NutriGeneInteract: An Interactive Tool for Nutrigenomics

A Major Project report submitted

In partial fulfilment of the requirement for the degree of

Master of Technology (M. Tech)

In

Bioinformatics

Submitted by

Mohini Yadav

(2K15/BIO/08)

Under the supervision of

Dr. Yasha Hasija

Department of Biotechnology

Delhi Technological University

(Formerly Delhi College of Engineering)

Shahbad Daulatpur, Main Bawana Road,

Delhi-110042, INDIA

DECLARATION

I declare that the project report entitled "**NutriGeneInteract: An Interactive Tool for Nutrigenomics**", submitted by Mohini Yadav is in the partial fulfilment of the requirement for the award of the degree of Master of Technology in Bioinformatics, Department of Biotechnology, Delhi Technological University. It is a record of original research work carried out by me under the supervision of **Dr. Yasha Hasija**, Department of Biotechnology, Delhi Technological University, Delhi.

The matter reported in this project report is original and has not been submitted for the award of any other degree.

Mohini Yadav

2K15/BIO/08

Department of Biotechnology

Delhi Technological University

(Formerly Delhi College of Engineering)

Date:

ACKNOWLEDGEMENT

First of all, I will like to express my gratitude to my open minded guide and mentor, Dr. Yasha Hasija, for her continuous encouragement, support and critical suggestions throughout of this work. There were times when she kept on having faith in me, when I was a little apprehensive about the progress of my work; I truly appreciate that she always gave us different opportunities to groom our knowledge and skills.

Also, I would like to express my gratitude to other faculty members for the valuable knowledge that they have imparted through their courses. I would also acknowledge the support and help provided by the staff of the department.

I would also like to express my pleasant appreciation to Lakshaya, for providing me with the initial data for Nutrition related Genes and Disorders. Also, my sincere thanks to my senior Ms. Isha Srivastava for helping me in project guidelines.

And finally, my parents and friends, for always giving me emotional support and determination. It is only because of them whatever I am able to do today and whatever good I will achieve in future.

Mohini Yadav

(2K15/BIO/08)

CONTENTS

NutriGeneInteract: An Interactive Tool for Nutrigenomics

Mohini Yadav

*Delhi Technological University, Delhi, India

Miniyad8100@gmail.com

ABSTRACT

Post-human genome disclosure observes the emergence of 'Nutigenomics' as one of the exciting scientific advancement influencing mankind around the world. Nutrition has the major impact in defining the cause-response interaction between nutrient and human health. Each person differs from the next by an average of over 3 million genetic variations in their DNA. This genetic diversity is responsible for many of the inter individual differences in food preferences, nutritional needs, and dietary responses between humans. The field of nutrigenetics aims to utilize this type of genetic information in order to personalize diets for optimal health.

Studies have demonstrated that network approaches helps when applied to integrate biological information can provide novel insights and pave the way for understanding and curing complex diseases. Here we, therefore, propose a new platform NutriGeneInteract that integrate different data and allow the creation of heterogeneous networks of gene, protein, disease and nutrition interaction levels. NutriGeneInteract also provides a user-friendly interface to integrate, visualize and analyse genome-scale biological networks for Nutrition related diseases ultimately allowing the research community to connect distinct spots in space to solve the puzzle of Nutritional effects on genetics. NutriGeneInteract is available online at **http://genomeinformatics.dtu.ac.in/NutriGeneInteract/Index.html.**

1. INTRODUCTION

Nutrigenomics is an emerging science which investigates a certain area of nutrition that uses molecular tools to search, access and understand the several responses obtained through a certain diet applied between individual and population groups (Sales et al., 2014). The term "nutrigenomics" was first described in 2001 from Pelegrin(2001) and then, it appears in 2002 in a review by Van Ommen and Stierum (2002). Nutrigenomics application to everyday life would be the future of the Nutrition science and a great series of tools for nutritionists, dietitians, doctors as well as any Health Professional that implicates nutrition therapy for the treatment of disease. It could possibly help into the prevention of diet-related diseases as well as the designing of nutritional strategies and the adverse or beneficial effects of some food or nutrients (Palou, 2007). Currently, personalised medicine and nutrition are not completely applied into the everyday routine of the patients and their carers; doctors, nurses, dietitians or nutritionists. In a recent review, it is described how nutrigenomics should be inserted in public health and how the need for the gene–diet–disease interaction was created in the last years (Neeha and Priyamvadah, 2013).

Nutrigenomics was born due to the need to pass from Epidemiology and Physiology to Molecular Biology and Genetics (Neeha and Priyamvadah, 2013). It is very important to evaluate the genes, the gene/protein network and the method in order to research and check on the influence of the nutrients on gene/protein expression (Neeha and Priyamvadah, 2013).

Traditionally, nutritional science was mainly concentrated on nutrient deficiencies and their effects on health and disease. However, over the past few decades, research emphasis has gradually shifted to the link between (over)-nutrition and chronic diseases, including cancer, obesity, cardiovascular disease, and diabetes (M. Muller, 2003 ; J.M. Ordovas, 2004). Nutrigenomics may provide the strategies for the development of safe and effective dietary interventions against the obesity epidemic. According to the World Health Organization, more than 60% of the global disease burden will be attributed to chronic disorders associated with obesity by 2020. Meanwhile in the US, the prevalence of obesity has doubled in adults and tripled in children during the past three decades. In this regard, a number of natural dietary supplements and micronutrients have been studied for their potential in weight management. Inherited character, genetic make-up and availability of nutritious food is not only factors symbolic to health status, but it is much more complex interaction between individual's genome (entire set of genes present) and environmental factors faced over lifetime. In addition,

nutritional level, concentration of bioactive and their ability to influence health status are some of the crucial factors, which need to be addressed in studies aiming food-nutrient-health interaction. Thus, the science involved in development of dietary supplement, nutraceuticals or functional food, especially for disease prevention is highly complex.

Alteration in functioning of particular gene or their protein product during a particular disease is actually influenced by subsequent modification in nutritional composition of food (WHO 1990). The phenomenon of interaction between nutrients - genes - diseases is a highly complex. However, the scientific pursuit starting from isolation of DNA, interpretation of its structure and finally revelation of human genome help to resolve many such complexities. The realization of these scientific breakthroughs was only possible due to parallel technological advancements, especially in the area of 'Omics' and 'Bioinformatics' (Ovesná J et al 2008). Post human genome era led to surfacing of 'Nutrigenomics' a nuptials between nutrition and genomics. It brings together the science of nutrition, genomics, computational biology & bioinformatics, and molecular medicine to tackle these chronic diseases. It enables nutritionist to understand potential of food fortification or supplementation of diet with particular nutrient and its molecular effect in human health and reduce the risk of life style related disease. It render help in development of designed food for individuals with specific genetic backgrounds and also enable policymakers to strategize the development of safer and more effective dietary interventions.

2. REVIEW OF LITERATURE

This section introduces the commonly used network terminology and definitions followed by the applications of network biology in the medicine.

2.1 NETWORK THEORY

A network is a set of nodes connected by interactions, and can represent anything from social interactions between pupils in a school class, to traffic between airports. Organizing complex systems into networks can give a clear overview of the system, and effectively identify important components.

The analysis of complex networks constitutes a field of science known as network theory. Network theory is a powerful tool for systems biologists. Disease-associated genes can be organized into networks using various biological sources such as physical protein-protein interactions (PPIs), literature co-citation or co-expression. Those networks can then be dissected to identify disease pathways, mechanisms, clinical markers and drug targets. Two concepts are of particular interest in this kind of analysis: modularity and centrality.

2.1.1 Modularity

In a modular network, the nodes are divided into groups that share many interactions internally but few interactions with other groups. Many real world networks, such as social networks and gene networks tend to be highly modular. Identifying modules in a large, complex network can reduce thousands of nodes to a handful of modules, thereby drastically simplifying the analysis.

The most extreme version of module is termed clique. A clique is a set of nodes that are completely connected with each other. All cliques in a network that are not subsets of another clique, are termed as maximal cliques.

A method to determine if a gene is part of a module or not is termed clustering coefficient (see Table 1 for the description of properties of a node in a network). The clustering coefficient of a node is defined by how many of the node's interactors that are connected with each other. Average clustering coefficient of a network calculated as the mean of the clustering coefficients of all the nodes in the network gives a measure of cohesiveness in the network which is also commonly referred to as the extent of modularity. The higher the clustering coefficient greater is the modular nature of the network. To compare the extent of cohesiveness in a network often

clustering coefficients of the real networks are compared with random networks with similar size and degree distribution.

Table 1. *Different local properties which can be defined for a node in complex networks.*

Property	Description
Incoming or Indegree degree	Indegree refers to the number of incoming connections to a node of interest. In other words, indegree is the number of arrows that flow into the node under investigation.
Outgoing or Outdegree degree	Out degree refers to the number of edges which start from a node of interest and point to other nodes in the network and is valid for directed networks where there is direction associated with each edge represented.
Connectivity or Degree	Degree or connectivity of a node refers to the total number of interactions it has in a network – the higher the connectivity (i.e., hub nodes) the more the number of targets it interacts with. In directed networks degree simply corresponds to the sum of in and out degrees of a node.
Clustering Coefficient	Clustering coefficient of a node reflects the extent to which the neighbors of a given node are interconnected among themselves to what is expected theoretically and indicates the cohesiveness

2.1.2 Centrality

Centrality is a way to prioritize nodes within a network. A central node is well-connected with the rest of the network; removing it will strongly affect the integrity of the network. The measures of centrality are: Degree, Betweenness and Closeness.

