DEVELOPMENT OF A SYSTEMS BASED MODULAR APPROACH TO LINK COMPLEX DISEASES

by

Deniz Rende

B.S., Chemical Engineering, Boğaziçi University, 2001

M.S., Chemical Engineering, Boğaziçi University, 2004

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

Graduate Program in Chemical Engineering

Boğaziçi University

2010

ACKNOWLEDGEMENTS

I would like to express my sincere gratitute to my dissertation advisor, Prof. Betül Kırdar, for her everlasting support and encouragement during this study and my entire high education years. As my mentor, Prof. Kırdar will always be present in my future work. I sincerely thank Assoc. Prof. Ebru Toksoy Öner for her guidance following my senior years. I would like to acknowledge the members of the dissertation committee, Prof. Z. İlsen Önsan, Assist. Prof. Elif Özkırımlı Ölmez and Assist. Prof. K. Yalçın Arga, for their time and productive comments on this work. This research was financed by Boğaziçi University Research Fund, project 09HA502D and State Planning Organization; project DPT-09K120520. Boğaziçi University Foundation kindly financed this work in several conferences.

I would like to thank Prof. Fahir Borak for supporting my work and for permitting much needed time to complete my research. Part of the conference costs for this study to be presented to a broader audience was kindly provided by Yeditepe University. I would like to thank especially Assist. Prof. Seyda Bucak for contributing my research life with her ambition and motivation. I would like to express my kind regards to the members of Department of Chemical Engineering, Yeditepe University for their understanding throughout my PhD studies. At the onset of my career I am indebted to the members of the department for the professional experience on the value of detachment and priorities.

I feel lucky to be a member of the Biosystems Engineering Research Group at Bogazici University. Past and present members, Pınar, Emrah, Hilal, Çağrı, Tunahan, Yalçın, Duygu, Ayça, Esra and Elif provided friendly and supporting environment and fruitful discussions in the last eight years. I will miss being a part of this community. I am also very happy to participate in the studies in Computational nanoMaterials and bioNetworks Research Group at Yeditepe University. I sincerely thank to Assist. Prof. Nihat Baysal, for his encouragement and support. During our ever-lasting discussions, I have learned a lot, particularly how to be calm and patient; without his help and support, I wouldn’t complete this work. My group friends, Semra, Merve, Binnaz and Leyla, provided an exciting research environment and brought joy to my life.

I am deeply indebted to my parents, Ayşen and Fikri Rende, for their never-ending support from the very beginning. Mom, this one is the last; there are no more schools, your belief in the value of education and your continuous encouragement brought me so far. Dad, without your support by all means, I wouldn't even dream to complete this work. My sister, Sevinç, my support and joy guru; you are the most hard-working, creative and enthusiastic person I have ever met: "I smile because you are my sister; I laugh because you have nothing to do about it". Our cats, Fıstık, Mırmır and Boncuk, always brought joy to my life; I hope we will have more time together. Last, but not least, I bless every moment I shared with my boy, Pisi. He passed the rainbow bridge on February 4th, 2010. My grief for him is eternal.

ABSTRACT

DEVELOPMENT OF A SYSTEMS BASED MODULAR APPROACH TO LINK COMPLEX DISEASES

A novel systems based modular approach was developed in this study to investigate the shared underlying mechanisms of three prevalent complex disorders: cardiovascular disease (CVD), Type 2 diabetes (T2D) and Alzheimer's disease (AD). The disease related functional linkage networks were constructed and analyzed in every aspects of modularization. The fundamental biological processes were deciphered through integration of various levels of "omics" data. The identification of the links between these diseases was accomplished through the shared proteins and biological processes.

The cardiovascular disease functional linkage network (CFN) containing 1536 proteins and 3345 linkages was constructed using proteins encoded by 234 genes associated with cardiovascular disease. The modular architecture of this network and its integration with "bibliomics" indicated that blood coagulation, lipid metabolism and renin-angiotensin systems were found to be linked to CVD as expected. The modularization in CFN was anticipated to exert a potential to develop a general perspective, hence the functional modules enumerated from CFN were evaluated in terms of shared pathways, co-localization, co-expression and associations with diseases. The members of the top scoring functional modules were assembled in a condensed network representing the fundamental biological processes. The approach was later applied to the other two complex disorders. The functional linkage networks include 2734 proteins / 14823 linkages (TDFN) and 1587 proteins / 7785 linkages (ADFN) for Type 2 diabetes and Alzheimer's disease, respectively. The individual fundamental processes and signalling cascades observed in each of these diseases are converged at blood coagulation, glucose homeostasis, lipid metabolism, oxidative stress and inflammation. The central roles of the proteins generating inflammatory response in all disease networks indicate that inflammation might be the missing link between CVD, T2D and AD.

ÖZET

KOMPLEKS HASTALIKLARI BİRBİRİNE BAĞLAMAK İÇİN SİSTEM BAZLI MODÜLER BİR YAKLAŞIM GELİŞTİRİLMESİ

Bu çalışmada ilk defa sistem bazlı modüler bir yaklaşım geliştirilerek üç yaygın kompleks hastalığın, kalp hastalıkları (CVD), Tip 2 diyabet (T2D) ve Alzheimer hastalığı (AD), altında yatan ortak biyolojik mekanizmalar incelenmiştir. Bu hastalıklarla ilgili işlevsel etkileşim ağları kurulmuş ve bu ağlardaki modüler yapılar her açıdan analiz edilmiştir. Değişik düzeylerdeki "omiks" verileri birleştirilerek temel biyolojik süreçler ortaya çıkarılmıştır. Üç hastalıkta da paylaşılan proteinler ve biyolojik süreçler ile hastalıkların arasındaki ilişki tanımlanmıştır.

Kalp hastalıklarının işlevsel etkileşim ağı (CFN), hastalıkla ilişkili olduğu gösterilen 234 genin şifrelediği proteinlerden başlayarak 1536 protein ve 3345 etkileşim ile kurulmuştur. Bu ağdaki modüler yapılar ve bu yapıların "bibliomik" bilgisi ile birleştirilmesi sonucunda kan pıhtılaşması, yağ metabolizması ve renin-anjiyotensin sistemi gibi biyolojik süreçlerin CVD ile beklenen ilişkisi gösterilmiştir. CFN'de bulunan modüler yapılar genel bir yaklaşım oluşturmak bakımından bir potansiyel olarak görülmüş, dolayısıyla CFN'de bulunan modüler yapılar ortak biyolojik yolağları, lokalizasyon, ekspresyon ve diğer hastalıklarla ilişkiler açısından değerlendirilmiştir. Yüksek skorlu modüller içinde yer alan proteinlerin daha yoğun bir haritada toplanması ile hastalığın temelinde yatan biyolojik süreçler gösterilmiştir. Bu yaklaşım diğer iki kompleks hastalık için de uygulanmıştır. İşlevsel etkileşim ağları Tip 2 diyabet ve Alzheimer hastalığı için sırasıyla 2734 protein / 14823 etkileşim (TDFN) ve 1587 protein / 7785 etkileşim (ADFN) ile oluşturulmuştur. Burada gözlemlenen temel biyolojik süreçler ve sinyal ileti mekanizmaları, kan pıhtılaşması, glukoz dengesi, yağ metabolizması, oksidatif stres ve enflammasyon süreçlerinde ortaklık göstermektedir. Enflammasyon yaratan proteinlerin üç hastalık ağında da merkezi roller üstlenmesi, enflammasyonun CVD, T2D ve AD arasındaki ortak bağlantı olabileceğine işaret etmektedir.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS / ABBREVIATIONS

b_m	Betweenness for node *m*
$b(k)$	Betweenness distribution with respect to degree *k*
$<C>$	Average clustering coefficient for a network
C_i	Clustering coefficient for node *i*
$C(k)$	Clustering coefficient distribution with respect to degree *k*
d_i	Number of proteins associated with a disease
$DO_{i,j}$	Disease overlapping score for diseases *i* and *j*
f^k	Relative frequency of the classification in the cluster
F_i	Genetic Algorithm model
$<k>$	Average degree for a network
k	Degree
l	Edges
n	Total number of classes in the classification scheme
$n(k)$	Cumulative degree distribution with respect to degree *k*
N	Nodes
$P(k)$	Probability of finding *k* links
R_i	Functionlal module score for classification *i*
S	General term for scoring
Q	Modularity
α	Model coefficient
β	Model coefficient
γ	Degree exponent in scale-free networks
γ_c	Cumulative degree exponent in scale-free networks
$\Gamma(i,j)$	Shortest paths between nodes *i* and *j*
$\Gamma(i,m,j)$	Shortest paths between nodes *i* and *j* passing through *m*
μ	Average of the arrays
σ	Standard deviation

