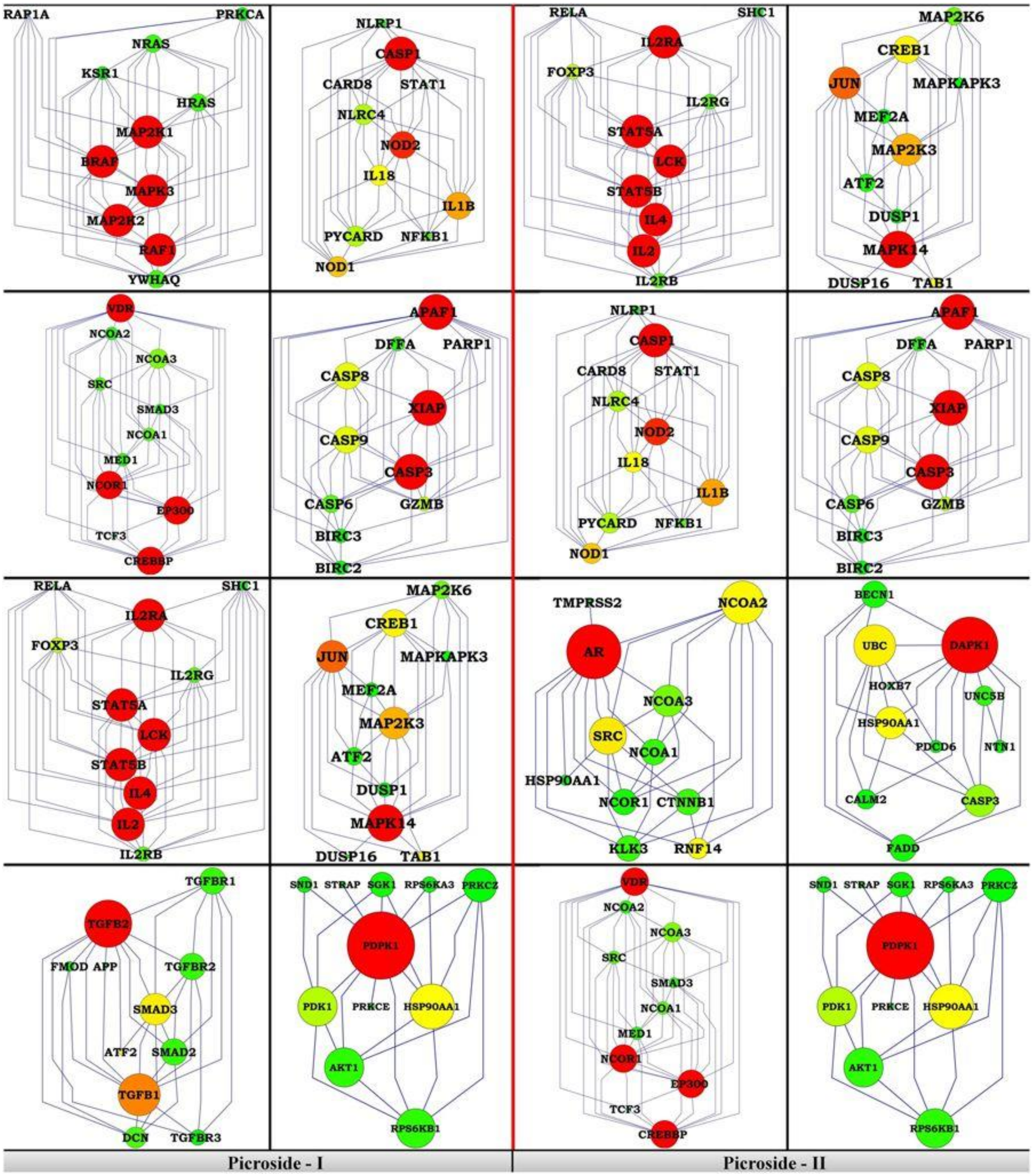


**Figure 4. 2** Co-expression network module construction

### 4.4.1 Pathway Analysis

A number of pathways were found to be involved in diseases by the genes highlighted through our analysis. Some of the pathways from various diseases remained consistant while studying global network. But, KEGG mapping have revealed the selected genes based pathway induction at given below with literature support. Such signalings are a potential regulators in carcinogenesis (Figure 4.3). Role of signaling specifically in hepatocellular carcinoma is indicated below as Picrosides are well known for liver-related diseases.

- Receptor-based Death-Associated Protein Kinase 1 (DAPK1) activation

**Literature Support:** *Li, Ling, et al. "DAPK1 as an independent prognostic marker in liver cancer." PeerJ 5 (2017): e3568.*

- Transforming Growth Factor Beta (TGFβ) signaling

**Literature Support:** *Moon, Hyuk, et al. "Transforming growth factor-β promotes liver tumorigenesis in mice via up-regulation of a snail." Gastroenterology 153.5 (2017): 1378-1391.*

- Interleukin (IL) 2, four signaling

**Literature Support:** *Wu, Zhitong, et al. "Association Between IL-4 Polymorphisms and Risk of Liver Disease: An Updated Meta-Analysis." Medicine 94.35 (2015).*

*Gabeen, Abdulwahab Ali, et al. "Potential immunotherapeutic role of interleukin-2 and interleukin-12 combination in patients with hepatocellular carcinoma." Journal of hepatocellular carcinoma 1 (2014): 55.*

- Cytokine signaling

**Literature Support:** *Molina, Manuel Flores, et al. "Type 3 cytokines in liver fibrosis and liver cancer." Cytokine (2018).*