A. Network degree

The most straightforward definition of centrality is degree (sometimes called connectivity), which corresponds to the number of interactions a node has. Calculation of the degree allows determining the degree distribution P(k), which gives the probability that a selected node has exactly k links. P(k) is obtained by counting the number of nodes N(k) with k = 1, 2, 3 ... links and dividing by the total number of nodes N. Determining the degree distribution allows distinguishing different kind of graphs. For instance, a graph with a peaked degree distribution (Gaussian distribution) indicates that the system has a characteristic degree with no highly connected nodes (most of the nodes have average degree). This is typical of random, non-natural, networks. By contrast, a power-law degree distribution indicates the presence of few nodes having a very high degree. And this behaviour has been shown to be non-random. Networks displaying a degree distribution approximating a power-law are called scale-free networks. Many networks have been identified to show the characteristics of the scale free networks.

Nodes with a high degree are commonly referred to as hubs and they hold together several nodes with lower degree. This behaviour gives these networks a kind of robustness against random node deletion/failure. Additionally it also gives an opportunity to identify few network influential nodes in disease related biological networks as potential therapeutic targets.

In biological terms, the degree allows an immediate evaluation of the regulatory relevance of the node. For instance, in signaling networks, proteins with very high degree are interacting with several other signaling proteins, thus suggesting a central regulatory role of the proteins, that is they are likely to be regulatory hubs. 10

However, degree is a local centrality measure. It is often important to know a node's centrality with respect to the entire network. Two common global centrality measures are betweenness and closeness (Junker et al., 2006).

B. Betweenness Centrality

Betweenness is calculated by first identifying the shortest paths between all nodes in the network. The betweenness centrality of a node n is calculated considering couples of nodes (s and t) and counting the number of shortest paths linking node s and node t and passing through the node n . Then, the value is related to the total number of shortest paths linking s and t. Thus, a node can be traversed by only one path linking s and t, but if this path is the only connecting s and t the node i will score a higher betweenness value. Nodes with many shortest paths passing through them receive a high betweenness and are sometimes referred to as bottlenecks.

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

C. Closeness Centrality

Closeness is calculated in a similar way. A node's closeness is a measure of its average distance to all other nodes in the network. The closeness centrality of a node is calculated by computing the sum of the shortest path between the node and all other nodes in the graph, and then dividing by the number of nodes. Once this value is obtained, its reciprocal is calculated, so higher values assume a positive meaning in term of node proximity. Notably, high values of closeness centrality should indicate that all other nodes are in proximity to the node. In contrast, low values of closeness should indicate that all other nodes are distant from the node. 11

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

The inverse of closeness is called average shortest path length. Since closeness is based on the average distance to all other nodes in the network, some nodes could still be very distant from a node with high closeness. To compensate for this, closeness should be complemented with

eccentricity, which is the distance from a given node to the farthest node in the network. Note that unlike for degree, betweenness and closeness, a low eccentricity value implies high centrality.

2.2 NETWORKS IN MOLECULAR BIOLOGY

Development of several high throughput approaches in the last decade have not only increased the amount of information that we could gather to reveal important insights on the transcriptional, post-transcriptional or functional organization of an organism but they have also enabled us to start our journey to uncover the principles which hold them together. This is mainly because of the extent of information that has been possible to be collected by interrogating the cell's environment at different levels of detail. For instance, availability of modern techniques now enable us to identify the set of protein-protein interactions, genetic interactions, metabolic maps and small molecule interactions at a whole-organism level. Perhaps the most common form of interaction graphs which have been studied since the early days of genome sequencing are protein interactions.

2.2.1 Protein-Protein Interactions

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

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

computational modeling to determine which proteins may bind to each other based on their (predicted) structural properties. Some of the most popular and widely used experimental and computational techniques are explained below:

a. Yeast two hybrid (Y2H) (Fields and Song, 1989) **:** It is the genetic method that uses the transcriptional activity as a measure of protein-protein interaction. Two hybrid proteins are created corresponding to the proteins A and B between which the interaction is to be identified. One protein(say A) is bind with the DNA binding domain. Other protein(B) is fused with the transcription activation domain. These two hybrids are expressed in a cell containing reporter genes. The positive expression of the reporter genes identifies the interaction between the proteins A and B.

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

c. Co-occurrence: This method exploits the co-occurrence of homologous pairs of genes across multiple genomes. The fact that a pair of genes express together across many 13 different species suggest that these genes are functional associated or physically interacting.

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

1. Biomolecular Interaction Network Database - BIND

2. Database of Interacting Proteins - DIP

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

4. General Repository for Interaction Datasets - GRID

5. Human Protein Reference Database - HPRD

6. Molecular Interactions Database –MINT

PPI Networks

The protein-protein interaction network is generally represented as undirected graph G(V,E), where the set of nodes V are the proteins. An edge (p1, p2)->E is present if there is an interaction between the two proteins p1 and p2. Multiple sources of protein interaction

networks from different studies and databases represent the protein interaction networks differently.

2.2.2 Metabolic Networks

Another class of networks which are commonly studied is that of metabolic networks. They comprise of representing the metabolites and enzymes involved in catalyzing metabolic reactions as the nodes and edges in a directed network. The metabolic network maps are likely the most comprehensive of all biological networks. Most of the work on understanding metabolic networks relies on either manually curated or semi-automated metabolic databases such as the kyoto encyclopedia of genes and genomes (KEGG) and Metacyc which are available for a wide range of model organisms (Caspi et 14 al., 2008; Grossetete et al., Kanehisa et al., 2008). Recently, Duarte et al. published a comprehensive literature-based genome-scale metabolic reconstruction of human metabolism with 2,766 metabolites and 3,311 metabolic and transport reactions. An independent manual construction by Ma et al. contains nearly 3,000 metabolic reactions, organized into about 70 human-specific metabolic pathways.

2.2.3 Regulatory Networks

Mapping of the human regulatory network is in its infancy, making this network perhaps the most incomplete among all biological networks. Data generated by experimental techniques, such as ChIP-on-chip and ChIP-Sequencing, have started to be collected in databases such as Universal Protein Binding Microarray Resource for Oligonucleotide Binding Evaluation (UniPROBE) and JASPAR. Literature-curated and predicted protein-DNA interactions have been compiled in various databases, such as TRANSFAC and the B-cell interactome (BCI). Human post-translational modifications can be found in databases such as Phospho.ELM, PhosphoSite, and phosphorylation site database (PHOSIDA).

3. MATERIALS USED

3.1 Cytoscape

Cytoscape (Figure 2) is an open source bioinformatics software platform for visualizing molecular interaction networks and integrating with gene expression profiles and other state data. Additional features are available as plugins. Plugins are available for network and molecular profiling analyses, new layouts, additional file format support and connection with databases and searching in large networks.

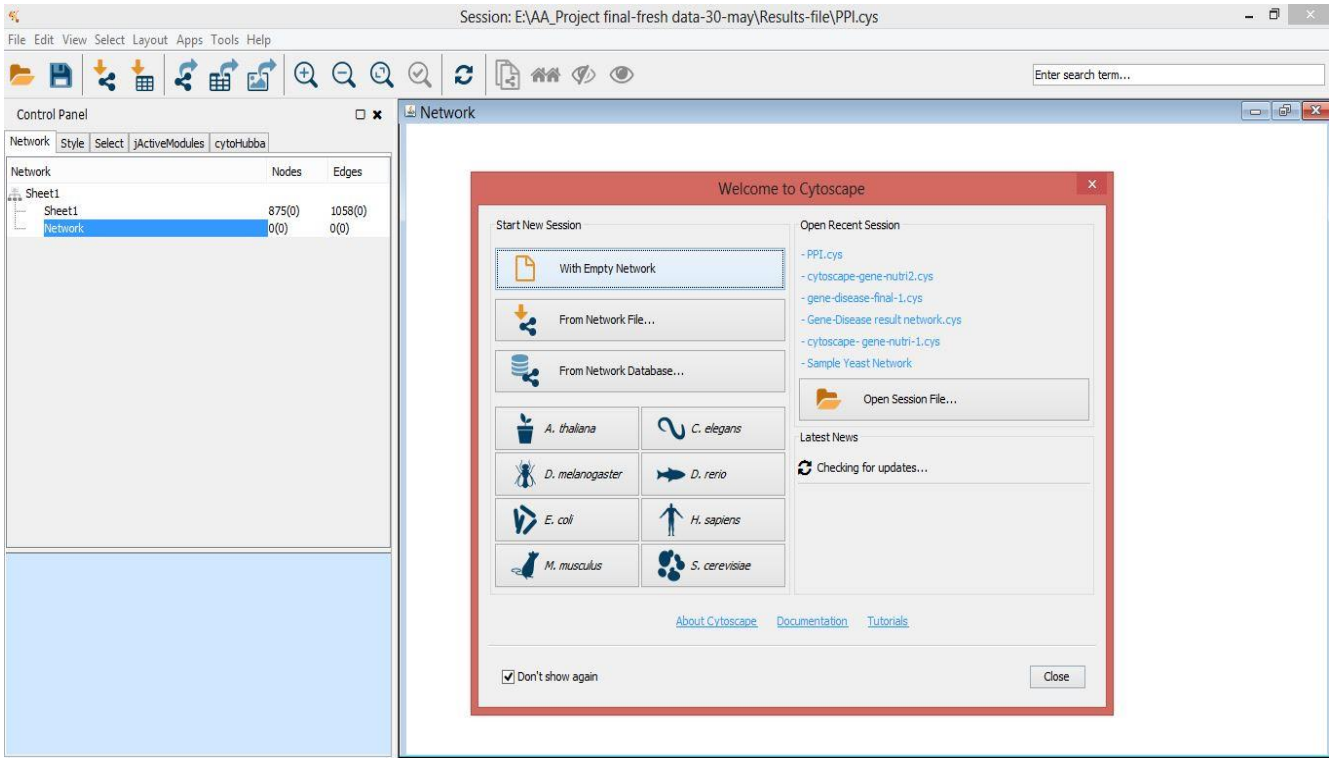


Figure 2. *Window of Cytoscape, Different panels for different purposes viz visualization, control panel, table panel and result panel.*