Aβ	Amyloid-beta
AD	Alzheimer's disease
ADFN	Alzheimer's Disease Functional Linkage Network
CFN	Cardiovascular Disease Functional Linkage Network
CVD	Cardiovascular disease
GO	Gene Ontology
MeSH	Medical Subject Headings
OMIM	Online Mendelian Inheritance of Men
PCC	Pearson correlation coefficient
T2D	Type 2 diabetes
TDFN	Type 2 Diabetes Functional Linkage Network

1. INTRODUCTION

A novel systems based modular approach was developed in this study to have a general understanding to the complex disease interference. Systems biology approaches to diseases arise from a simple hypothesis that genes contributing to a common disorder have an increased tendency for their products to be linked at various levels of functionality, including protein-protein interaction, co-expression, co-regulation and share Gene Ontology terms [1]. Complex diseases have long been known to emerge from an impaired function of a single protein or a protein cluster that alter the general functionality. Statistically significant pathogenic overlap between the complex disorders results possibly from the variations in linked genes encoding proteins that are a part of a functional module. Hence, systems based approaches have found a wide range of applications for the identification of the putative proteins and the understanding of the underlying biological processes. The identification of disease-causing genes not only facilitates the understanding of the protein function that provides direct insight into the progression of the disease but also points out potential drug targets for further research.

The modular approach developed in this study was used to investigate three prevalent diseases: cardiovascular disease, Type 2 diabetes and Alzheimer's disease. Cardiovascular disorders (CVD) consist of a group of heterogeneous disorders affecting vascular system, heart and brain and are among the main causes of death and disability worldwide. Type 2 diabetes is the most common form of diabetes, accounting for over 90 per cent of the diabetes cases. It is characterized by a combination of impaired insulin secretion and action. Alzheimer's disease, which is one of the prevalent chronologic diseases in developed countries, is a neurodegenerative disorder characterized by a decline in cognitive abilities and memory loss due to deposits of the Aβ peptide and protein tau tangles. In the last decades, significant efforts have been expanded to understand the contribution of genetic factors to the development of these diseases with the hope that discovering these genetic factors will provide fundamental insights for pathogenesis, diagnosis and treatment [2, 3]. Most of these studies revealed the importance of underlying biological pathways and shared genes among the disease under investigation. Thereby, systems biology based approaches have recently found wide applications for the

identification of the molecular mechanisms underlying complex disorders and their relationships with other complex disorders including cancer [4], asthma [5], aging [6, 7].

Protein-protein interactions are essential for all biological processes. Although, large scale protein interaction maps have been generated for several model organisms, only a few interaction datasets are available for mammals. Human protein interactions have been used to identify potential disease causing proteins [8-10]. Functional linkage networks are also relevant from a systems biology point of view, the general organization principles can be conveyed using these networks. Protein-protein interactions from high throughput experiments are reported and deposited in publicly available databases. However, functional relatedness can be achieved at any level of interaction; including physical interaction as well as co-expression, co-regulation and phenotypic behavior. Functionally related genes usually act in the form of modules of highly interacting proteins encoded by these genes. These modules are considered as building blocks of biological systems and their interactions may shed light into the complex function of the whole system.

The idea of community structure in networks has been applied in various research fields including social communities [11, 12], the internet [13] and ecosystems [14]. Modularization in yeast protein interaction networks received much attention for gene annotation, protein function prediction, identification of regulators and novel proteins in molecular pathways [15-17]. There are numerous algorithms proposed to identify dense subgraphs and functional modules [18-21]. The Bron-Kerbosch algorithm [22] is a rigorous clique partitioning algorithm that aims to enumerate maximal cliques within a network. The implementation is easy compared to some other clique enumeration algorithms [23]. The algorithm has been applied to various networks ranging from social networks to large scale proteomic networks to find overlapping cliques [24-27].

In this study, an integrative modular network approach, where genes were organized into functional modules based on the topological characteristics of the constructed network, was developed to investigate three complex disorders, cardiovascular diseases, Type 2 diabetes and Alzheimer's disease for the identification of shared genes and statistically significant co-occurrence with other disorders. The proposed approach was initiated from the modular architecture of the cardiovascular disease related functional

linkage network. Initial results gathered through incorporation of this network with the semantic information available in biomedical literature indicated that the modular structures present in this network are potentially useful to delineate the association of the disease with other complex disorders. For a better understanding of the modular structure of the cardiovascular disease functional linkage network, the proteins in the functional modules were scored and evaluated in terms of shared pathways, co-localization, co-expression and associations with similar diseases. The assembly of top scoring overlapping members in the functional modules revealed the fundamental biological processes present in the pathophysiology of the disease. Other complex diseases that have pronounced associations with the proteins included in this assembly were linked to each other through shared proteins. The significant associations of the complex diseases supported the strong coherence of cardiovascular diseases with a number of metabolic, neurological and muscoskeletal diseases and several cancers arising from epithelial cells. This proposed methodology was also applied to other disease related functional linkage networks: Type 2 diabetes and Alzheimer's disease. The results indicated the major biological processes involved in these diseases, as well as persuasive associations with other complex disorders. This study provides further evidence that systems based modular approaches are powerful tools to develop a comprehensive understanding for human diseases.

2. BACKGROUND INFORMATION

2.1. Cardiovascular Disease as a Model System

Cardiovascular disorders (CVD) consist of a group of heterogeneous disorders affecting vascular system, heart and brain and are among the main causes of death and disability worldwide. Although environmental factors and lifestyle, such as smoking, diet and physical activity [28-30] play important roles in the prognosis of these disorders; age [31], sex [32] and underlying medical conditions, such as obesity, diabetes, hypertension and hypercholesterolemia are also considered to be important risk factors [33-36]. In recent years, there has been considerable effort to understand the contribution of genetic factors to the development of the disease with the hope that discovering these genetic factors will provide fundamental insight for pathogenesis, diagnosis and treatment of cardiovascular diseases [2, 3]. Most of these studies revealed the importance of several genes related to the homeostatic balance of coagulation-fibrinolysis, renin-angiotensin systems and lipid metabolism resulting in a detailed list of large number of genes that are significantly associated or non-associated with CVD. Association of CVD with other complex disorders such as Alzheimer's disease [37], obesity, diabetes [38] as well as cancer and inflammation [39] was also the subject of several recent reports. Computational biology approaches provide the fundamental insights and development of biomarkers for cardiovascular disease [40]. The lipoprotein metabolism, coagulation cascade, renin-angiotensin mechanism, oxidative stress and inflammatory processes are the fundamental cellular processes that are involved in the progression of the disease.

2.1.1. Lipoprotein Metabolism in CVD

One of the prevalent forms of cardiovascular diseases is atherosclerosis, which is characterized by the accumulation of lipid particles and cells of the immune system in subendothelial regions, leading to narrowing of the artery and, following plaque rupture, to thrombosis. The initiation of the process involves the accumulation of low density lipoproteins (LDLs) which are modified by oxidation or enzymatic activity and form agglomerates, leading to an increase phagocytosis by macrophages. A key early step in its

development is the accumulation and subsequent oxidation of LDLs within the arterial intima. *In vitro* studies have shown that oxidized LDL (oxLDL) promotes leukocyte recruitment and activation, lipid accumulation and cell death. Unregulated uptake of atherogenic lipoproteins by macrophages resulted in the generation of foam cells (Figure 2.1.a). The accumulation of foam cells leads to the formation of fatty streaks, which are widely considered to be the initial indication leading to the development of complex atherosclerotic lesions (Figure 2.1.b). Later, vascular smooth muscle cells contribute to this process by secreting large amounts of extracellular-matrix components, such as collagen. The presence of collagen increases the retention and aggregation of atherogenic lipoproteins. Monocytes and leukocytes are recruited that propagate chronic inflammation. As the plaque grows, while the size of the passage remains constant, the artery enlarges (Figure 2.1.c). Upon the death of foam cells, crystalline cholesterol is accumulated. In addition, smooth muscle cells form a fibrous cap covering the agglomerate preventing the plaque removal by blood. This process contributes to the formation of a necrotic core within the plaque and further promotes the recruitment of inflammatory cells (Figure 2.1.d). When the thrombus is large enough, it may block the artery causing acute coronary syndrome or myocardial infarction [41]. The plaque formation in the artery involves many components of the immune system: macrophages that develop into foam cells, T-cells and cytokines that are secreted by cells within atherosclerotic plaques, including interleukins (IL1, IL2, IL6, IL8, IL10, IL12), tumor-necrosis factor (TNF), interferon-γ (IFNG) and platelet-derived growth factor (PDGF) [42].