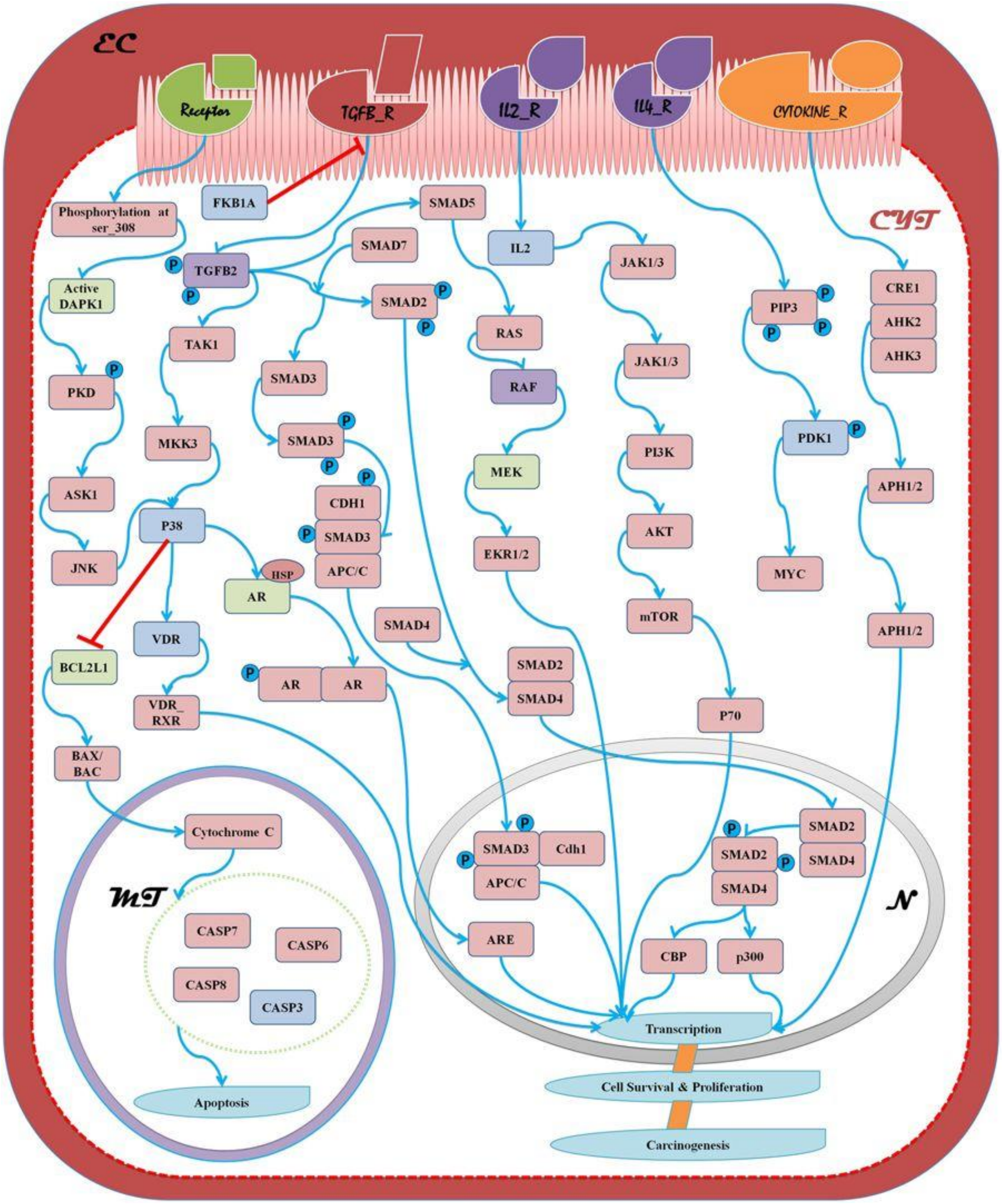


**Figure 4. 3** Global pathway construction on the basis of receptor signaling

#### 4.4.1.1 Death-Associated Protein Kinase-mediated signaling

Calcium/calmodulin-dependent serine/threonine kinases (CDK) regulating various cell death related pathways by using JNK route [127]. Autophagy pathway

#### 4.4.1.2 Transforming Growth Factor Beta mediated signaling

TGFβ is major signaling in various cancers as it regulates both cell death as well as cell differentiation pathway through various signaling cascades. Major activation routes utilize SMAD complex for transcription activation through various interacting nodes which includes FKB1A, CASP1, CASP3, and TGFB2. But inhibiting TGFβ, all downstream signaling can be inhibiting which can be referred from pathway reconstruction image as well as Table 4.1.

#### 4.4.1.3 Interleukin mediated signaling

IL2 and IL4 signaling help in the differentiation of T helper 2 (TH2) cells and immunoglobulin E (IgE) [128]. Usually, mitogen-activated protein kinase (MAPK), phosphoinositide 3-kinase (PI3K), signal transducers and activators of transcription (STAT), and mammalian target of rapamycin (mTOR) signally plays a crucial role in cancer-associated pathway regulation through interleukin routes [129] (Table 4.1).

#### 4.4.1.4 Cytokine-mediated signaling

The release of cytokine has a crucial role in cell differentiation and apoptotic path regulation as it is involved in various immunity related responses in a variety of cells [130]. The JAK-STAT targeting can inhibit complete role and can help in the progression of cancer in various cell lines [130]. MAPKs considered as a central point for controlling cell differentiation pathways at transcriptional activation.

We have reconstructed the combined pathways of above-stated signaling and through PatchDock stated the structural target inhibition. But, short listing of biomarkers needs to be done by cancer patient profiles and combinatorial therapies. BRAF, FKB1A, CASP1, CASP3, TGFB2, IL2 and MAPKs considered as key targets which can be inhibited by Picrosides. These targets can further be used for experimental validation.

The presented study suggests a novel path to identify the key targets for any medicinal herbs. Also, target screening can be done for any diseases by our network biology inferred analysis

**Table 4. 1 Blind docking for Picroside I and II against shortlisted targets from human protein library**

| PDB ID | UNIPLOT | SCORE | AREA | ACE |
|---|---|---|---|---|
| **1BL4** | FKB1A | 1328 | 451.7 | -346.89 |
| **1DB1** | VDR | 2190 | 396.9 | -402.49 |
| **1ICE** | CASP1 | 1290 | 490.5 | -456.81 |
| **1NMS** | CASP3 | 1624 | 387.5 | -378.14 |
| **1PW6** | IL2 | 1720 | 416.1 | -439.09 |
| **1NXK** | P49137 | 1486 | 386.8 | -484.19 |
| **1TFG** | TGFB2 | 3130 | 346.8 | -418.27 |
| **1MQ6** | FA10 | 1300 | 375.8 | -459.61 |
| **2PE1** | PDPK1 | 1896 | 461.9 | -562.56 |
| **2RGS** | Q16539 | 1568 | 455.9 | -614.34 |
| **3C4C** | BRAF1 | 1472 | 427.3 | -488.98 |
| **3FV8** | P53779 | 1366 | 450.6 | -429.62 |

B)

| PDB ID | UNIPLOT | SCORE | AREA | ACE |
|---|---|---:|---:|---:|
| **1BL4** | FKB1A | 1950 | 319.5 | -341.15 |
| **1BMQ** | CASP1 | 2334 | 468.1 | -494.43 |
| **1DB1** | VDR | 3352 | 454.5 | -392.19 |
| **1GS4** | ANDR | 1850 | 363.7 | -386.21 |
| **1IG1** | DAPK1 | 1226 | 427.5 | -459.77 |
| **1KV2** | Q16539 | 1502 | 451.9 | -526.44 |
| **1PMN** | MK10 | 2156 | 474.4 | -480.14 |
| **1PY2** | IL2 | 1478 | 386.5 | -431.36 |
| **1RHJ** | CASP3 | 2770 | 391.2 | -326.53 |
| **1S9J** | MP2K1 | 1272 | 378.3 | -462.76 |
| **1NXK** | P49137 | 1958 | 476.7 | -456.24 |
| **2JRI** | FA10 | 2920 | 448.1 | -368.14 |

|  |  |  |  |  |
|---|---|---|---|---|
| **2PEI** | PDPK1 | 1660 | 445.9 | -547 |
| **2YXJ** | Q07817 | 1612 | 421.3 | -462.2 |