3.2 Cytoscape plugins

Cytoscape aims to satisfy the unique needs of biologists needing to interactively explore biological networks, such as metabolic pathways or gene regulatory networks, in the context of corresponding experimental data. Cytoscape differentiates itself from other network visualization tools by enabling and encouraging active third-party development of add-on visualization and analysis Apps. A large number of Cytoscape Apps (known previously as plugins) designed for biological networks are available from the Cytoscape App Store.

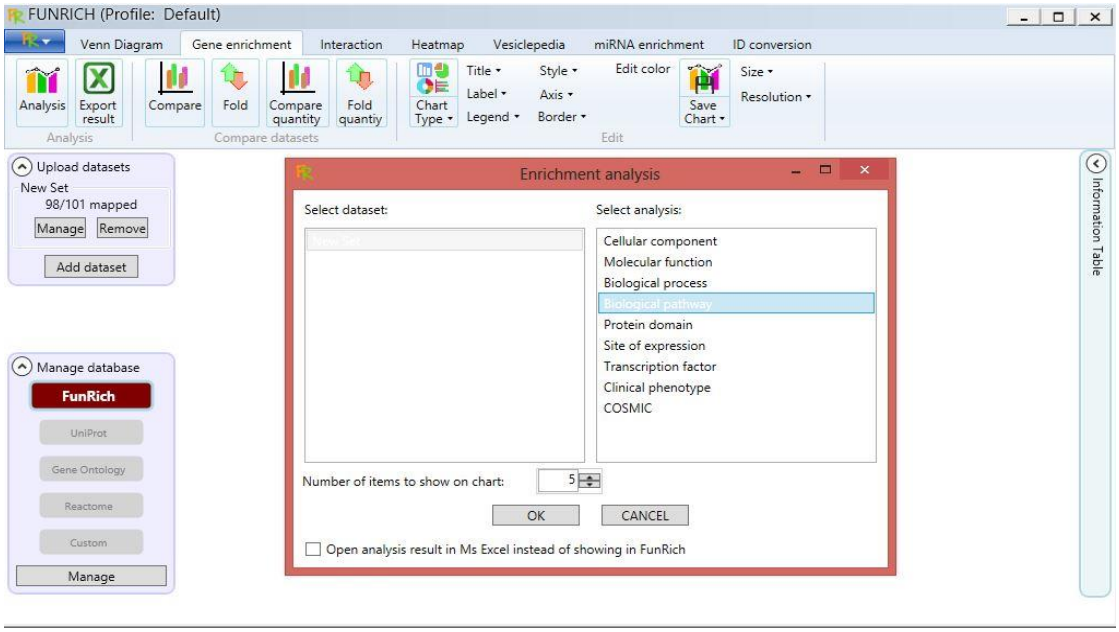
CFinder, **cytoHubba**, **ClusterMaker** are the apps used in the network analysis and identify Clique, Hub Genes and Cluster respectively.

3.3 FunRich

FunRich is an individual software tool used mainly for functional enrichment and interaction network analysis of genes and proteins. Besides, the results of the analysis can be depicted graphically in the form of Venn, Bar, Column, Pie and Doughnut charts. Currently, FunRich tool is designed to handle variety of gene/protein data sets irrespective of the organism. Users can not only search against default background database, but can also load customized database against which functional enrichment analysis can be carried out.

FunRich database option currently supports the enrichment analysis of the following categories:

• Biological process
• Cellular component
• Molecular function
• Protein domains
• Site of expression (normal tissues, cancer tissues, cell types and cell lines)
• Biological pathways
• Transcription factors
• Clinical synopsis phenotypic terms



3.4 JAVA Script

JavaScript is most commonly used as a client side scripting language. This means that JavaScript code is written into an HTML page. When a user requests an HTML page with JavaScript in it, the script is sent to the browser and it's up to the browser to do something with it. For the development of tool NutriGeneInteract we use JAVA Script as main language. The network of Gene Disease and Gene Nutrients form with the help of JAVA Script.

```
window.addEventListener("DOMContentLoaded", init);
function init() {
   document.getElementById("btn_ppi_click").addEventListener("click", main);
}

function main() {
   var checkedValue;
   var inputElements = document.getElementsByTagName('input');
   for (var i = 0; inputElements[i]; ++i) {
      if (inputElements[i].checked) {
         checkedValue = inputElements[i].value;
         console.log("Checked Value ", checkedValue);
         findNodes(checkedValue);
         findEdge(checkedValue);
      }
   }
   console.log("Node Data is ", nodes);
   if (nodes && nodes.length > 0) {
      localStorage.myjsonstore = JSON.stringify(nodes);
      location.href = "PPI-Network.html";
   }
}
var nodes = [];

function findNodes(id) {
```

```
    for (var i = 0; i < json.length; i++) {
        if (json[i].data.id == id || json[i].data.source == id) {
            nodes.push(json[i]);
        }
    }
    console.log("After Filter ", nodes);
}

function findEdge(id) {
    for (var j = 0; j < json.length; j++) {
        if (json[j].data.id == id || json[j].data.source == id) {
            nodes.push(json[j]);
        }
    }
    console.log("After Filter ", nodes);
}
```

3.5 jQuery

jQuery is very compact and well written JavaScript code that increases the productivity of the developer by enabling them to achieve critical UI functionality by writing very small amount of code. jQuery JavaScript file can be downloaded from http://jquery.com/. jQuery usually comes as a single JavaScript file containing everything comes out of the box with jQuery. It can be included within a web page using the following mark-up:

```
<script type="text/javascript" src="jQuery-1.4.1-min.js"></script>
```

3.6 JSON

JavaScript Object Notation or **JSON** is an open-standard file format that uses human-readable text to transmit data objects consisting of attribute–value pairs and array data types. It is a very common data format used for asynchronous browser/server communication, including as a replacement for XML in some AJAX-style systems.

"data": {
"id": 201100
, "idInt": 201100
, "name": "SLC6A4"
, "score": 0.001269877
, "query": true
, "gene": true
}
, "position": {
"x": 481.0169597
, "y": 384.8210888
}
, "group": "nodes"
, "removed": false
, "selected": false
, "selectable": true
, "locked": false
, "grabbed": false
, "grabbable": true
, "classes": "fn10473 fn944 fn941 fn1059 fn823 fn696 fn635 fn847 fn939 fn693 fn669 fn798 fn697 fn862 fn657 fn876 fn626 fn967 fn695 fn696 fn1024 fn10022 fn681 fn361 fn279 fn278 f569 fn741 fn85 fn43"
}

3.7 HTML and CSS

HTML is stands for Hyper Text Markup Language. Hyper Text means "Text within Text". A text has a link within it, is a hypertext. A markup language is a programming language that is used make text more interactive and dynamic. It can turn a text into images, tables, links etc. CSS stands for Cascading Style Sheets, CSS describes how HTML elements are to be displayed on screen, paper, or in other media. It saves a lot of work and can control the layout of multiple web pages all at once.

3.8 IDE Bracket

It is a java developing IDE used for the development of NutriGeneInteract tool. Bracket can be used for all the coding languages like JAVA, JavaScript, jQuery, HTML, CSS etc.