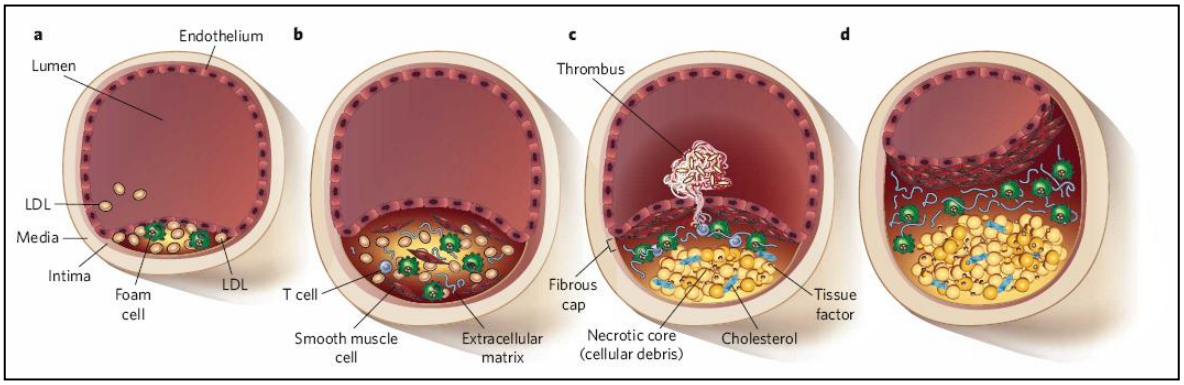

Figure 2.1. Development of atherosclerosis [41]

Upon the discovery of the inhibitors of HMG-CoA reductase (3-hydroxy-3-methylglutaryl coenzyme A reductase), also known as statins, there is a marked reduce in

the progression of atherosclerosis and the incidence of coronary and cerebrovascular events. Statins inhibit cholesterol biosynthesis and result in accelerated clearance of low density plasma lipoproteins (LDLs), which is the main lipid component of cholesterol. Statins received remarkable attention to understand the molecular mechanisms that regulate plasma LDL concentrations. It has long been known that the amount of functional LDL receptor present at the hepatocyte surface is one of the most important factors influencing plasma LDL concentration. However, LDL-receptor regulation is more complicated than anticipated. Genetic mutations that cause the rare disease autosomal recessive hypercholesterolemia result in disruption of LDL-receptor recycling and in a substantial reduction in the number of LDL receptor molecules at the hepatocyte surface, thus markedly increasing plasma cholesterol concentrations. LDL plasma concentrations are also dependent on the hepatic production rates of very-low-density lipoproteins (VLDLs), the metabolic precursor of LDLs. Indeed, hepatic overproduction of VLDLs is a common finding in individuals with insulin resistance or Type 2 diabetes and is also the basis of familial combined hyperlipidaemia, a common genetic lipoprotein disorder. Upstream transcription factor 1 (USF1) has been genetically associated with familial combined hyperlipidaemia, although the molecular mechanisms underlying the effect of USF1 on VLDL production are unclear. Studies of humans with low LDL concentrations have also provided important insights into the regulation of VLDL and LDL production. Mutations in the gene encoding APOB, which is the key structural protein component of VLDLs and LDLs, can result in low LDL concentrations, at least in part by reducing VLDL production. As is the case for LDLs, the main lipid component of high density lipoproteins (HDLs) is cholesterol. However, in contrast to LDL cholesterol, plasma concentrations of HDL cholesterol are inversely associated with atherosclerotic disease [41]. HDL-cholesterol levels are inversely correlated with the risk of coronary heart disease. A low level of HDL is the most common lipid abnormality observed in patients with coronary artery disease. Clinical trials demonstrated that there is a reduced incidence of coronary events in association with an increase in the plasma HDL levels [43]. The efflux of cholesterol from foam cells is mediated by HDL or its apolipoproteins and represents a crucial step in the prevention or reversal of atherosclerosis. Therefore the potential regulators of HDL have been extensively studied [44-47].

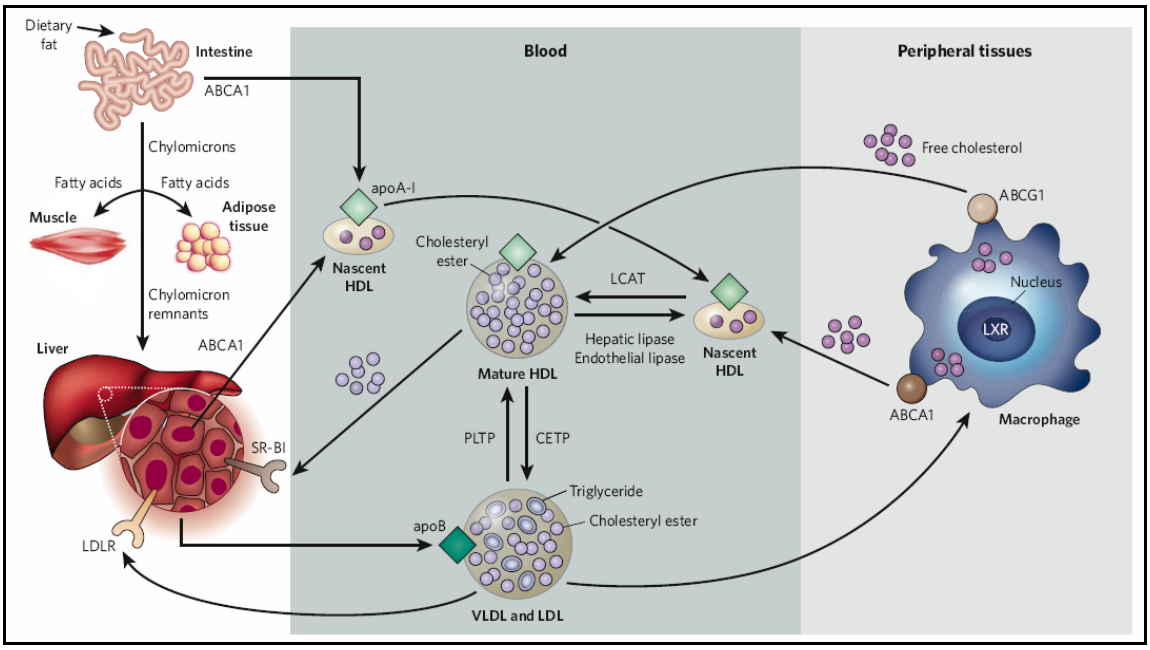

Figure 2.2. Lipoprotein metabolism has a key role in atherogenesis

Hence, the lipoprotein metabolism involves the transport of lipids, particularly cholesterol and triglycerides, in the blood. The intestine absorbs dietary fat and packages it into chylomicrons (large triglyceride-rich lipoproteins), which are transported to peripheral tissues through the blood. In muscle and adipose tissues, the enzyme lipoprotein lipase breaks down chylomicrons, and fatty acids enter these tissues. The chylomicron remnants are subsequently taken up by the liver. The liver loads lipids onto APOB and secretes VLDLs, which undergo lipolysis by lipoprotein lipase to form LDLs. By contrast, HDLs are generated by the intestine and the liver through the secretion of lipid-free APOA1. APOA1 then recruits cholesterol from these organs through the actions of the transporter ABCA1, forming nascent HDLs, and this protects APOA1 from being rapidly degraded in the kidneys. In the peripheral tissues, nascent HDLs promote the efflux of cholesterol from tissues, including from macrophages. HDLs can promote the activity of nitric-oxide synthase 3 (NOS3; also known as eNOS) and thereby increase the bioavailability of nitric oxide [29].

2.1.2. Coagulation Cascade

Blood coagulation and platelet-mediated primary haemostasis have evolved as important defense mechanisms against bleeding. The coagulation system is triggered in

response to rupture of endothelium, which allows exposure of blood to the extravascular tissue. At sites of vascular injury, activation of blood coagulation results in the generation of high concentrations of thrombin that activate platelets and coagulate blood. Anticoagulant mechanisms ensure careful control of coagulation and, under normal conditions; they prevail over the procoagulant forces. The efficient coagulation system is controlled by several anticoagulant mechanisms, ensuring that the clotting process remains a local process. Depending on the factors affecting the initiation of the coagulation process, the cascade is mainly composed of two paths: extrinsic and intrinsic pathway.