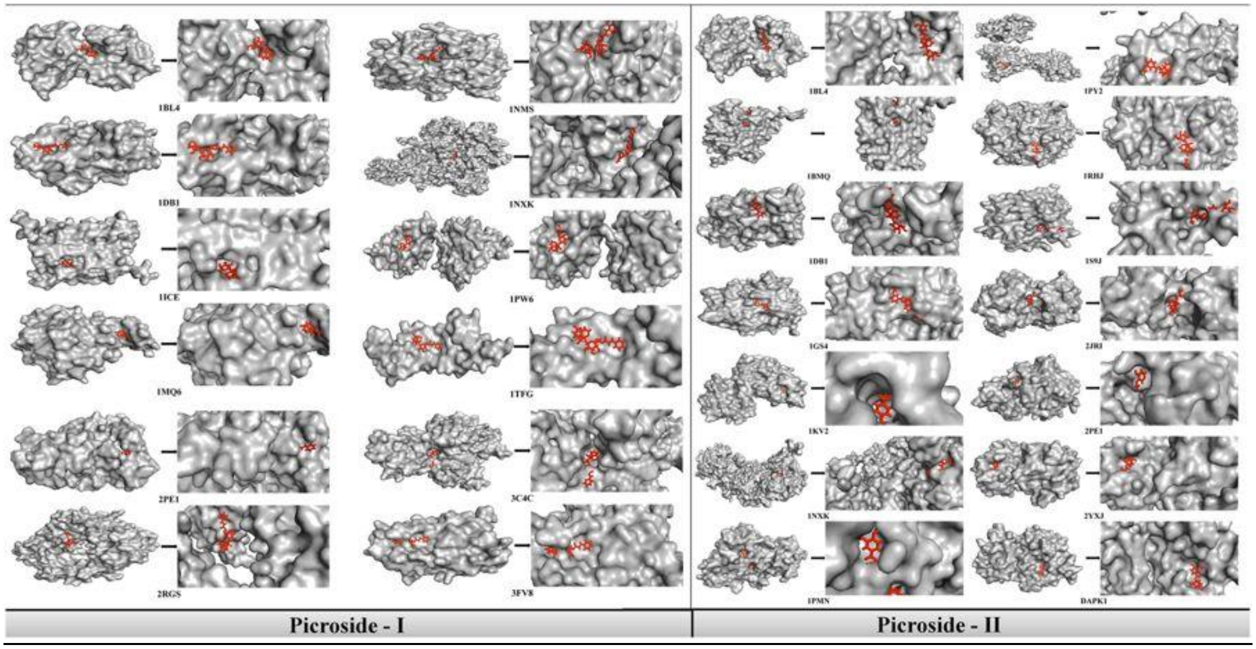


**Figure 4. 4** Docking studies of Picroside I and II.

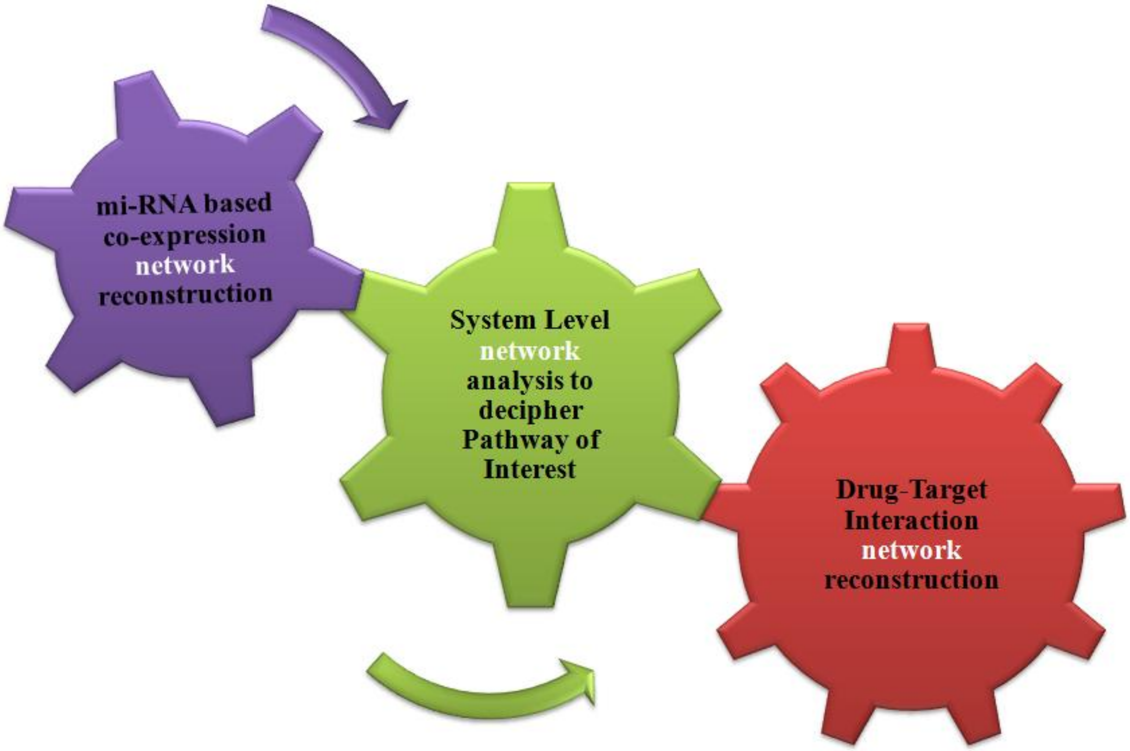
framework. Network-based drug-target studies can help computational and experimental biologist to screen candidate drug-target interactions.

## 4.5 Conclusion

There is a need to collaborative study effects of the genes instead of focusing on one gene. As the collaborative effect of the genes state regulation of the pathways with control at triggering points. With the help of our pipeline, researchers can construct a global network for their datasets and especially target on the pathway of interest using network biology approach. Our framework can be used for any medicinally important herbs or metabolite specific targeting in the human body. Integration of the data refining and literature based support emphasized on the robustness of our method. Molecular interaction using docking based approach also revealed the same molecules which were highlighted by the network-based approach. Selected targets from our analysis can be considered as potential biomarkers for experimental validation.

# CHAPTER 5

## CONCLUSION & FUTURE PROSPECTS


mi-RNA based co-expression network reconstruction
System Level network analysis to decipher Pathway of Interest
Drug-Target Interaction network reconstruction

The outcome of the research work in this thesis has helped in the understanding of various cellular and molecular mechanisms through biological networks. A biological network is a set of metabolic and physical processes that determine the physiological and biochemical properties of the cell. Biological networks have tremendous importance in medicine especially in Inborn errors of metabolism cause acute symptoms, metabolic diseases (obesity, diabetes), and metabolic enzymes are becoming viable drug targets. Along with these, biological networks plays a major role in bioengineering applications viz. designing strains for production of biological products, generation of bio-fuels, etc.

Huge amount of data is getting generated in the field of biotechnology through various omics approaches. There is a need to align a complete dataset in the form of a network to deduce significant results at the phenotypic level. There have been great advances in the techniques and methods in the area of computational biology, but no method or facility exists to analyze different parameters at one platform. Several of the existing approaches in this direction, however, are data-driven and thus lack potential to be generalized and extrapolated to different species. Thus new algorithms or pipelines built on hypotheses and data rather than data-only are necessary for integration of omics data and discovery of biologically meaningful patterns. So, to provide help for experimental biologists, we have introduced three different pipelines which are helpful for network reconstruction by differential and co-expression networks.

**Objective 1:** Use of miRNA identification and target prediction techniques gives input to construct biological networks. Moreover, the function level annotation of selected targets can be inferred under network biology framework. Correlation between genotype and phenotype can be established by using various enrichment analysis. With the availability of large-scale omics data, identification of candidate genes for future interventions becomes comparatively easy. Predicted targets through our miRNA analysis framework can further experimentally be validated. This framework can aid in the comparative transcriptomic analysis to understand key regulatory mechanisms and identifying regulatory nodes.

**Objective 2:** The second objective of our network study is to develop a framework to identify the pathway of interest from large datasets. No doubt there are various statistical and machine learning tools available to prioritize candidate biomarkers by specific parameters, but there is no such method or utility available which construct the global network and then narrow down to the pathway of interest. The proposed framework will help in global network construction, identifying randomly occurred nodes, sub-network construction and finally give insight about the roadmap for the pathway of interest. So, our framework can be used for the construction of module-based network which reveals various pathways which are dominant in specific signaling of interest and miRNA framework implications are also utilized in this pipeline. Further, we were interested in studying drug-target interaction to control the disease by comparing and targeting pathway module entities.