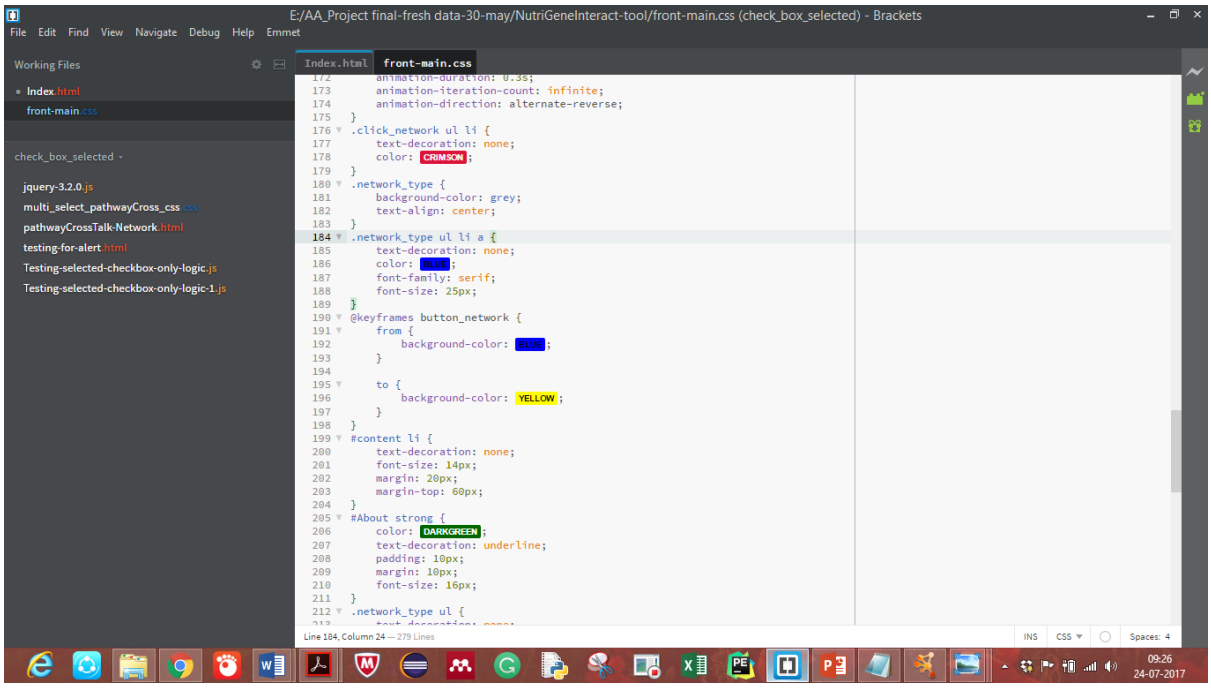


Figure: Bracket IDE

4. METHODOLOGY

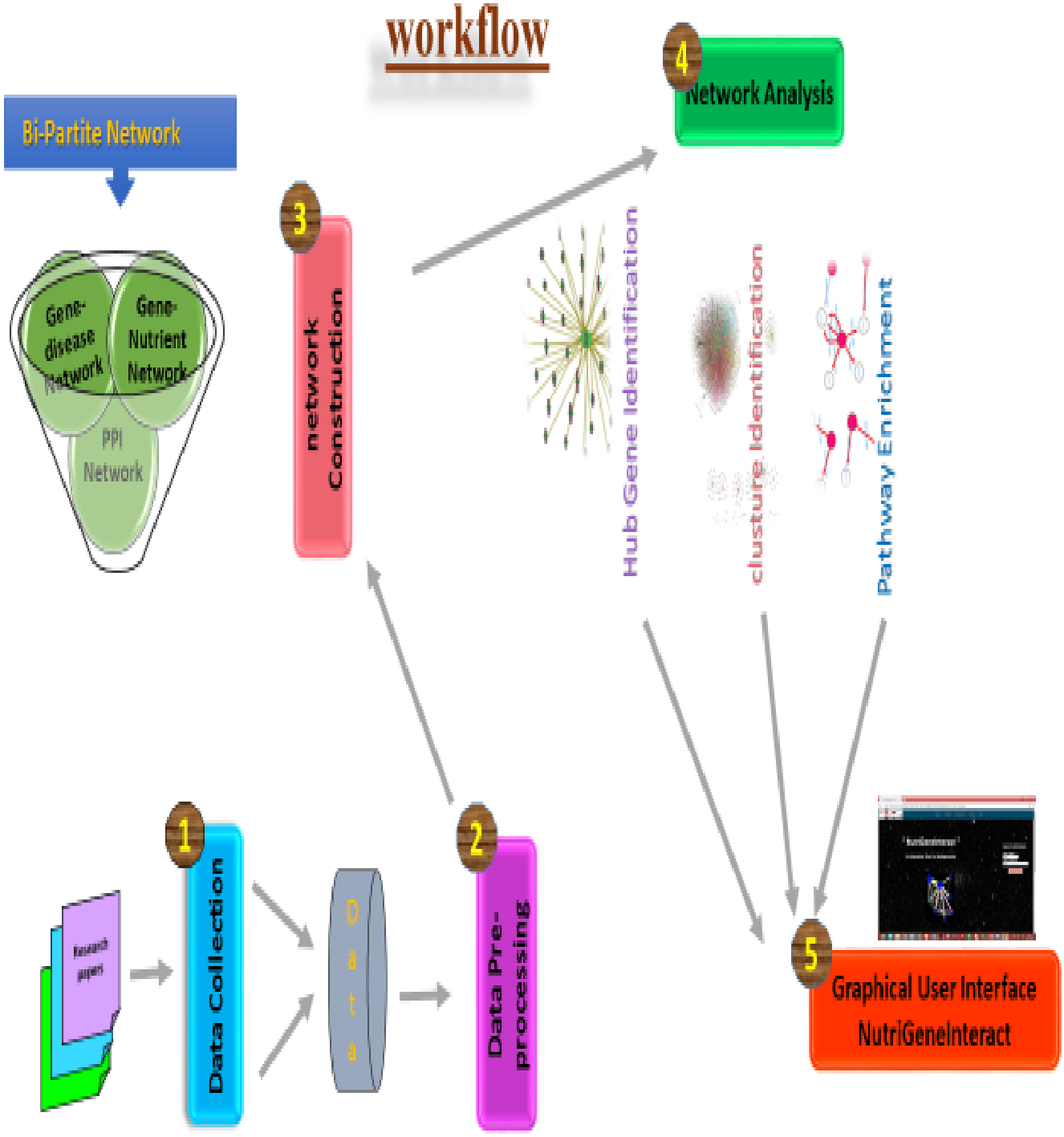


Figure.3: Workflow of the methodology

4.1 Data Collection and Pre-processing

4.1.1 Nutrient-Gene-Disease associations

The initial data for the study was obtained from literature as nutrigenomics is only decades old branch so there is not available database regarding Nutrigenomics. There are handful of databases related to nutrigenomics or Nutrigenetics but no one give the information of genes and metabolic nutrients or the connection of nutrients with the diseases. All the research paper which has work on the genes and diseases which caused due to imbalanced diet or nutritional elements are included in this study. There are so many diseases like metabolic syndrome, Obesity, Type 2 diabetes mellitus, Type 1 diabetes, cardiovascular disease, Weight gain, Glucose tolerance which caused due to mutation in gene and effect of nutrients. These genes, related diseases and affecting nutrients data has been collected form the literature. The data collected from the literature contains 171 genes related to 51 different diseases and affected with 91 nutrients.

4.1.2 Protein-Protein Interaction Network

To construction PPI network, we obtained 1124 protein-protein interactions (PPIs) relating to 825 proteins from the Human Protein Reference Database (HPRD) database. It is known to be one of the most reliable databases for PPI data. Many databases are available for obtaining protein-protein interaction data, such as Bio grid, MINT, STRING, MIPS and HPRD. Since Human Protein Reference Database is one of the most widely used database for research purposes and all the interaction data is experimental so we choose HPRD for fetching PPI data.

4.2 Construction and visualization of different networks

The Branch of Nutrigenomics is only decades old, earlier people were not aware of the science of nutrients and impact of nutrients on genetics. Unlike other fields there is less research and data is available. The Human Genome Project add great help to the nutrigenomics and nutrigenetics. It allows scientists to discover and study mutual relations between nutrition, genes, and diseases (El-Sohemy 2007).

4.3 Hub Genes Identification

Hubs were identified by two approaches. One of them was identifying hubs based on degree, betweeness centrality, bottleneck and maximum clique component (MCC) score. In this

approach, the top 10 hubs were identified which can be targeted by drugs and cure more than one Nutrients Related Disease's.

4.4 Cluster Identification

One common task in biological network analysis is to identify clusters (or modules) of biological molecules that share similar properties. For instance, a cluster of genes whose expression changes similarly to external stimuli may have related function and participate in the same biological processes. The clusterMaker App (Morris *et al.*, 2011) provides many frequently used network clustering algorithms.

4.5 Clique Identification

CFinder, a platform-independent, stand-alone application locating overlapping groups of densely interconnected nodes in graphs and illustrate its use on the network of gene associations. CFinder is efficient for networks with millions of nodes and, as a by-product of its search, the full clique overlap matrix of the network is determined.

4.6 Functional Enrichment Analysis

To explore the hypothesis that the associate genes characterize different pathways to Nutrient related diseases, we performed Functional enrichment analysis with the use of FunRich tool.

4.7 NutriGeneInteract: An Interactive Tool for Nutrigenomics

At the end we integrated all the information obtained from the analysis of different network studies of Nutrigenomics and made that available as a web interface.

5. RESULTS & DISCUSSION

5.1 Disease gene Bi-partite network

Disease-Gene bipartite network was constructed such that the circular nodes represented Nutrition related diseases and the arrow shape nodes represented genes implicated in diseases. Figure represents disease-Gene network. As it can be clearly observed that many genes are only connected to single diseases. However, there are few genes which connect more than one disease and hence can be concluded that they have pleiotropic effects. And multiple genes associated with single disease exhibit epistatic effect on the development of that disease.