The initiation of the extrinsic pathway in the coagulation system is a result of exposure of tissue factor (TF) to blood and the subsequent binding and activation of factor VII (FVII). TF serves as a cofactor to the enzyme FVIIa, the TF-FVIIa complex efficiently activating factor IX (FIX) and factor X (FX). The subsequent reactions occur on the surface of negatively charged phospholipid membranes, exposed on activated platelets, where the blood coagulation proteins bind and assemble into enzymatically active complexes. Thus, FIXa binds to its cofactor FVIIIa forming the tenase (FIXa-FVIIIa) complex that activates additional FX, whereas FXa together with FVa form the prothrombinase (FXa-FVa) complex that efficiently converts prothrombin (PT) to thrombin. FVIIIa and FVa serve as important cofactors to the enzymes FIXa and FXa, respectively. Indeed, without the cofactors and the negatively charged phospholipid, the efficiency of the two enzymes FIXa and FXa is negligible, further ensuring that the enzymatic reactions remain localized. The generated thrombin positively feedback-activates the coagulation system by converting circulating precursors FVIII and FV into their active forms. The whole system is designed to provide massive amplification of an initiating stimulus and if not appropriately controlled, it would rapidly convert circulating blood into a clot [48]. On the other hand, the intrinsic pathway is an alternative mechanism by which the coagulation system can be initiated. It involves factor XII, high-molecular-weight kininogen, prekallikrein, and factor XI; it results in the activation of factor XI [49].

Control of blood coagulation is achieved by several anticoagulant mechanisms and at all levels of the system. Tissue factor pathway inhibitor (TFPI) regulates the very initiation of coagulation. It binds and inhibits newly formed FXa that is associated with the TF-FVIIa complex. The activated coagulation enzymes can all be inhibited by the circulating

serine protease inhibitor (serpin) antithrombin (AT), in particular when the enzymes are not engaged with their respective cofactors. Another mechanism of control is achieved by regulation of the two cofactors FVIIIa and FVa by the protein C (PC) anticoagulant system [48].

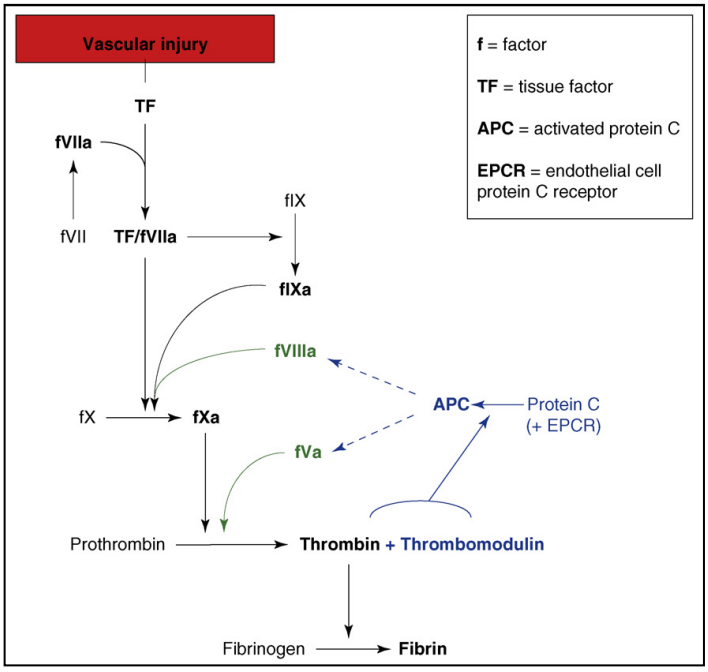

Figure 2.3. Schematic representation of the coagulation cascade and protein C pathway

After vascular injury or inflammation, tissue factor (TF) is expressed on the endothelial surface and initiates a series of proteolytic reactions that result in generation of thrombin and formation of a fibrin clot. Binding of thrombin to thrombomodulin (TM), accelerates activation of protein C, which is presented to the thrombin–TM complex by the endothelial cell protein C receptor (EPCR). Activated protein C (APC) proteolytically inactivates cofactors Va and VIIIa, thus preventing propagation of the cascade and inhibiting thrombin generation [50]. The figure also highlights the dual role of thrombin as both a procoagulant and anticoagulant factor [48].

The PC pathway has dual functions. Traditionally, it has been regarded as an important regulator of coagulation because it maintains the fluidity of the vasculature to prevent thrombosis. However, several anti-inflammatory and cytoprotective activities of the PC pathway are becoming recognized. The major components of the PC pathway include thrombomodulin (TM), the endothelial cell protein C receptor (EPCR) and PC. TM

is highly expressed on the surface of the microvasculature, whereas EPCR is more prominently found on endothelial cells of larger vessels. PC is produced in the liver and is readily detectable in the systemic circulation. The PC pathway was originally described as a major anti-coagulant system (Figure 2.4). As clot formation progresses, TM binds thrombin, thereby directly inhibiting its procoagulant potential while augmenting the conversion of PC to its active form at the same time. In fact, the thrombin–TM complex increases the rate of PC activation by more than a 1000-fold. This process is facilitated by localization of PC on the endothelial cell surface by binding to EPCR, which enhances activation of PC 20-fold *in vivo*. Activated PC (APC) then exerts its anti-coagulant function by binding protein S and inactivating factors Va and VIIIa, thus inhibiting further thrombin generation [50].

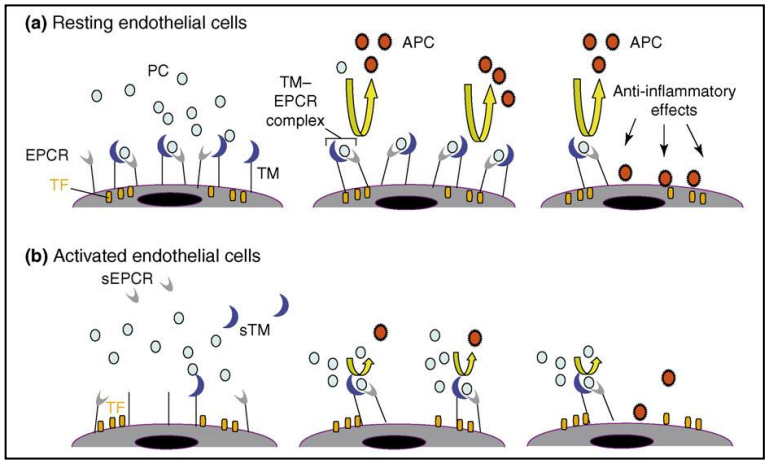

Figure 2.4. Overview of the protein C (PC) pathway [50]

Besides its role as a critical regulator of coagulation activity, the PC pathway is emerging as an important mediator of inflammation. These anti-inflammatory effects occur independently of the pathway's anti-coagulant function and appear to result from modification of the response of the endothelium and immune cells to an inflammatory stimulus. Genetic manipulation of individual components of the PC pathway has underscored the vital role of this system in controlling inflammation. Each component of the pathway, individually or together, possesses significant anti-inflammatory properties [50].

2.1.3. Renin-Angiotensin Mechanism

The circulating renin-angiotensin system (RAS) consists of angiotensinogen (AGT) synthesized in the liver and cleaved by renin, an aspartyl protease secreted from juxtaglomerular cells in the afferent arterioles of the kidney. Angiotensin I (ANGI), the decapeptide product of the reaction, is further converted by ACE to angiotensin II (ANGII), the active octapeptide. ANGII exerts its physiological effects in humans by binding to two distinct receptors, AGTR1 and AGTR2. AGTR1 mediates the classical vasoconstrictor action of ANGII, whereas AGTR2 has opposing effects. Renin is the first and limiting step of this catalytic cascade. In addition, ANGII inhibits renin secretion through a negative feedback mechanism [51].