**Objective 3:** The third pipeline will help in the specifically targeting pathway of interest through drug-target interaction. There is a need to collaborative study effects of the genes instead of focusing on one gene. As the collaborative effect of the genes state regulation of the pathways with control at triggering points. With the help of our pipeline, researchers can construct a global network for their datasets and especially target on the pathway of interest using network biology approach. Our framework can be used for any medicinally important herbs or metabolite specific targeting in the human body. Integration of the data refining and literature based support emphasized on the robustness of our method. Molecular interaction using docking based approach also revealed the same molecules which were highlighted by the network-based approach. Selected targets from our analysis can be considered as potential biomarkers for experimental validation.

Constructed three pipelines miRNA-Pathway, miRNA-Pathway of Interest and Pathway of interest to drug target interaction network represents a systematic way to deal with high throughput data. Network biology based methods represents a interdisciplinary way to deal with high throughput data and identify various information which was hidden in many data layers. Developed methods have been employed to carry on various analysis of plants, complex diseases in human-like cancer, neurodegenerative disorders, etc. Overall, global picture of the biological network can be constructed, and regulation using regulatory elements or drugs at a network level can be performed through proposed pipelines.

# REFERENCES

## References

[1] J. Travers and S. Milgram, “An Experimental Study of the Small World Problem,” *Sociometry*, vol. 32, no. 4, pp. 425–443, 1969.

[2] D. J. Watts and S. H. Strogatz, “Collective dynamics of ‘small-world’ networks,” *Nature*, vol. 393, no. 6684, pp. 440–442, Jun. 1998.

[3] B. Bollobás, A. Saito, and N. C. Wormald, “Regular factors of regular graphs,” *J. Graph Theory*, vol. 9, no. 1, pp. 97–103, Mar. 1985.

[4] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A. L. Barabási, “The large-scale organization of metabolic networks,” *Nature*, vol. 407, no. 6804, pp. 651–654, Oct. 2000.

[5] H. Jeong, S. P. Mason, A.-L. Barabási, and Z. N. Oltvai, “Lethality and centrality in protein networks,” *Nature*, vol. 411, no. 6833, pp. 41–42, May 2001.

[6] A.-L. Barabási and R. Albert, “Emergence of Scaling in Random Networks,” *Science*, vol. 286, no. 5439, pp. 509–512, Oct. 1999.

[7] R. Albert, H. Jeong, and A.-L. Barabási, “Error and attack tolerance of complex networks,” *Nature*, vol. 406, no. 6794, pp. 378–382, Jul. 2000.

[8] R. Pastor-Satorras and A. Vespignani, “Epidemic dynamics and endemic states in complex networks,” *Phys. Rev. E*, vol. 63, no. 6, May 2001.

[9] F. Liljeros, C. R. Edling, L. A. Amaral, H. E. Stanley, and Y. Aberg, “The web of human sexual contacts,” *Nature*, vol. 411, no. 6840, pp. 907–908, Jun. 2001.

[10] L. H. Hartwell, J. J. Hopfield, S. Leibler, and A. W. Murray, “From molecular to modular cell biology,” *Nature*, vol. 402, no. 6761 Suppl, pp. C47-52, Dec. 1999.

[11] E. V. Koonin, Y. I. Wolf, and G. P. Karev, “The structure of the protein universe and genome evolution,” *Nature*, vol. 420, no. 6912, pp. 218–223, Nov. 2002.

[12] G. R. Bock and J. A. Goode, *“In Silico” Simulation of Biological Processes*. John Wiley & Sons, 2003.

[13] S. S. Shen-Orr, R. Milo, S. Mangan, and U. Alon, “Network motifs in the transcriptional regulation network of Escherichia coli,” *Nat. Genet.*, vol. 31, no. 1, pp. 64–68, May 2002.

[14] R. E. al et, “Hierarchical organization of modularity in metabolic networks. - PubMed - NCBI.” [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/12202830. [Accessed: 09-Sep-2018].

[15] R. Pastor-Satorras, M. Rubi, and A. Diaz-Guilera, *Statistical Mechanics of Complex Networks*. Springer Science & Business Media, 2003.

[16] A. Wagner and D. A. Fell, “The small world inside large metabolic networks,” *Proc. Biol. Sci.*, vol. 268, no. 1478, pp. 1803–1810, Sep. 2001.

[17] A. Barve, J. F. M. Rodrigues, and A. Wagner, “Superessential reactions in metabolic networks,” *Proc. Natl. Acad. Sci.*, vol. 109, no. 18, pp. E1121–E1130, May 2012.

[18] W. Y. al et, “Scale-free networks in biology: new insights into the fundamentals of evolution? - PubMed - NCBI.” [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/11835273. [Accessed: 09-Sep-2018].

[19] G. P. Karev, Y. I. Wolf, A. Y. Rzhetsky, F. S. Berezovskaya, and E. V. Koonin, “Birth and death of protein domains: a simple model of evolution explains power law behavior,” *BMC Evol. Biol.*, vol. 2, p. 18, Oct. 2002.

[20] S. Wuchty, “Scale-free behavior in protein domain networks,” *Mol. Biol. Evol.*, vol. 18, no. 9, pp. 1694–1702, Sep. 2001.

[21] G. Apic, J. Gough, and S. A. Teichmann, “An insight into domain combinations,” *Bioinforma. Oxf. Engl.*, vol. 17 Suppl 1, pp. S83-89, 2001.

[22] G. Apic, J. Gough, and S. A. Teichmann, “Domain combinations in archaeal, eubacterial and eukaryotic proteomes,” *J. Mol. Biol.*, vol. 310, no. 2, pp. 311–325, Jul. 2001.

[23] S. Wuchty, “Interaction and domain networks of yeast,” *Proteomics*, vol. 2, no. 12, pp. 1715–1723, Dec. 2002.

[24] P. J. al et, “Mapping protein family interactions: intramolecular and intermolecular protein family interaction repertoires in the PDB and yeast. - PubMed - NCBI.” [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/11273711. [Accessed: 09-Sep-2018].

[25] I. Xenarios, L. Salwínski, X. J. Duan, P. Higney, S.-M. Kim, and D. Eisenberg, “DIP, the Database of Interacting Proteins: a research tool for studying cellular networks of protein interactions,” *Nucleic Acids Res.*, vol. 30, no. 1, pp. 303–305, Jan. 2002.

[26] H. W. Mewes *et al.*, “MIPS: a database for genomes and protein sequences,” *Nucleic Acids Res.*, vol. 30, no. 1, pp. 31–34, Jan. 2002.

[27] P. Uetz *et al.*, “A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae,” *Nature*, vol. 403, no. 6770, pp. 623–627, Feb. 2000.

[28] T. Ito, T. Chiba, R. Ozawa, M. Yoshida, M. Hattori, and Y. Sakaki, “A comprehensive two-hybrid analysis to explore the yeast protein interactome,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 98, no. 8, pp. 4569–4574, Apr. 2001.

[29] Z. N. Oltvai and A.-L. Barabási, “Systems biology. Life’s complexity pyramid,” *Science*, vol. 298, no. 5594, pp. 763–764, Oct. 2002.