High number of mutations is observed to be linked with obesity. Type 2 diabetes, cardiovascular disease and metabolic syndrome disease share many variants, which reflect some biological relationship leading to the co-manifestation of these diseases.

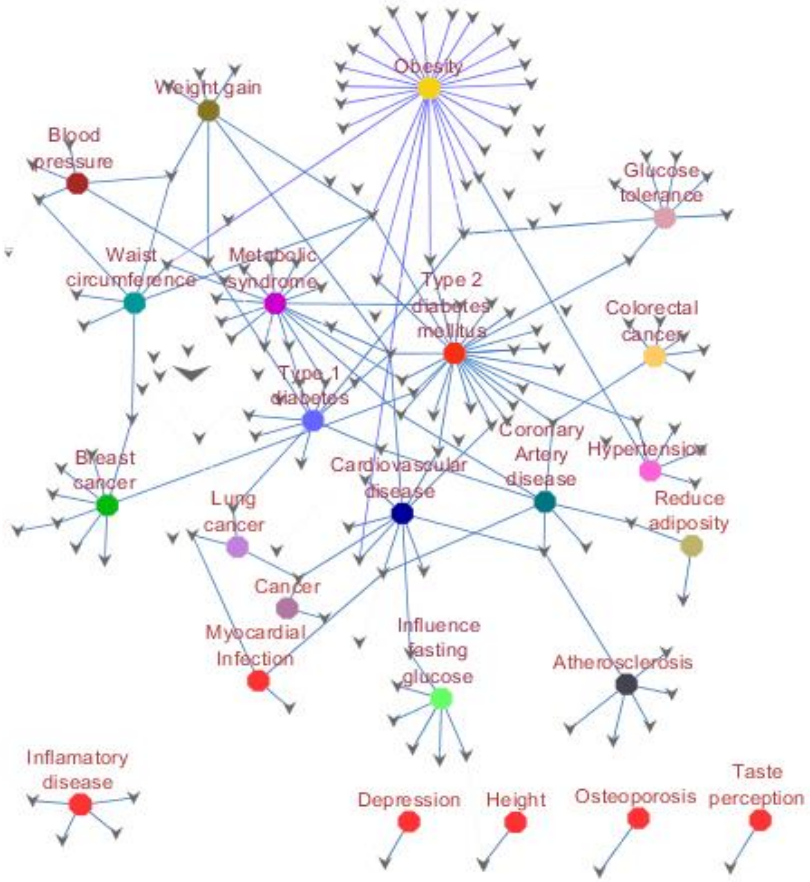


Figure.4: *Disease-SNP bipartite network (ARD- circles; Genes – triangle).*

5.2 **Nutrient-Gene network**

Nutrients effect on genetics is not very well studied. Many nutrients effect genes and their functionality. It is showing that maximum genes which has taken for the study are connected to mainly insulin, glucose, cholesterol and triglycerides.

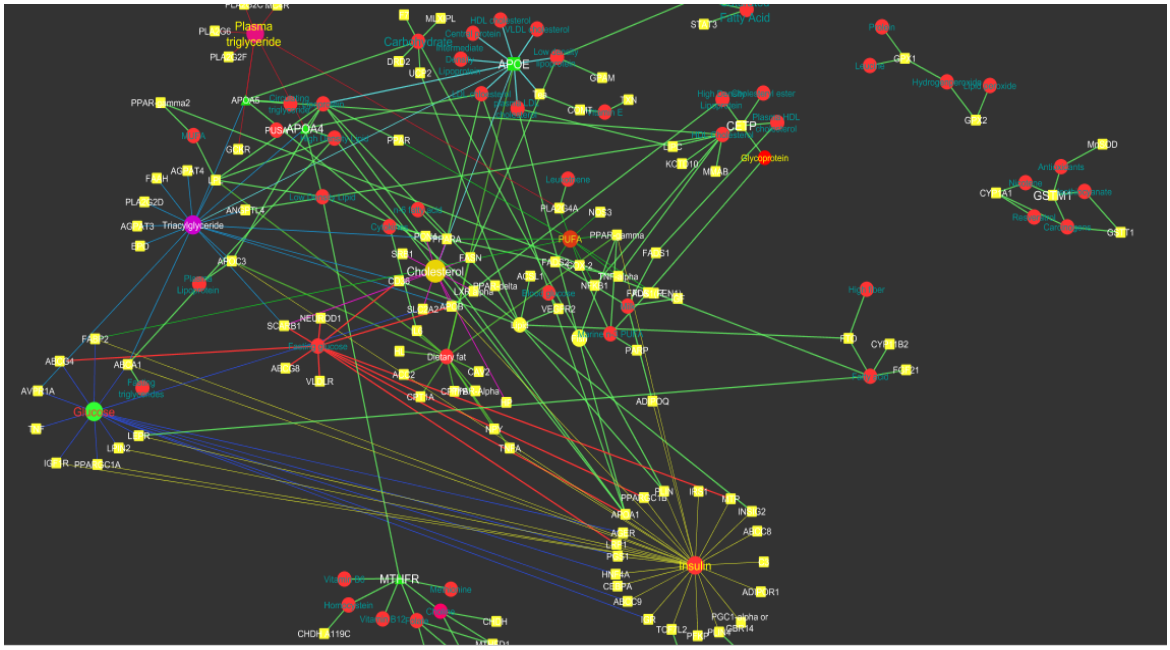


Figure.5: Nutrient-gene network (circle represent Nutrients, square represent genes)

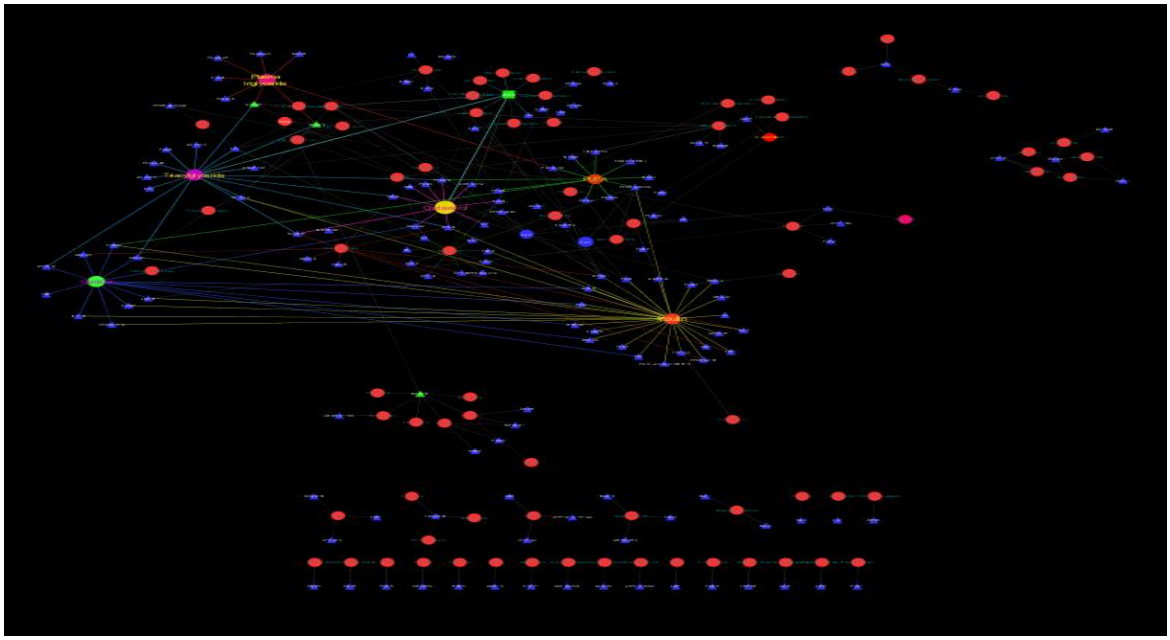


Figure.6: Full view of Nutrient-Gene network

5.3 Nutrient related PPI network

Protein-protein interaction network visualise that genes STAT3, LRP1, PPARA, IRS1, NFKB1, IGF1R shows maximum number of interaction with other genes.