As the renin–angiotensin system plays such a large role in the development of hypertension and, subsequently, atherosclerosis, studies have examined whether blocking this system can prevent the onset of endothelial changes. In the tissue renin-angiotensin system, ANGII causes the degradation of bradykinin, which stimulates nitric oxide production and prostaglandin release. ACE, which is identical to kininase II, also causes bradykinin degradation. Thus, ACE inhibition's benefits appear to be due to the resulting bradykinin accumulation. There is abundant information suggesting a role for ACE-inhibitor therapy in reducing myocardial hypertrophy, vascular hypertrophy, atherosclerosis progression, plaque rupture, and thrombosis after plaque rupture [52, 53].

2.1.4. Oxidative Stress in CVD

Oxidants, including reactive oxygen species (ROS), are constantly produced in cells through normal metabolic processes. Oxidative or oxidant stress occurs when the balance of oxidants within the cell exceeds the levels of antioxidants present. An increased level of ROS can lead to damage of macromolecules within the cell; and it is this damage to lipids, proteins, and DNA that can give rise to pathological consequences. There is considerable overlap not only in the pathology but also in the etiology and underlying molecular mechanisms of oxidant stress-dependent diseases [54].

2.1.5. Inflammation Pathophysiology in CVD

Cytokines are key regulatory glycoproteins that function in inflammatory / immunological processes which modulate all aspects of vascular inflammation by altering the proliferation, differentiation and function of an extensive array of cell types. Cytokines target specific receptors to function. 200 cytokines are characterized until recently, but the due to their complex biological function, it is hard to generate a classification. Structurally, cytokines were categorized into nine families as given in Table 2.1.

Table 2.1. Cytokine classification and implications in atherosclerosis [55]

Cytokines	**Family**	**Implication in atherosclerosis**
IL-1α, IL-1β, IL-1ra, IL-18	IL-1 family	Pro-atherogenic except IL-1ra
IL-2, IL-4, GM-CSF	IL-2 family	Pro-atherogenic. IL-4 levels elevated in aortic aneurysms but precise categorization of IL-4v and GM-CSF as atherogenic or protective is not possible
IL-6, IL-12	IL-6 family	Pro-atherogenic
CXCL1-16, CXCL-28, CXCL1/fractalkine, IL-8, MCP-1	Chemokines	Pro-atherogenic
IL-10	IL-10 family	Protective
IL-17	IL-17 family	Pro-atherogenic
IFN-γ	Interferon family	Pro-atherogenic
TGF-β	TGF-β	Protective
TNF-α, TNF-β, LT-β	Tumor necrosis factor family	Pro-atherogenic

The involvement of cytokines in the development of cytokines is at the initiation of atherosclerosis: endothelial injury or dysfunction. Endothelial cells play a pivotal role in atherogenesis. They modify LDL, directing macrophage uptake via the scavenger receptors (SRs). Endothelial cells express Toll-like receptors (TLRs), which comprise a family of pattern recognition receptors, also associated with macrophages. Endothelial cell TLR ligation induces NF-kB activation with consequent gene expression of leukocyte adhesion molecules (LAMs), IL1, and other inflammatory mediators. Endothelial cells also express SRs, such as CD36, which can engulf modified lipoproteins. Thus, through innate immunoreactive receptor engagement, augmentation of EC antigen-presenting function occurs. Although less efficient antigen presenters than macrophages, ECs can present foreign antigens to specific T-cells, thereby providing an early link between innate and adaptive immunity [56].

2.1.6. Apoptosis Pathway

Apoptosis is a highly regulated programmed cell death and initiated through the activation of death receptor pathway via a complex signal transduction mechanism starting from plasma membrane resulting in the activation of caspase cascade. The well known death receptor pathway members, such as FAS, and TNFR1 have been implicated to contribute to cardiovascular disease. The ligand for TNFR1 is TNF was reported to exhibit elevated levels with congestive heart failure patients. In addition to TNFR1, FAS and FASL have been reported to induce apoptosis in cardiac myocytes. There is evidence suggesting that Fas pathway contributes to cell death in the later stages of infarction [57]. Moreover, Monocyte chemoattractant protein-1 (MCP1) was reported to play an important role in initiating coronary heart disease by recruiting macrophages and monocytes to the arterial wall [58].

Recently, the involvement of apoptosis in cardiovascular disease has been reviewed [57]. It has been reported that macrophages and smooth muscle cells undergo apoptosis in unstable atherosclerotic plaques which can lead to rupture of the plaque and thrombosis [59]. There is also accumulating evidence that apoptosis plays an important role in both acute and chronic loss of cardiac myocytes after a myocardial infarction [60].

2.2. Type 2 Diabetes as a Model System

Diabetes is a heterogeneous group of metabolic disorders, which is characterized by glucose intolerance and the development of these disorders can be due to both environmental and genetic contributions. Single gene forms of diabetes are relatively rare and account for approximately five per cent of diagnosed cases with maturity-onset diabetes of the young (MODY) the most common. This group of disorders is of pancreatic β-cell origin and is caused by mutations in a variety of genes. In the early 1990s, the first MODY gene was identified as that encoding the glycolytic enzyme glucokinase. The majority of diabetic disorders are multifactorial. Type 1 diabetes (T1D), or insulin-dependent diabetes mellitus (IDDM), is caused by auto-immune destruction of the β-cells of the pancreas, rendering the pancreas unable to synthesize and secrete insulin T1D

accounts for between five per cent and 10 per cent of all cases of diabetes with the major susceptibility gene mapping to the HLA region of chromosome 6.

Type 2 diabetes (T2D), or non-insulin dependent diabetes mellitus (NIDDM), is the most common form of the disease world-wide, accounting for over 90 per cent of diabetes cases, and it has been projected that by 2025, the prevalence of this form of the disease will affect 366 million people world-wide [61]. Type 2 diabetes is characterized by a combination of impaired insulin secretion and insulin action, both of which precede and predict the onset of disease. Although environmental factors, such as dietary habits and sedentary life [62], play important roles in the progress of the disease, it is now well-known phenomenon that the disease susceptibility is influenced by genetic factors. Despite strenuous efforts over the last two decades had been embarked on the identification of genetic variants that contribute to individual differences in predisposition of T2D, susceptibility genes are mostly identified through genome-wide analysis [63].

2.2.1. Insulin Signaling Pathway

Insulin resistance, a reduced ability of a tissue to respond to insulin, is an important component resulting in the development of Type 2 diabetes and the metabolic syndrome, and can occur with obesity, inflammation and increasing age. Insulin is the most potent anabolic hormone known, and promotes the synthesis and storage of carbohydrates, lipids and proteins, while inhibiting their degradation and release into the circulation. Insulin stimulates the uptake of glucose, amino acids and fatty acids into cells, and increases the expression or activity of enzymes that catalyze glycogen, lipid and protein synthesis, while inhibiting the activity or expression of those that catalyze degradation (Figure 2.5) [64] .

Insulin signaling is one of the central signaling cascades (Figure 2.6). The insulin receptor is a tyrosine kinase that can autophosphorylate, and promotes the phosphorylation of the other cellular proteins. Upon phosphorylation, these proteins interact with other signaling molecules, resulting in allocation through diverse signaling pathways, such as PI(3)K, RAS and MAPK cascade. These pathways coordinate the regulation of vesicle trafficking, protein synthesis, enzyme activation / inactivation, and gene expression, which results in the regulation of glucose, lipid and protein metabolism [64, 65]. The growing body of

evidence that supports a role for specific members of the suppressor of cytokine signaling (SOCS) family of proteins in the development of leptin and insulin resistance through their ability to inhibit leptin and insulin signaling pathways [65] and target IRS1 and IRS2 for proteosomal degradation [66].