[30] P. Holme, M. Huss, and H. Jeong, “Subnetwork hierarchies of biochemical pathways,” *Bioinforma. Oxf. Engl.*, vol. 19, no. 4, pp. 532–538, Mar. 2003.

[31] M. Girvan and M. E. J. Newman, “Community structure in social and biological networks,” *Proc. Natl. Acad. Sci.*, vol. 99, no. 12, pp. 7821–7826, Jun. 2002.

[32] F. Crick, “Central Dogma of Molecular Biology,” *Nature*, vol. 227, no. 5258, pp. 561–563, Aug. 1970.

[33] A. Jewett *et al.*, “Strategies to Rescue Mesenchymal Stem Cells (MSCs) and Dental Pulp Stem Cells (DPSCs) from NK Cell Mediated Cytotoxicity,” *PLoS ONE*, vol. 5, no. 3, Mar. 2010.

[34] J. M. Peregrín-Alvarez, X. Xiong, C. Su, and J. Parkinson, “The Modular Organization of Protein Interactions in Escherichia coli,” *PLoS Comput. Biol.*, vol. 5, no. 10, p. e1000523, Oct. 2009.

[35] S. H. Strogatz, “Exploring complex networks,” *Nature*, 08-Mar-2001. [Online]. Available: https://www.nature.com/articles/35065725. [Accessed: 09-Sep-2018].

[36] A. P. Burgard, E. V. Nikolaev, C. H. Schilling, and C. D. Maranas, “Flux coupling analysis of genome-scale metabolic network reconstructions,” *Genome Res.*, vol. 14, no. 2, pp. 301–312, Feb. 2004.

[37] S. Klamt and J. Stelling, “Combinatorial complexity of pathway analysis in metabolic networks,” *Mol. Biol. Rep.*, vol. 29, no. 1–2, pp. 233–236, 2002.

[38] E. Dekel and U. Alon, “Optimality and evolutionary tuning of the expression level of a protein,” *Nature*, vol. 436, no. 7050, pp. 588–592, Jul. 2005.

[39] R. U. Ibarra, J. S. Edwards, and B. O. Palsson, “Escherichia coli K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth,” *Nature*, vol. 420, no. 6912, pp. 186–189, Nov. 2002.

[40] J. Nielsen and S. Oliver, “The next wave in metabolome analysis,” *Trends Biotechnol.*, vol. 23, no. 11, pp. 544–546, Nov. 2005.

[41] A.-L. Barabási and Z. N. Oltvai, “Network biology: understanding the cell’s functional organization,” *Nat. Rev. Genet.*, vol. 5, no. 2, pp. 101–113, Feb. 2004.

[42] V. Kumar, A. Bansal, and R. S. Chauhan, “Modular Design of Picroside-II Biosynthesis Deciphered through NGS Transcriptomes and Metabolic Intermediates Analysis in Naturally Variant Chemotypes of a Medicinal Herb, Picrorhiza kurroa,” *Front. Plant Sci.*, vol. 8, 2017.

[43] A. Bansal and P. A. Srivastava, “Transcriptomics to Metabolomics: A Network Perspective for Big Data,” *IGI Glob.*, pp. 188–206, 2018.

[44] K. Jindal and A. Bansal, “APOEε2 is Associated with Milder Clinical and Pathological Alzheimer’s Disease,” *Ann. Neurosci.*, vol. 23, no. 2, pp. 112–112, 2016.

[45] A. Bansal and J. Ramana, “TCGDB: A Compendium of Molecular Signatures of Thyroid Cancer and Disorders,” *J. Cancer Sci. Ther.*, vol. 7, no. 7, Jul. 2015.

[46] A. E. Pasquinelli, “MicroRNAs and their targets: recognition, regulation and an emerging reciprocal relationship,” *Nat. Rev. Genet.*, vol. 13, no. 4, pp. 271–282, Apr. 2012.

[47] S. Liu *et al.*, “The Host Shapes the Gut Microbiota via Fecal MicroRNA,” *Cell Host Microbe*, vol. 19, no. 1, pp. 32–43, Jan. 2016.

[48] A. H. Buck *et al.*, “Exosomes secreted by nematode parasites transfer small RNAs to mammalian cells and modulate innate immunity,” *Nat. Commun.*, vol. 5, p. 5488, Nov. 2014.

[49] A. B. Rodgers, C. P. Morgan, N. A. Leu, and T. L. Bale, “Transgenerational epigenetic programming via sperm microRNA recapitulates effects of paternal stress,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 112, no. 44, pp. 13699–13704, Nov. 2015.

[50] N. D. Mendes, A. T. Freitas, and M.-F. Sagot, “Current tools for the identification of miRNA genes and their targets,” *Nucleic Acids Res.*, vol. 37, no. 8, pp. 2419–2433, May 2009.

[51] I. Vashisht, P. Mishra, T. Pal, S. Chanumolu, T. R. Singh, and R. S. Chauhan, “Mining NGS transcriptomes for miRNAs and dissecting their role in regulating growth, development, and secondary metabolites production in different organs of a medicinal herb, Picrorhiza kurroa,” *Planta*, vol. 241, no. 5, pp. 1255–1268, May 2015.

[52] T. R. Singh, A. Gupta, and P. Suravajhala, “Challenges in the miRNA research,” *Int. J. Bioinforma. Res. Appl.*, vol. 9, no. 6, pp. 576–583, Jan. 2013.

[53] M. Selbach, B. Schwanhäusser, N. Thierfelder, Z. Fang, R. Khanin, and N. Rajewsky, “Widespread changes in protein synthesis induced by microRNAs,” *Nature*, vol. 455, no. 7209, pp. 58–63, Sep. 2008.

[54] V. Kumar, R. S. Chauhan, and C. Tandon, “Biosynthesis and therapeutic implications of iridoid glycosides from Picrorhiza genus: the road ahead,” *J. Plant Biochem. Biotechnol.*, vol. 26, no. 1, pp. 1–13, Jan. 2017.

[55] V. Kumar, N. Malhotra, T. Pal, and R. S. Chauhan, “Molecular dissection of pathway components unravel atisine biosynthesis in a non-toxic Aconitum species, A. heterophyllum Wall,” *3 Biotech*, vol. 6, no. 1, Dec. 2016.

[56] A. Conesa *et al.*, “A survey of best practices for RNA-seq data analysis,” *Genome Biol.*, vol. 17, p. 13, 2016.

[57] F. Maghuly, R. C. Ramkat, and M. Laimer, “Virus versus Host Plant MicroRNAs: Who Determines the Outcome of the Interaction?,” *PLOS ONE*, vol. 9, no. 6, p. e98263, Jun. 2014.

[58] S.-D. Hsu *et al.*, “miRTarBase: a database curates experimentally validated microRNA-target interactions,” *Nucleic Acids Res.*, vol. 39, no. Database issue, pp. D163-169, Jan. 2011.

[59] T. Vergoulis *et al.*, “TarBase 6.0: capturing the exponential growth of miRNA targets with experimental support,” *Nucleic Acids Res.*, vol. 40, no. Database issue, pp. D222-229, Jan. 2012.

[60] F. Xiao, Z. Zuo, G. Cai, S. Kang, X. Gao, and T. Li, “miRecords: an integrated resource for microRNA-target interactions,” *Nucleic Acids Res.*, vol. 37, no. Database issue, pp. D105-110, Jan. 2009.