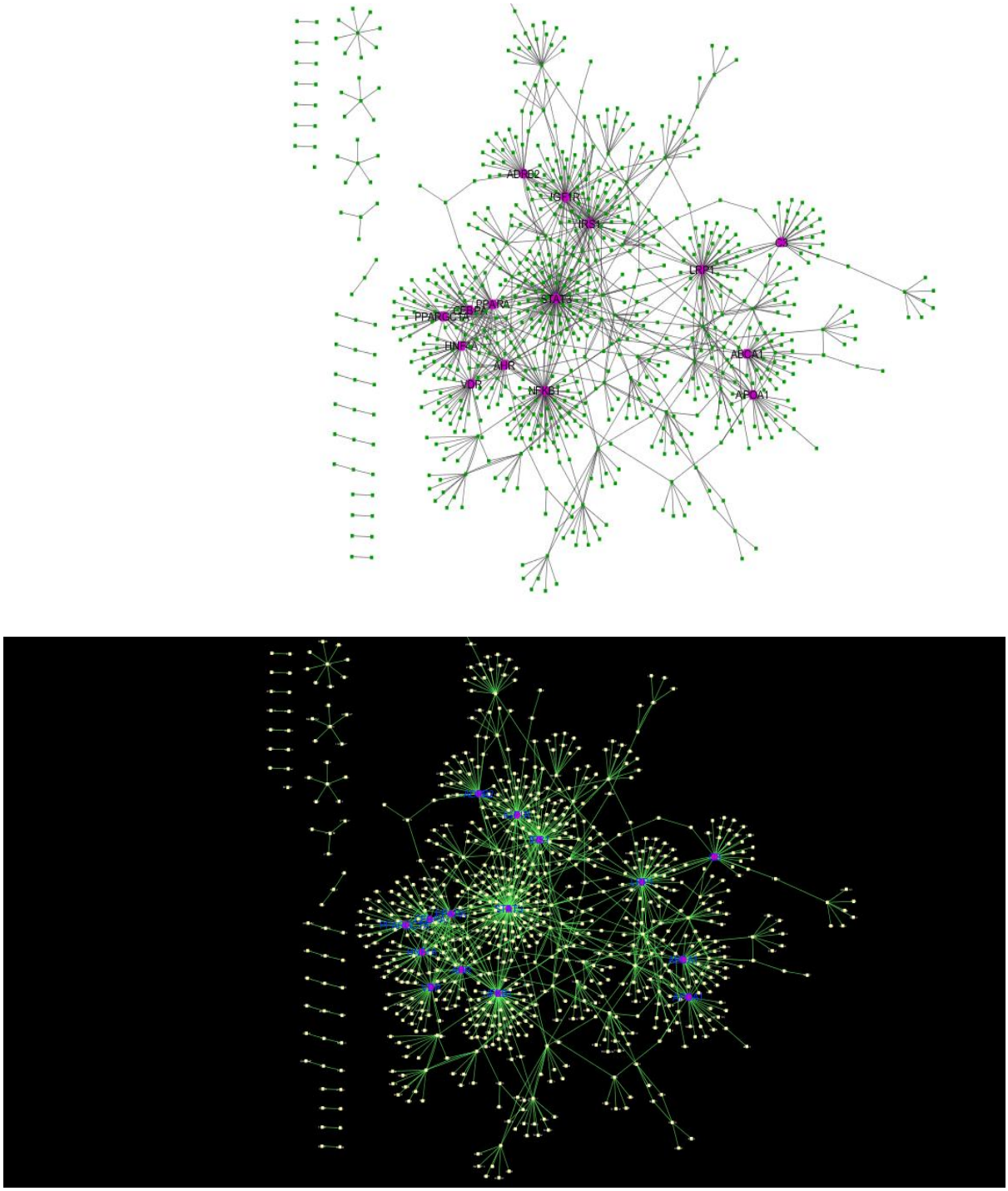
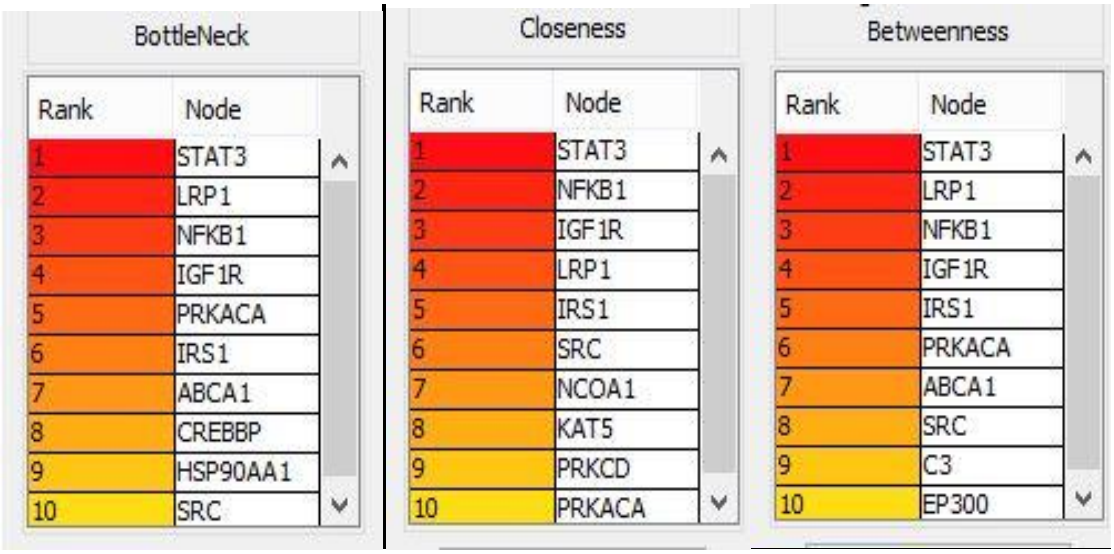


Figure.7: PPI networks of Nutrient related genes

5.4 Hub Gene Analysis of Nutrition related diseases



BottleNeck

Rank	Node
1	STAT3
2	LRP1
3	NFKB1
4	IGF1R
5	PRKACA
6	IRS1
7	ABCA1
8	CREBBP
9	HSP90AA1
10	SRC

Closeness

Rank	Node
1	STAT3
2	NFKB1
3	IGF1R
4	LRP1
5	IRS1
6	SRC
7	NCOA1
8	KAT5
9	PRKCD
10	PRKACA

Betweenness

Rank	Node
1	STAT3
2	LRP1
3	NFKB1
4	IGF1R
5	IRS1
6	PRKACA
7	ABCA1
8	SRC
9	C3
10	EP300

Figure 8: Hub gene analysis of nutritional diseases

5.5 Finding clique

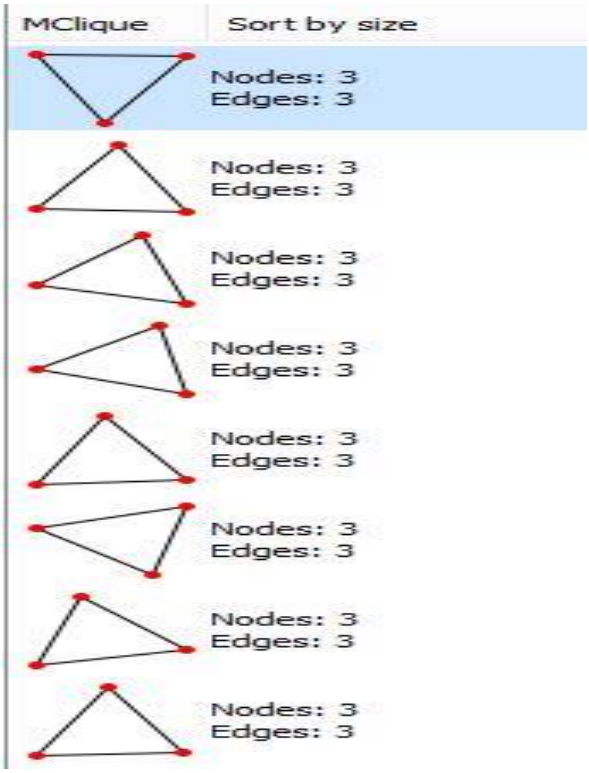
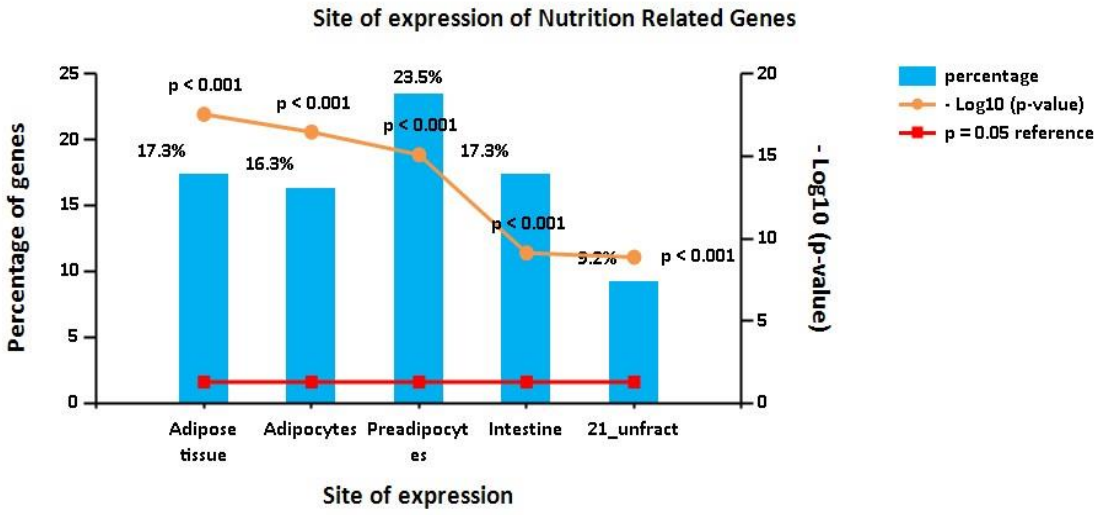
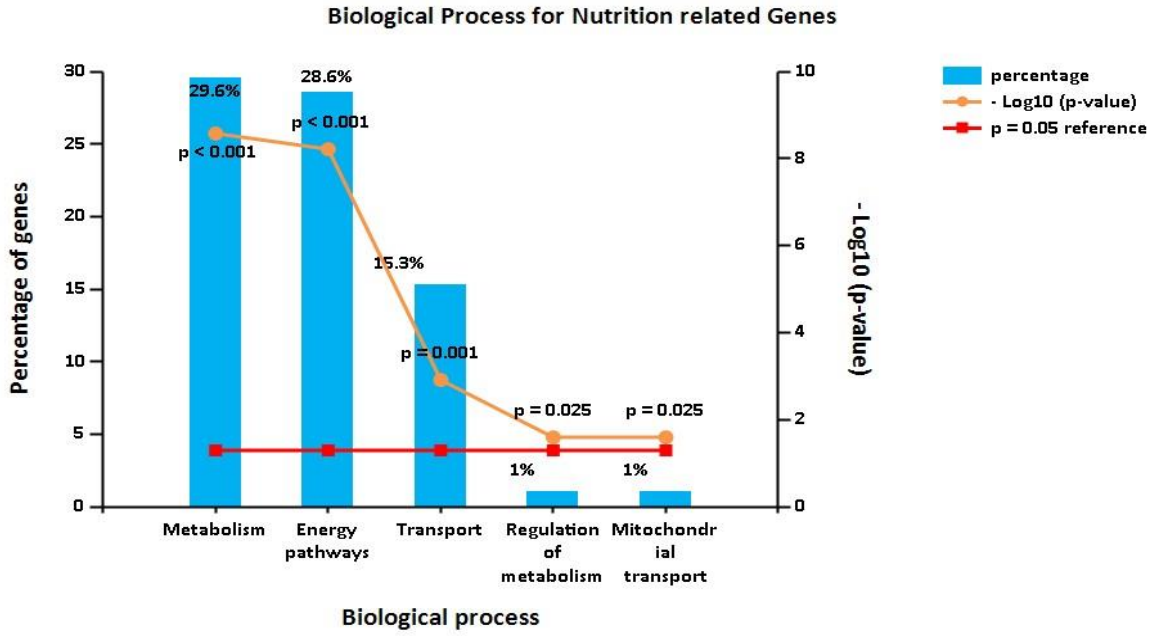