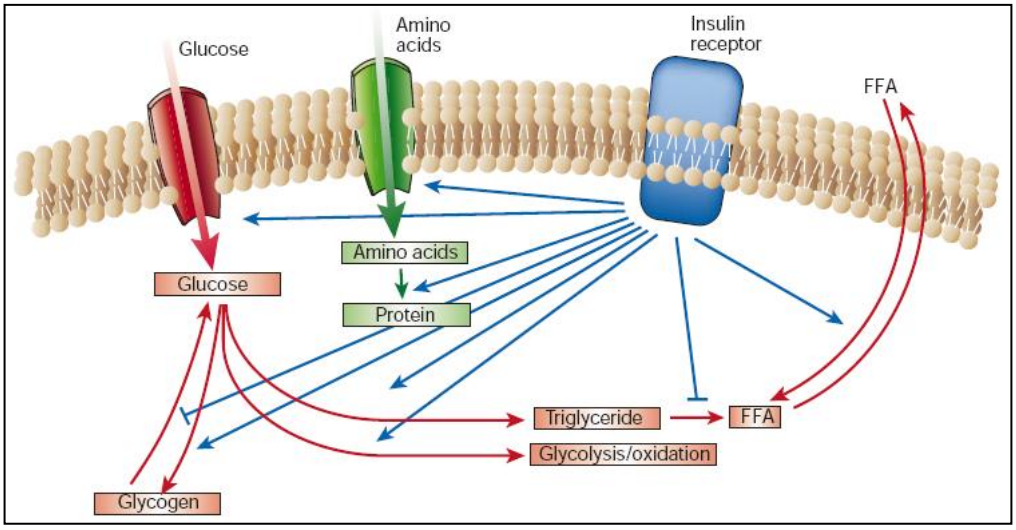

Figure 2.5. The regulation of metabolism by insulin [64]

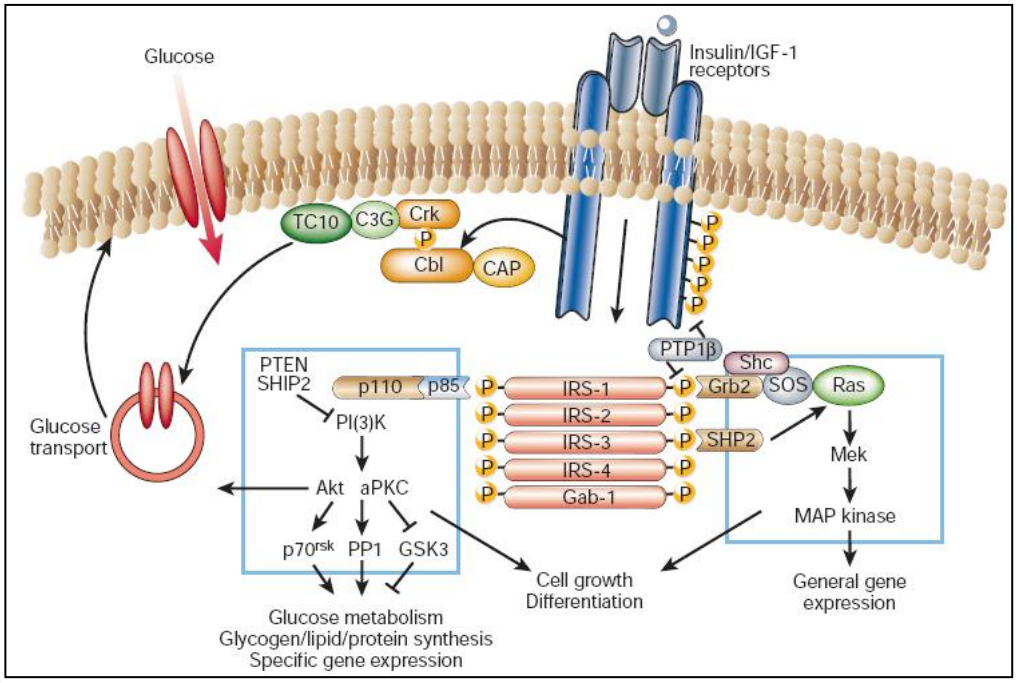

Figure 2.6. Signal transduction in insulin action [64]

2.2.2. Leptin Signaling Pathway

Leptin activates cytokine-like signal transduction by stimulation of the JAK–STAT pathway via its receptor (LRb) (Figure 2.7) [67].

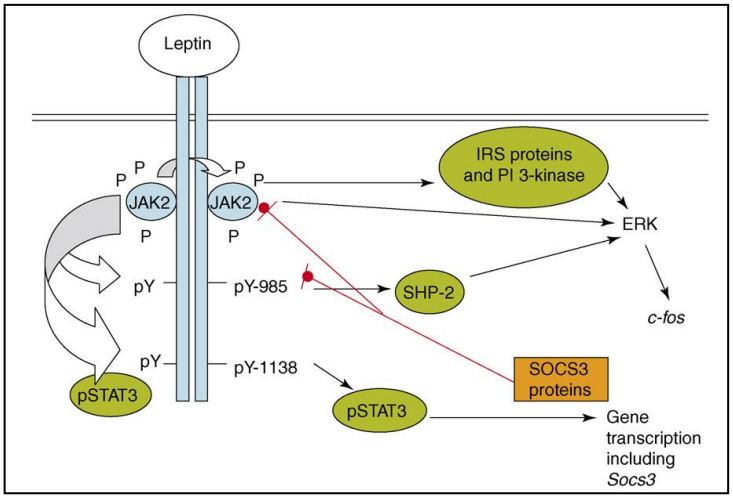

Figure 2.7. Leptin receptor signaling [65]

Similar to other cytokine receptors, the leptin receptor lacks intrinsic tyrosine kinase catalytic activity and in close contact with JAK2 and STAT3. This pathway appears to be crucial in the action of leptin on energy expenditure. STAT3 activation also results in the transcription of the signaling inhibitor SOCS3. Other leptin receptor signals arise from the JAK2 phosphorylation sites themselves. Leptin receptor signaling also results in the activation of the IRS-phosphorylation–PI3-kinase pathway, probably through phosphorylation sites on JAK2. Current evidence suggests that the effects of leptin on reproduction, growth, and perhaps glucose homeostasis and immune function might be independent of the STAT3 pathway. SOCS3 was proposed as a potential mediator of leptin resistance [68], which attenuates leptin receptor signaling [65]. IL6 also directly influences human adipocyte metabolism by decreasing the activity of lipoprotein lipase (LPL), an enzyme that regulates uptake of circulating triglycerides into adipocytes. The investigation of influence of IL6 on leptin production in adipocytes concluded that IL6 can modulate leptin production and lipid metabolism in human adipose tissue [69].

2.2.3. Lipoprotein Metabolism in T2D

The alteration in the lipoprotein profiles in T2D is due to an elevation of very low density lipoprotein accompanied by a decrease in the high density lipoprotein. The other main source of circulating triglycerides is the chylomicrons, which carry dietary triglycerides. Triglycerides within VLDL and chylomicrons undergo lipolysis. The resultant lipoproteins, which are low-density lipoproteins (LDL) and chylomicron remnants, are removed from the circulation primarily via receptors in the liver [70]. In diabetic mouse models, the lipoprotein metabolism is regulated by angiopoietin-like protein 3 (ANGPTL3), by inhibiting lipolysis of triglyceride-rich lipoproteins. ANGPTL3 may induce an endogenous inhibitor of lipolysis, such as the apolipoproteins CI and CIII or APOE. Another possible mechanism is the action of the ANGPTL3 itself is a lipolysis inhibitor, either by attaching to the lipoprotein and preventing its lipolysis or by directly inhibiting the rate-limiting enzyme for the lipolysis of triglyceride rich proteins, LPL [71].

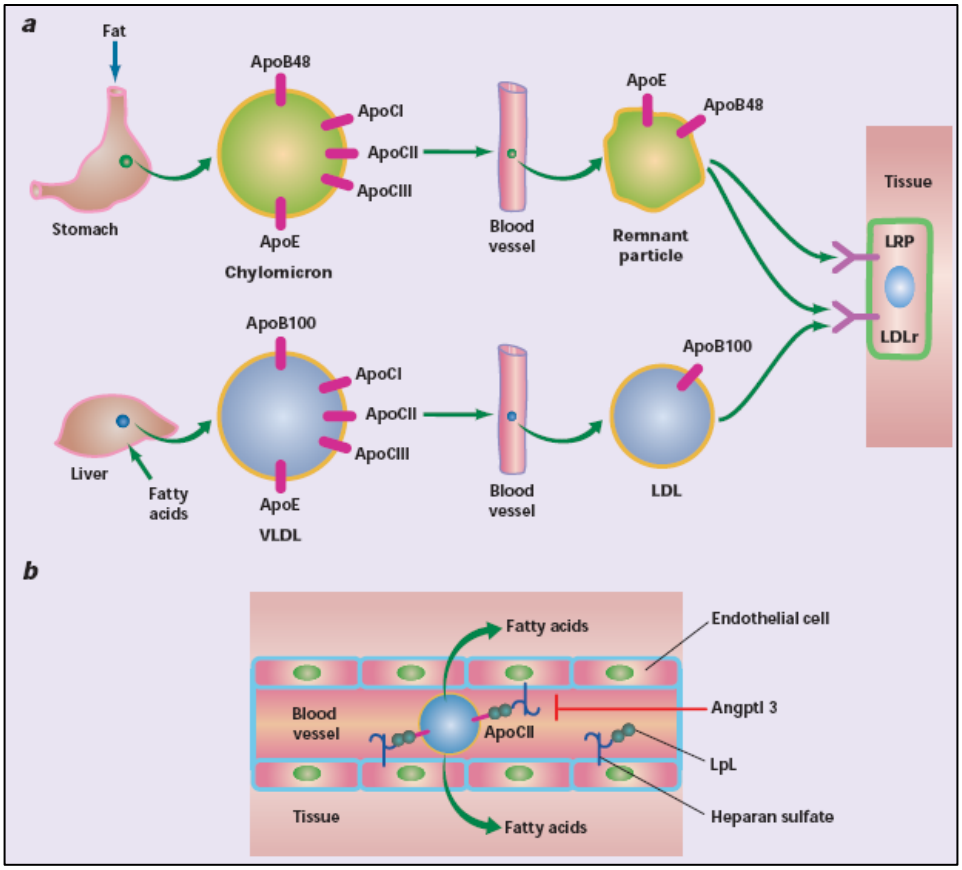

Figure 2.8. Representation of major steps in the metabolism of triglyceride-rich lipoproteins (a) metabolism (b) lipolysis [70]