[61] J.-H. Yang, J.-H. Li, P. Shao, H. Zhou, Y.-Q. Chen, and L.-H. Qu, “starBase: a database for exploring microRNA-mRNA interaction maps from Argonaute CLIP-Seq and Degradome-Seq data,” *Nucleic Acids Res.*, vol. 39, no. Database issue, pp. D202-209, Jan. 2011.

[62] E. Dai *et al.*, “EpimiR: a database of curated mutual regulation between miRNAs and epigenetic modifications,” *Database J. Biol. Databases Curation*, vol. 2014, p. bau023, 2014.

[63] J. L. Rukov, R. Wilentzik, I. Jaffe, J. Vinther, and N. Shomron, “Pharmaco-miR: linking microRNAs and drug effects,” *Brief. Bioinform.*, vol. 15, no. 4, pp. 648–659, Jul. 2014.

[64] Q. Jiang *et al.*, “miR2Disease: a manually curated database for microRNA deregulation in human disease,” *Nucleic Acids Res.*, vol. 37, no. Database issue, pp. D98-104, Jan. 2009.

[65] X. Liu *et al.*, “SM2miR: a database of the experimentally validated small molecules’ effects on microRNA expression,” *Bioinforma. Oxf. Engl.*, vol. 29, no. 3, pp. 409–411, Feb. 2013.

[66] A. Ruepp, A. Kowarsch, and F. Theis, “PhenomiR: microRNAs in human diseases and biological processes,” *Methods Mol. Biol. Clifton NJ*, vol. 822, pp. 249–260, 2012.

[67] J. M. Schmiedel *et al.*, “MicroRNA control of protein expression noise,” *Science*, vol. 348, no. 6230, pp. 128–132, Apr. 2015.

[68] A. Sood and R. S. Chauhan, “Comparative NGS Transcriptomics Unravels Molecular Components Associated with Mosaic Virus Infection in a Bioenergy Plant Species, Jatropha curcas L.,” *Bioenergy Res.*, 2016.

[69] B. Zhang, X. Pan, and T. A. Anderson, “Identification of 188 conserved maize microRNAs and their targets,” *FEBS Lett.*, vol. 580, no. 15, pp. 3753–3762, Jun. 2006.

[70] S. Griffiths-Jones, R. J. Grocock, S. van Dongen, A. Bateman, and A. J. Enright, “miRBase: microRNA sequences, targets and gene nomenclature,” *Nucleic Acids Res.*, vol. 34, no. suppl 1, pp. D140–D144, Jan. 2006.

[71] Z. Zhang *et al.*, “PMRD: plant microRNA database,” *Nucleic Acids Res.*, vol. 38, no. suppl 1, pp. D806–D813, Jan. 2010.

[72] M. Zuker, “Mfold web server for nucleic acid folding and hybridization prediction,” *Nucleic Acids Res.*, vol. 31, no. 13, pp. 3406–3415, Jul. 2003.

[73] X. Dai and P. X. Zhao, “psRNATarget: a plant small RNA target analysis server,” *Nucleic Acids Res.*, vol. 39, no. Web Server issue, pp. W155-159, Jul. 2011.

[74] J. X. Hu, C. E. Thomas, and S. Brunak, “Network biology concepts in complex disease comorbidities,” *Nat. Rev. Genet.*, vol. 17, no. 10, pp. 615–629, Oct. 2016.

[75] A. Conesa, S. Götz, J. M. García-Gómez, J. Terol, M. Talón, and M. Robles, “Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research,” *Bioinformatics*, vol. 21, no. 18, pp. 3674–3676, Sep. 2005.

[76] S. J. C. Gosline *et al.*, “Elucidating microRNA regulatory networks using transcriptional, post-transcriptional and histone modification measurements,” *Cell Rep.*, vol. 14, no. 2, pp. 310–319, Jan. 2016.

[77] Y. Meng, C. Shao, H. Wang, and M. Chen, “The Regulatory Activities of Plant MicroRNAs: A More Dynamic Perspective,” *Plant Physiol.*, vol. 157, no. 4, pp. 1583–1595, Dec. 2011.

[78] A. Zewail *et al.*, “Novel functions of the phosphatidylinositol metabolic pathway discovered by a chemical genomics screen with wortmannin,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 100, no. 6, pp. 3345–3350, Mar. 2003.

[79] J. E. F. Butler and J. T. Kadonaga, “The RNA polymerase II core promoter: a key component in the regulation of gene expression,” *Genes Dev.*, vol. 16, no. 20, pp. 2583–2592, Oct. 2002.

[80] B. M. Lange, T. Rujan, W. Martin, and R. Croteau, “Isoprenoid biosynthesis: The evolution of two ancient and distinct pathways across genomes,” *Proc. Natl. Acad. Sci.*, vol. 97, no. 24, pp. 13172–13177, Nov. 2000.

[81] I. Jonkers and J. T. Lis, “Getting up to speed with transcription elongation by RNA polymerase II,” *Nat. Rev. Mol. Cell Biol.*, vol. 16, no. 3, pp. 167–177, Mar. 2015.

[82] V. Kumar, N. Sharma, H. Sood, and R. S. Chauhan, “Exogenous feeding of immediate precursors reveals synergistic effect on picroside-I biosynthesis in shoot cultures of Picrorhiza kurroa Royle ex Benth,” *Sci. Rep.*, vol. 6, p. 29750, Jul. 2016.

[83] V. Kumar *et al.*, “An insight into conflux of metabolic traffic leading to picroside-I biosynthesis by tracking molecular time course changes in a medicinal herb, Picrorhiza kurroa,” *Plant Cell Tissue Organ Cult. PCTOC*, vol. 123, no. 2, pp. 435–441, Nov. 2015.

[84] V. Kumar, K. Shitiz, R. S. Chauhan, H. Sood, and C. Tandon, “Tracking dynamics of enzyme activities and their gene expression in Picrorhiza kurroa with respect to picroside accumulation,” *J. Plant Biochem. Biotechnol.*, vol. 25, no. 2, pp. 125–132, Apr. 2016.

[85] R. Shankar, A. Bhattacharjee, and M. Jain, “Transcriptome analysis in different rice cultivars provides novel insights into desiccation and salinity stress responses,” *Sci. Rep.*, vol. 6, p. 23719, Mar. 2016.

[86] R. N. Trigiano, *Plant Pathology Concepts and Laboratory Exercises, Second Edition*. CRC Press, 2007.

[87] M. S. Hussain, S. Fareed, S. Ansari, M. A. Rahman, I. Z. Ahmad, and M. Saeed, “Current approaches toward production of secondary plant metabolites,” *J. Pharm. Bioallied Sci.*, vol. 4, no. 1, pp. 10–20, 2012.

[88] Y.-W. Sun, C.-S. Tee, Y.-H. Ma, G. Wang, X.-M. Yao, and J. Ye, “Attenuation of Histone Methyltransferase KRYPTONITE-mediated transcriptional gene silencing by Geminivirus,” *Sci. Rep.*, vol. 5, p. 16476, Nov. 2015.

[89] H. Sanfaçon, “Investigating the role of viral integral membrane proteins in promoting the assembly of nepovirus and comovirus replication factories,” *Front. Plant Sci.*, vol. 3, Jan. 2013.

[90] M. Kanehisa, Y. Sato, M. Kawashima, M. Furumichi, and M. Tanabe, “KEGG as a reference resource for gene and protein annotation,” *Nucleic Acids Res.*, vol. 44, no. D1, pp. D457-462, Jan. 2016.