Figure 9: Depiction of nodes & edges

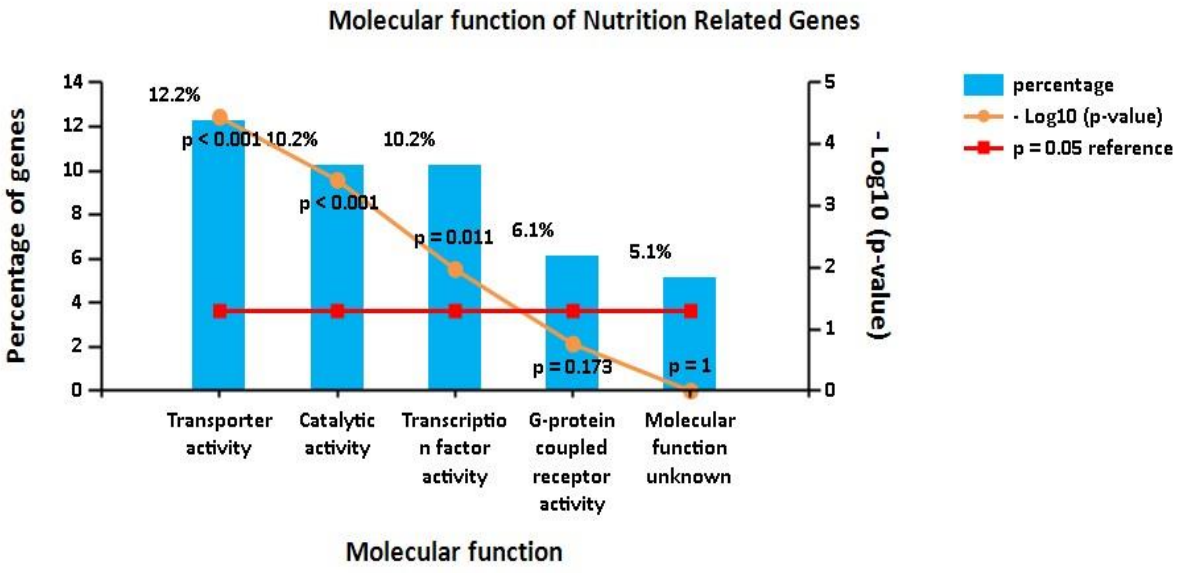
5.6 Functional enrichment analysis



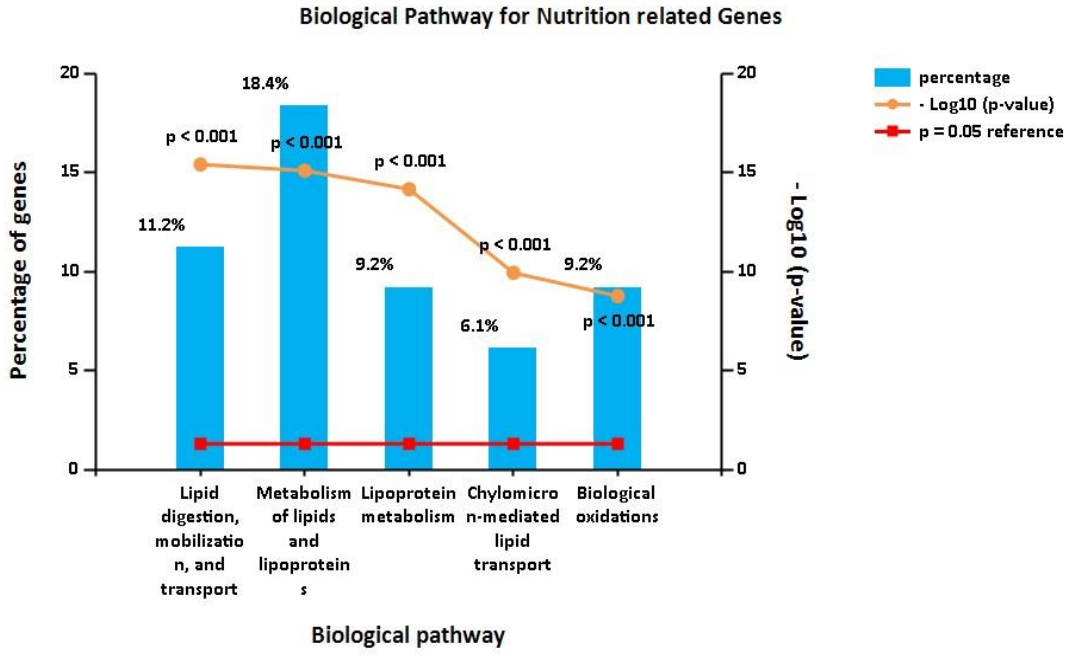
17.3 % genes are expressed in Adipose Tissue 23.5 % in Preadipocytes, another 17.3% in Intestine and 16.3% are expressed in Adipocytes



29.6% genes are responsible for metabolism, 28.6 % for energy pathways, and 15.3% for Transport



12.2% genes are responsible for Transporter activity, 10.2% for Catalytic activity,another 10.2% for Transcription Factor activity



18.4% genes are responsible for Metabolism of lipids and lipoproteins, 11.2% for lipid digestion, mobilization and transport, 9.2% for lipoprotein metabolism and another 9.2% for Biological oxidation

Figure.10: FunRich results

5.7 NutriGeneInteract:

NutriGeneInteract (http://genomeinformatics.dtu.ac.in/NutriGeneInteract/Index.html) have a front page with different informative tabs like About the tool, links about related database and websites.

Search panel: there is a search panel which shows the option to search for Genes and related Nutrients or diseases.

Display Panel: it shows the network of selected genes with the nutrients or disease.