Metabolism of triglyceride-rich lipoproteins is initiated with the uptake of dietary fat. The two classes of lipoproteins whose major lipid is triglyceride are chylomicrons and VLDL. These particles are initially secreted into the lymphatics. Once in the bloodstream, much of the triglyceride is hydrolyzed and the smaller particles are removed from the bloodstream by LDLr and LRP. APOE serves as a ligand for these receptors. VLDL contains triglycerides assembled in the liver. Some of these fatty acids are derived from adipose tissue when hormone-sensitive lipase is activated. VLDL triglyceride is also lipolyzed within the bloodstream and the remaining lipid, primarily cholesterol and cholesteryl ester, circulates as LDL. APOB100, the major structural protein of LDL, contains a LDL receptor-binding region; this region is not contained in APOB48, the truncated form of APOB found in chylomicrons. APOCI and APOCIII are smaller proteins that modulate lipolysis and interaction of triglyceride-rich lipoproteins with receptors.

2.2.4. β-Cells

According to the studies in the last decade, it is now a well known phenomenon that obesity is one of the major risk factors for the development of Type 2 diabetes [72] and is presented to increase the risk for Type 2 diabetes through the mechanism of associated insulin resistance [73]. A substantial increase in the β-cell mass was observed in nondiabetic obese patients. Regulation of the β-cell mass appears to involve in the β-cell replication and apoptosis, therefore alterations in these pathways β-cell formation or increase rates of β-cell death could lead to a decrease in the β-cell mass. However, the studies performed on β-cell mass in Type 2 diabetes indicated that β-cell mass is decreased in both obese and lean humans with Type 2 diabetes compared with their non-diabetic age- and weight-matched counterparts [74].

Many studies have suggested that β-cell dysfunction results from prolonged exposure to high glucose, elevated free fatty acid levels, or a combination of both. Cells are particularly sensitive to ROS because they are low in free-radical quenching (antioxidant) enzymes. The key role of increased glucose metabolism in producing impaired β-cell function through oxidative stress has recently been confirmed. Intracellular ROS increased

15 minutes after exposure to high glucose, and this effect was blunted by inhibitors of the mitochondrial function [75].

2.2.5. Oxidative Stress in T2D

Oxidative stress refers to a persistent imbalance between excessive production of ROS(reactive oxygen species) and/or RNS (reactive nitrogen species) and limited antioxidant defences [76]. The occurrence of Type 2 diabetes is associated with an elevated level of oxidative stress, which results from hyperglycaemia through glycoxidation and sorbitol system activation and from limitation of the hexose monophosphate shunt, leading to a decrease in glutathione synthesis. The accumulation of the products of the oxidative stress, especially reactive oxygen species (ROS), can cause the damage to biological macromolecules: proteins, lipids and DNA. Diminished activity of glutathione peroxidase, catalase and superoxide dismutase and decreased antioxidant activity was observed in diabetic patients, which contribute to their vulnerability to oxidative stress-related diseases, including atherosclerosis [77].

The most important tissues involved in the pathogenesis of insulin resistance are muscle and adipose tissue. When caloric intake exceeds the energy expenditure, the substrate-induced increase in citric acid cycle activity generates an excess of mitochondrial NADH (mNADH) and ROS. To protect themselves against harmful effects of ROS, cells may reduce the formation of ROS and/or enhance ROS removal. Prevention of ROS formation is accomplished by preventing the build-up of mNADH by inhibiting insulin stimulated nutrient uptake and preventing the entrance of energetic substrates (pyruvate, fatty acids) into the mitochondria [75].

2.2.6. DNA Repair Pathway

There is growing evidence that diabetes can be also associated with cancer. This association can involve cancer appearance and/or its progression. The involvement of DNA repair pathway in T2D was compared in a cohort study. T2D was suggested to be associated not only with the elevated level of oxidative DNA damage but also with the increased susceptibility to mutagens and the decreased efficacy of DNA repair. These

findings implicated that the cellular responses to DNA damage in diabetes in diabetes contribute to its association with cancer [77].

There is strong evidence that insulin regulates the synthesis of the DNA repair enzyme XPD, playing a pivotal role in nucleotide excision repair. A long exposure to glucose at a high concentration may result in downregulation of the insulin-dependent increase in XPD mRNA levels and increasing the extent of DNA damage [78]. Results of these studies also implicate that the degree of insulin resistance should be taken into account when discussing DNA damage and repair in diabetes. Increased susceptibility to mutagens and impaired DNA repair can contribute to the genomic instability in diabetes patients and, in consequence, to cancer [77].

2.3. Alzheimer's Disease as a Model System

Alzheimer's disease (AD) is a disorder of the brain region that controls the human central nervous system and the cerebral cortex, which is characterized by a memory deterioration, progressive cognitive deterioration, emergence of neuropsychiatric symptoms and behavioral disturbances, impairment of activities of daily living, and loss of independent function [79, 80]. Aβ precursor protein (APP), presenilin 1 and 2 (PSEN1, PSEN2), and apolipoprotein E (APOE) have been implicated in the pathogenesis of AD [81]. The hallmarks of AD are tangles (abnormal paired helical filaments of the protein tau, which normally binds to microtubules) and plaques (extracellular deposits composed primarily of β-amyloid protein (Aβ)) within the brain [82].

Amyloid precursor protein, encoded by APP, is an integral membrane protein with a single membrane-spanning domain, a large extracellular N terminus and a shorter cytoplasmic C terminus. The normal functions of APP in the body are only partly understood. Amyloid peptide, is a 40–42 amino acid peptide; the 42 amino acid form is the most toxic and prone to aggregation, and is the primary component of diffuse or neuritic plaques [83]. Processing of APP occurs by two major protease pathways. Cleavage of APP at the N-terminus of the Aβ region by β-secretase, and at the C-terminus by γ-secretase, represents the amyloidogenic pathway for processing of APP to form Aβ [84]. While Aβ production seems to be a fairly straightforward process involving two proteases (β- and γ-

secretases), APP processing in fact is a quite intricate and complex mechanism constituting multiple sites of feedback regulation [85].

The mutations in the genes presenilin 1 (PSEN1) and presenilin 2 (PSEN2) have been associated with AD patients [86]. The proteins encoded by PSEN1 and PSEN2 are located in intracellular membranes such as nuclear envelop, the endoplasmic reticulum and Golgi apparatus. They are primarily expressed in neurons and are ubiquitously expressed within the brain. Presenilin proteins have also been proposed to function in the control of apoptosis. PSEN1 is required for proper formation of the axial skeleton and is involved in normal neurogenesis and survival of progenitor cells and neurons in specific brain regions. PSEN1 is also involved in γ-secretase activity and binding of PSEN proteins to APP may play an important role in inducing intercellular signaling. Two conserved transmembrane aspartate residues in PSEN1 are critical for Aβ production, suggesting that PSEN1 either functions as an essential cofactor for g-secretase, or is itself g-secretase [87]. Like PSEN1, PSEN2 contains two transmembrane apparatus critical for g-secretase activity. Furthermore, gene deletion of PSEN1 shows that it required for normal proteolytic cleavage of β amyloid precursor protein to generate Aβ [88]. The exact mechanism of cleavage of APP at the residues 40-42 as well as how it is affected by inherited mutations in presenilins remains elusive [89], however, some of the presenilin-interacting proteins can be subdivided into functional groups, including those involved in the γ-secretase protein complex, Wnt signalling, cell–cell adhesion, vesicular transport, apoptosis and calcium signaling [90].