[91] S. M. Gasser and E. Li, “Epigenetics and disease: pharmaceutical opportunities. Preface,” *Prog. Drug Res. Fortschritte Arzneimittelforschung Progres Rech. Pharm.*, vol. 67, pp. v–viii, 2011.

[92] A. Bansal, T. R. Singh, and R. S. Chauhan, “A novel miRNA analysis framework to analyze differential biological networks,” *Sci. Rep.*, vol. 7, no. 1, p. 14604, Nov. 2017.

[93] J. Menche *et al.*, “Integrating personalized gene expression profiles into predictive disease-associated gene pools,” *Npj Syst. Biol. Appl.*, vol. 3, no. 1, p. 10, Mar. 2017.

[94] M. C. Guzman-Karlsson, J. P. Meadows, C. F. Gavin, J. J. Hablitz, and J. D. Sweatt, “Transcriptional and epigenetic regulation of Hebbian and non-Hebbian plasticity,” *Neuropharmacology*, vol. 80, pp. 3–17, May 2014.

[95] G. Joshi-Tope *et al.*, “Reactome: a knowledgebase of biological pathways,” *Nucleic Acids Res.*, vol. 33, no. Database issue, pp. D428-432, Jan. 2005.

[96] H. Zhou, M. Gao, and J. Skolnick, “Comprehensive prediction of drug-protein interactions and side effects for the human proteome,” *Sci. Rep.*, vol. 5, Jun. 2015.

[97] X. Zhou, J. Menche, A.-L. Barabási, and A. Sharma, “Human symptoms–disease network,” *Nat. Commun.*, vol. 5, p. 4212, Jun. 2014.

[98] D. Gomez-Cabrero *et al.*, “From comorbidities of chronic obstructive pulmonary disease to identification of shared molecular mechanisms by data integration,” *BMC Bioinformatics*, vol. 17, no. 15, p. 441, Nov. 2016.

[99] M. Kitsak *et al.*, “Tissue Specificity of Human Disease Module,” *Sci. Rep.*, vol. 6, p. 35241, Oct. 2016.

[100] G. Basler, Z. Nikoloski, A. Larhlimi, A.-L. Barabási, and Y.-Y. Liu, “Control of fluxes in metabolic networks,” *Genome Res.*, p. gr.202648.115, May 2016.

[101] A. Bansal and P. A. Srivastava, “Transcriptomics to Metabolomics: A Network Perspective for Big Data,” *Httpservicesigi-Glob.-1-5225-2607-0ch008*, pp. 188–206, 2018.

[102] S. D. Ghiassian *et al.*, “Endophenotype Network Models: Common Core of Complex Diseases,” *Sci. Rep.*, vol. 6, p. 27414, Jun. 2016.

[103] E. Guney, J. Menche, M. Vidal, and A.-L. Barábasi, “Network-based in silico drug efficacy screening,” *Nat. Commun.*, vol. 7, p. 10331, Feb. 2016.

[104] A. Vinayagam *et al.*, “Controllability analysis of the directed human protein interaction network identifies disease genes and drug targets,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 113, no. 18, pp. 4976–4981, May 2016.

[105] F.-R. Meng, Z.-H. You, X. Chen, Y. Zhou, and J.-Y. An, “Prediction of Drug-Target Interaction Networks from the Integration of Protein Sequences and Drug Chemical Structures,” *Mol. Basel Switz.*, vol. 22, no. 7, Jul. 2017.

[106] X. Chen, “Editorial: Identifying Drug-target Interactions Based on Heterogeneous Biological Data - PART 1,” *Curr. Protein Pept. Sci.*, vol. 19, no. 5, pp. 428–429, 2018.

[107] X. Chen *et al.*, “NRDTD: a database for clinically or experimentally supported non-coding RNAs and drug targets associations,” *Database J. Biol. Databases Curation*, vol. 2017, 01 2017.

[108] X. Chen *et al.*, “Drug-target interaction prediction: databases, web servers and computational models,” *Brief. Bioinform.*, vol. 17, no. 4, pp. 696–712, 2016.

[109] X. Chen, M.-X. Liu, and G.-Y. Yan, “Drug-target interaction prediction by random walk on the heterogeneous network,” *Mol. Biosyst.*, vol. 8, no. 7, pp. 1970–1978, Jul. 2012.

[110] L. Wang, Z.-H. You, X. Chen, X. Yan, G. Liu, and W. Zhang, “RFDT: A Rotation Forest-based Predictor for Predicting Drug-Target Interactions Using Drug Structure and Protein Sequence Information,” *Curr. Protein Pept. Sci.*, vol. 19, no. 5, pp. 445–454, 2018.

[111] Y.-A. Huang, Z.-H. You, and X. Chen, “A Systematic Prediction of Drug-Target Interactions Using Molecular Fingerprints and Protein Sequences,” *Curr. Protein Pept. Sci.*, vol. 19, no. 5, pp. 468–478, 2018.

[112] L. Wang *et al.*, “A Computational-Based Method for Predicting Drug-Target Interactions by Using Stacked Autoencoder Deep Neural Network,” *J. Comput. Biol. J. Comput. Mol. Cell Biol.*, vol. 25, no. 3, pp. 361–373, Mar. 2018.

[113] J. Menche *et al.*, “Disease networks. Uncovering disease-disease relationships through the incomplete interactome,” *Science*, vol. 347, no. 6224, p. 1257601, Feb. 2015.

[114] S. D. Ghiassian, J. Menche, and A.-L. Barabási, “A DIseAse MOdule Detection (DIAMOnD) Algorithm Derived from a Systematic Analysis of Connectivity Patterns of Disease Proteins in the Human Interactome,” *PLOS Comput. Biol.*, vol. 11, no. 4, p. e1004120, Apr. 2015.

[115] R. Chander, N. K. Kapoor, and B. N. Dhawan, “Picroliv, picroside-I and kutkoside from Picrorhiza kurrooa are scavengers of superoxide anions,” *Biochem. Pharmacol.*, vol. 44, no. 1, pp. 180–183, Jul. 1992.

[116] Y. Dwivedi, R. Rastogi, N. K. Garg, and B. N. Dhawan, “Picroliv and its Components Kutkoside and Picroside I Protect Liver Against Galactosamine-Induced Damage in Rats,” *Pharmacol. Toxicol.*, vol. 71, no. 5, pp. 383–387, Nov. 1992.

[117] H. Gao and Y.-W. Zhou, “Anti-lipid peroxidation and protection of liver mitochondria against injuries by picroside II,” *World J. Gastroenterol.*, vol. 11, no. 24, pp. 3671–3674, Jun. 2005.

[118] P. Pratheeshkumar, Y.-O. Son, P. Korangath, K. A. Manu, and K. S. Siveen, “Phytochemicals in Cancer Prevention and Therapy,” *BioMed Res. Int.*, vol. 2015, 2015.

[119] Y. Kılıç *et al.*, “Effect of picroside II on hind limb ischemia reperfusion injury in rats,” *Drug Des. Devel. Ther.*, vol. 11, pp. 1917–1925, 2017.

[120] D. Rathee, M. Thanki, S. Bhuva, S. Anandjiwala, and R. Agrawal, “Iridoid glycosides-Kutkin, Picroside I, and Kutkoside from Picrorrhiza kurroa Benth inhibits the invasion and migration of MCF-7 breast cancer cells through the down regulation of matrix metalloproteinases: 1st Cancer Update,” *Arab. J. Chem.*, vol. 6, no. 1, pp. 49–58, Jan. 2013.