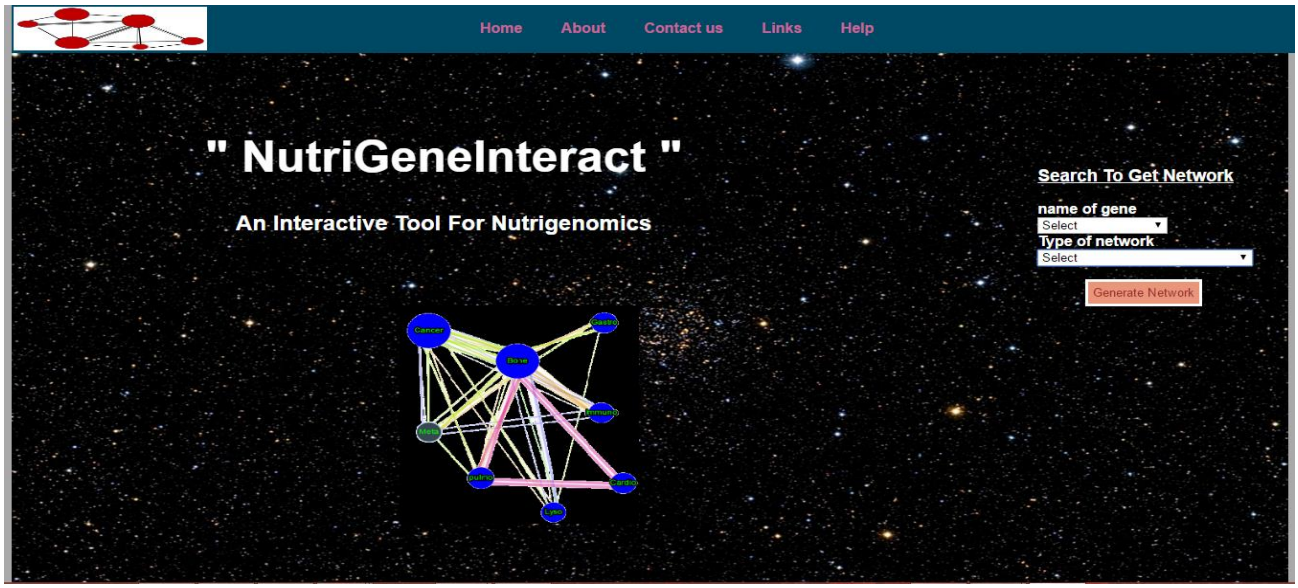


Figure 11: Front page of NutriGeneInteract Tool

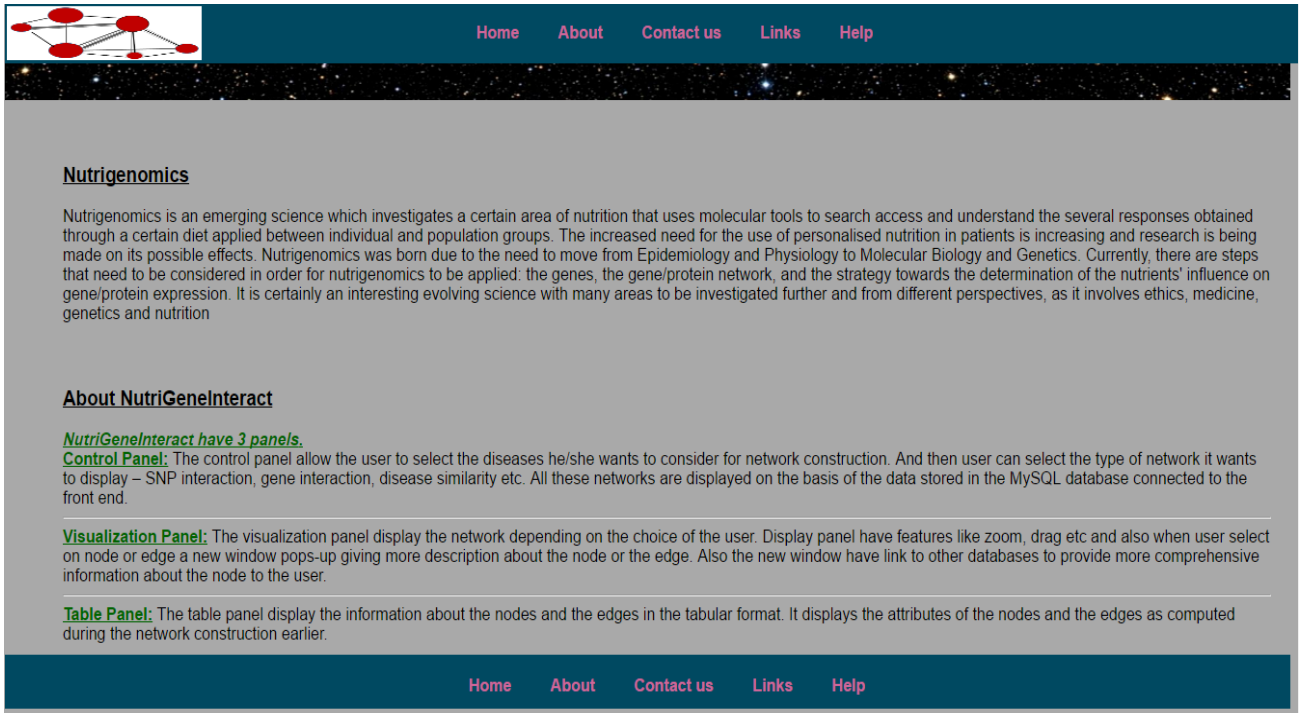


Figure 12: Introduction to NutriGeneInteract Tool

6. CONCLUSION

The Nutritional Genomics focuses on the interaction between bioactive food components and the genome, i.e Nutrigenetics and Nutrigenomics. The influence of nutrients on the genes expression is called Nutrigenomics, while the heterogeneous response of gene variants to nutrients, dietary components is called Nutrigenetics. Genetic variation is known to affect food tolerances among human subpopulations and may also influence dietary requirements and raising the possibility of individualizing nutritional intake for optimal health and disease prevention on the basis of an individual's genome. Nutrigenomics provides a genetic understanding for how common dietary components affect the balance between health and disease by altering the expression and structure of an individual's genetic makeup. Nutrigenetics describes that the genetic profile have impact on the response of body to bioactive food components by influencing their absorption, metabolism, and site of action.

NutriGeneInteract available as an online platform which allow to visualize and analyse the Nutrition related genes –disease and nutrient-gene networks in a user-friendly environment. NutriGeneInteract also allow the construction of multi-partite network for analysis and visualization to uncover genetic similarities among various ARDs and also direct the users to connect to other relevant databases/websites to retrieve more meaningful information.

In summary, by exploring the genes and genetic variations associated with Nutrient related diseases in context of network and pathways, we can unravel the unknown molecular mechanisms associated with Nutrition and can identify the novel genes and variants which are still unknown. The study will provide links between various Nutrition-related disorders, which can provide valuable perspective to physicians, dietitian and biomedical researchers. Moreover, the integration of this study with gene/miRNA expression profiles could further highlight the key players linking various Nutrition related diseases. Our multidimensional approach using network and pathway based analysis can be applicable to other diseases as well. Also, by exploiting the power of Bayesian networks, more statistically powerful enrichment of network can be done which is missing from the proposed plan but may be done in the future.

Nutrigenomics is a new field of research so have plenty of things to explore and discover.

7. REFERENCES

- El-Sohemy A. Nutrigenetics. Forum Nutr 2007;60:25–30.
- Ferguson LR. Nutrigenomics: integrating genomic approaches into nutrition research. Mol. Diagn. Ther. 2006;10:101–108.
- Ghosh D, Tapsell L (2008) Nutrigenomics: a genomic approach to human nutrition. In: Genomics: principles, applications and regulatory perspectives. CRC Press, USA, pp 337–344
- Joost HG, Gibney MJ, Cashman KD, Görman U, Hesketh JE, Mueller M, Ommen BV, Williams CM, Mathers JC. Personalised nutrition: status and perspectives. Br J Nutr. 2007;98:26–31. doi: 10.1017/S0007114507685195.
- J.M. Ordovas, V. Mooser, Nutrigenomics and nutrigenetics, Curr. Opin. Lipidol. 15 (2004) 101–108.
- Kogelman et. al. (2014). Identification of co-expression gene networks, regulatory genes and pathways for obesity based on adipose tissue RNA Sequencing in a porcine model, BMC Medical Genomics, 4, 7:57
- Kogelman, LJA and Kadarmideen, HN. (2014). Weighted Interaction SNP Hub (WISH) network method for building genetic networks for complex diseases using whole genome genotype data. BMC Systems Biology 8
- Koutsos A, Manaia A, Willingale-Theune J. Introdução aos Microarrays de ADN.Cambridge, UK: European Molecular Biology Laboratory (EMBL); 2009–2011. (Versão 2.3).
- M. Muller, S. Kersten, Nutrigenomics: goals and strategies, Nat. Rev., Genet. 4 (2003) 315–322.
- Morris JH, et al. clusterMaker: a multi-algorithm clustering plugin for Cytoscape. BMC Bioinformatics. 2011;12:436.
- Peregrin, T., 2001. The new frontier of nutrition science: nutrigenomics. J. Am. Diet. Assoc. 101 (11), 1306.
- Palou, A., 2007. From nutrigenomics to personalised nutrition. Gen. Nutr. 2, 5–7.

- Neeha, V.S., Priyamvadah, Kinth P., 2013. Nutrigenomics research: a review. J. Food Sci. Technol. 50 (3), 415–428.
- Sales, N.M.R., Pelegrini, P.B., Goerch, M.C., 2014. Nutrigenomics: definitions and advances of this new science. J. Nutr. Metab. 202759.
- Ovesná J., Slabý O., Toussaint O., Kodícek M., Marsík P., Pouchová V., Vaněk T. High throughput 'omics' approaches to assess the effects of phytochemicals in human health studies. Br. J. Nutr. 2008;99 E(Suppl. 1):ES127–ES134.
- Schlicker A., Albrecht M. (2010) FunSimMat update: new features for exploring functional similarity. Nucleic Acids Research, 38
- Paliyath G., Bakovic M., Shetty K. Functional Foods, Nutraceuticals, and, Degenerative Disease Prevention. UK: Wiley Blackwell; 2011.
- Van Ommen, B., Stierum, R., 2002. Nutrigenomics: exploiting systems biology in the nutrition and health arena. Curr. Opin. Biotechnol. 13 (5), 517–521.
- World Health Organization . Diet, nutrition and the prevention of chronic diseases. Report of a WHO Study Group. (WHO Technical Report Series, No. 797), Geneva.1990.