[121] T. Rolland *et al.*, “A proteome-scale map of the human interactome network,” *Cell*, vol. 159, no. 5, pp. 1212–1226, Nov. 2014.

[122] M. Tanabe and M. Kanehisa, “Using the KEGG database resource,” *Curr. Protoc. Bioinforma.*, vol. Chapter 1, p. Unit1.12, Jun. 2012.

[123] D. Schneidman-Duhovny, Y. Inbar, R. Nussinov, and H. J. Wolfson, “PatchDock and SymmDock: servers for rigid and symmetric docking,” *Nucleic Acids Res.*, vol. 33, no. Web Server issue, pp. W363–W367, Jul. 2005.

[124] J. Zhu *et al.*, “A pre-clinical pharmacokinetic study in rats of three naturally occurring iridoid glycosides, Picroside-I, II and III, using a validated simultaneous HPLC-MS/MS assay,” *J. Chromatogr. B Analyt. Technol. Biomed. Life. Sci.*, vol. 993–994, pp. 47–59, Jul. 2015.

[125] X. Liu *et al.*, “PharmMapper server: a web server for potential drug target identification using pharmacophore mapping approach,” *Nucleic Acids Res.*, vol. 38, no. Web Server issue, pp. W609-614, Jul. 2010.

[126] E. Eden, R. Navon, I. Steinfeld, D. Lipson, and Z. Yakhini, "GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists," *BMC Bioinformatics*, vol. 10, p. 48, Feb. 2009.

[127] P. Singh, P. Ravanan, and P. Talwar, "Death Associated Protein Kinase 1 (DAPK1): A Regulator of Apoptosis and Autophagy," *Front. Mol. Neurosci.*, vol. 9, Jun. 2016.

[128] W. E. Paul and J. Zhu, "How are $T_H$2-type immune responses initiated and amplified?," *Nat. Rev. Immunol.*, vol. 10, no. 4, pp. 225–235, Apr. 2010.

[129] M. Burotto, V. L. Chiou, J.-M. Lee, and E. C. Kohn, "The MAPK pathway across different malignancies: A new perspective," *Cancer*, vol. 120, no. 22, pp. 3446–3456, Nov. 2014.

[130] S. J. Thomas, J. A. Snowden, M. P. Zeidler, and S. J. Danson, "The role of JAK/STAT signalling in the pathogenesis, prognosis and treatment of solid tumours," *Br. J. Cancer*, vol. 113, no. 3, pp. 365–371, Jul. 2015.

# PUBLICATIONS

**Publications**

## Core Thesis oriented Publications

- **Bansal, Ankush**, Tiratha Raj Singh, and Rajinder Singh Chauhan. "A novel miRNA analysis framework to analyze differential biological networks." ***Scientific reports*** 7.1 (2017): 14604.
- **Bansal, Ankush**, Pulkit Anupam Srivastava, and Tiratha Raj Singh. "An integrative approach to develop computational pipeline for drug-target interaction network analysis." ***Scientific reports*** 8.1 (2018): 10238.
- **Bansal, Ankush**, and Tiratha Raj Singh. "A system level network analysis framework to snare transitions in metabolic pathways." ***Systems Biology and Applications***. [In-Process]

## Publications derived from outcome of thesis based pipelines

- **Bansal, Ankush**, Mehul Salaria, Tashil Sharma, Tsering Stobdan, and Anil Kant. "Comparative de novo transcriptome analysis of male and female Sea buckthorn." ***3 Biotech*** 8, no. 2 (2018): 96.
- Kumar, Varun, **Ankush Bansal**, and Rajinder S. Chauhan. "Modular design of picroside-II biosynthesis deciphered through NGS transcriptomes and metabolic intermediates analysis in naturally variant chemotypes of a medicinal herb, Picrorhiza kurroa." ***Frontiers in plant science*** 8 (2017): 564.
- Gangwar, Manali, Archit Sood, **Ankush Bansal**, and Rajinder Singh Chauhan. "Comparative transcriptomics reveals a reduction in carbon capture and flux between source and sink in cytokinin-treated inflorescences of Jatropha curcas L." ***3 Biotech*** 8, no. 1 (2018): 64.
- Vashisht, Ira, Tarun Pal, **Ankush Bansal**, and Rajinder Singh Chauhan. "Uncovering interconnections between kinases vis-à-vis physiological and biochemical processes contributing to picroside-I biosynthesis in a medicinal herb, Picrorhiza kurroa Royle ex. Benth." ***Acta Physiologiae Plantarum*** 40, no. 6 (2018): 115.

## Other Publications

- Panigrahi, Priya P., Ramit Singla, **Ankush Bansal**, Moacyr C. Junior, Vikas Jaitak, Ragothaman M. Yennamalli, and Tiratha Raj Singh. "*In silico* Screening and Molecular Interaction Studies of Tetrahydrocannabinol and its Derivatives with Acetylcholine Binding Protein." ***Current Chemical Biology*** 12, no. 2 (2018): 181-190.
- Kumar, Rutash†, **Ankush Bansal†**, Rohit Shukla, Tiratha Raj Singh, Pramod Wasudeo Ramteke, Satendra Singh, and Budhayash Gautam. "*In silico* screening of deleterious single nucleotide polymorphisms (SNPs) and molecular dynamics simulation of disease associated mutations in gene responsible for Oculocutaneous Albinism type 6 (OCA 6) disorder." ***Journal of Biomolecular Structure & Dynamics***. [In-Press]

## Book Chapters

- **Ankush Bansal**, and Pulkit Anupam Srivastava. "Transcriptomics to Metabolomics: A Network Perspective for Big Data." In *Applying Big Data Analytics in Bioinformatics and Medicine*, pp. 188-206. IGI Global, 2018.[Published]
- **Ankush Bansal**, Mehul Salaria, and Tiratha Raj Singh. "Tau Pathology: A Step Towards Understanding Neurodegenerative Disorders Network Complexity." In *Handbook of Research on Critical Examinations of Neurodegenerative Disorders*, pp. 217-234. IGI Global, 2019.[Published]
- Tiratha Raj Singh, **Bansal Ankush**, "Phylogenetic Analysis: Gene Duplication and Speciation." In *Encyclopedia of Bioinformatics and Computational Biology*, pp. 965-974. Elsevier, 2019 [Published]
- **Ankush Bansal**, Tiratha Raj Singh*, "Epigenome-Wide DNA Methylation Profiling in Alzheimer's Disease." In *Computational Epigenetics and Disease*. [In-Press]
- **Ankush Bansal**, Siddhant Kalra, Tiratha Raj Singh, "Modeling and Optimization of Molecular Bio-systems to Generate Predictive Models." In *Essentials of Bioinformatics Vol I.*[In-Press]

## Conferences

- **Bansal Ankush**, Anil Kant, and Tiratha Raj Singh, "A network analysis framework to decipher uni-directional information flow from differential transcriptome datasets for pathway of interest" in ***Indian Conference on Bioinformatics (Inbix'17)***, Birla Institute of Scientific Research (BISR), Jaipur, Rajasthan, India during Nov., 7-9, 2017.
- **Bansal Ankush**, Mehul Salaria, Tashil Sharma, and Anil Kant"Comparative miRNA mining of Sea buckthorn male and female transcriptomes" in ***Seabuckthorn for Improving Health and Sustainable Development of Himalayan Region***, Defence Institute of High Altitude Research (DIHAR-DRDO), Leh, Jammu and Kashmir, India during Sep., 22-24, 2017.