APOE, the major apolipoprotein of the central nervous system, has a central role in cholesterol transport. In the brain, APOE is mostly produced by astrocytes, and some binds to specific neuronal receptors for cholesterol uptake and has been thought to have a role in repair of nerve cells in response to damage. APOE has been implicated to have a specific impact on hippocampus degeneration, particularly by oxidative metabolism. Besides its role in lipid integrity and synaptic plasticity, APOE could also directly or indirectly interfere the status of some antioxidants in the brain [91]. APOE exists in three allelic variants, namely E2, E3 and E4, with a frequency of eight per cent, 77 per cent and 15 per cent, respectively. Biochemical tests have confirmed a direct interaction of APOE with Aβ

and with APP [91]. Isoform 4 (APOE4) is significantly associated with sporadic AD, whereas isoform 2 might protect against AD [83].

The neural microtubule (MT)-associated protein (MAP) tau is essential for the proper development and maintenance of the nervous system [92]. The major biological function of tau is to promote microtubule assembly and maintain the stability of the previously formed microtubules, which are essential for the axonal transport of the neurons. Additionally, the interaction of tau with diverse structural and functional proteins suggests that tau may play crucial roles not only in normal architecture but also in signal transduction of the neurons [93]. The abnormal hyperphosphorylation of tau in AD brain and resulting formation of paired helical filaments, which forms neurofibrillary tangles, is a hallmark lesion of this disease. Tau pathology, which is seen only as accumulation of abnormally hyperphosphorylated protein, is also seen in several other human neurodegenerative disorders [94]. However, tau increasingly seems obligatory to the processes leading to neurodegeneration in the tauopathies and lack, or a reduced amount, of tau is associated with neuroprotection and resistance to Aβ toxicity [95, 96]. Conversely, elevated tau expression enhances susceptibility to toxic stimuli and/or neurodegeneration and also results in increased production of neurotoxic amyloidogenic peptides [97]. It has been anticipated that tau phosphorylation is a pre-requisite for its aggregation, although this has yet to be proven. An alternative possibility is that tau aggregates before becoming phosphorylated, leaving it in a conformationally altered state that could protect the deposited tau from the action of protein phosphatases [98].

Initial studies on the disease focused on two major hypotheses to explain the molecular mechanism of disease: the cholinergic hypothesis and the amyloid cascade hypothesis. The former postulates a dysfunctional cholinergic system, which is the neurotransmission of acetylcholine, is sufficient to explain the memory loss in Alzheimer's dementia in parallel with the animal models. Indeed, reduced number of cholinergic markers, choline acetyltransferase and acetyl cholinesterase has been reported in the cerebral cortex of AD brain and degeneration of cholinergic neurons were observed in AD patients [99, 100]. Although cholinergic deficits cannot fully explain the overall neuropathological events observed in AD, it represents an important section for AD progression. The latter, amyloid hypothesis, states that the neurodegenerative process

observed in AD brains is initiated with the accumulation of proteinacious amyloid deposits. The presence of these amyloid plaque deposits is the essential observation underpinning the amyloid hypothesis [101]. The primary component of amyloid plaque is aggregated Aβ-peptide derived from the enzymatic processing of the Alzheimer precursor protein (APP) [81, 91]. Along with the amyloid cascade, other pathophysiological processes, including oxidative stress, inflammation, insulin signaling and cholesterol mechanism, received remarkable attention to generate a comprehensive understanding of the disease.

2.3.1. Amyloid Cascade

The amyloid cascade centralizes Aβ as the trigger of the pathological changes observed in the brains of AD patients, such as synapse loss, activation of inflammatory processes, the induction of neurofibrillary changes leading to the formation of helical filaments and, ultimately, neuronal death [102]. Accumulation of the hydrophobic Aβ40 and Aβ42 peptides results in aggregation and formation of insoluble plaques, which trigger a cascade of deleterious changes, resulting in neuronal death and thus causing AD [103].

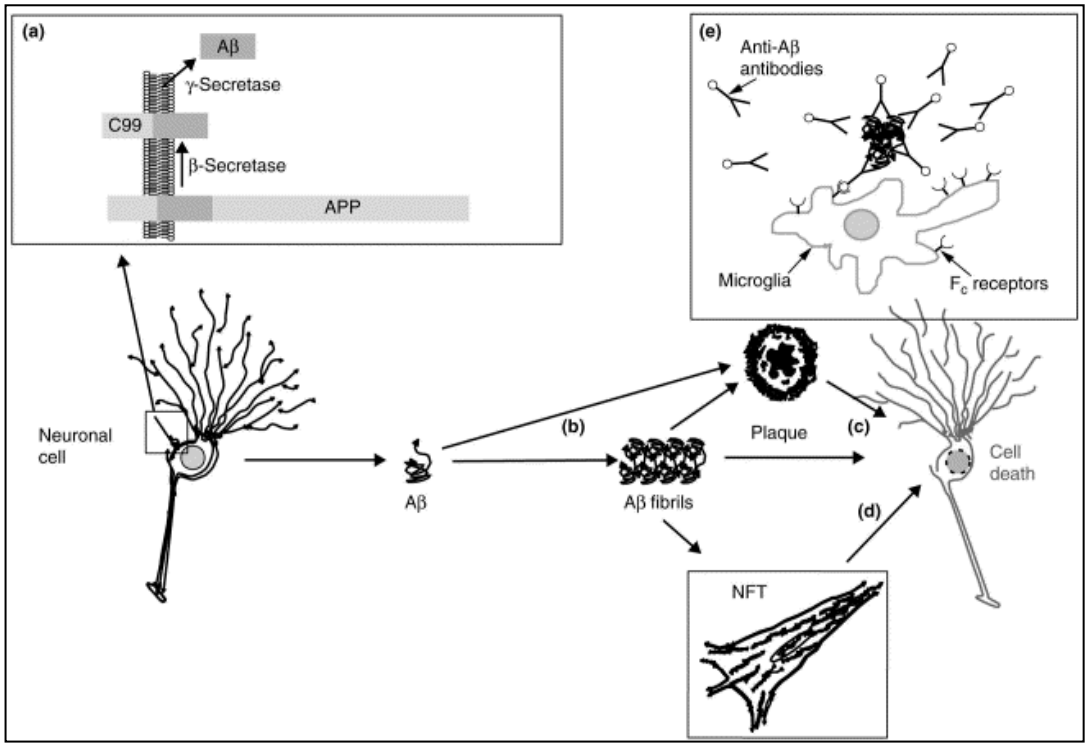

Figure 2.9. Illustration of the amyloid cascade hypothesis [102]

After processing of APP by β- and γ-secretase, Aβ is secreted (Figure 2.9.a) and aggregates into fibrils that deposit in senile plaques (Figure 2.9.b). Fibrils and plaques can induce neurotoxicity directly or indirectly, e.g. by the formation of neurofibrillary tangles (Figure 2.9.c-d). A possible treatment concept is immunization with Aβ, which can induce an F_c receptor mediated clearance process by phagocytic microglia cells (Figure 2.9.e) [102].

2.3.2. Oxidative Stress in AD

Oxidative stress is produced by free radicals, i.e. reactive oxygen species (ROS) that are generated by oxygen- and nitrogen-based molecules that have unpaired electrons. Because of unpaired electrons, free radicals are very unstable and highly reactive. To overcome unpaired state, the free radicals coupled with the electrons from other molecules. After losing the electrons, the donor molecules become unstable and are converted into free radicals themselves [84]. The free radicals produced in the body are toxic, and if not removed or neutralized, they react with lipids, proteins, and nucleic acids and damage cellular functions. Generally, oxidative damage to the cellular components results in alteration of the membrane properties such as fluidity, ion transport, enzyme activities, and protein cross-linking. Excessive oxidative damage eventually results in cell death. Experimental studies using animal models and human brain suggest that oxidative stress plays an important role in neuronal degeneration in AD. Soluble Aβ contributes to increased oxidative stress in AD, leading to increased lipid peroxidation, which has been implicated as an important factor in AD [84]. Other effects of Aβ-mediated oxidative stress are attenuation of functional hyperemia, learning and memory deficits in mice, increased levels of heme oxygenase-1 and heat shock protein 72 [104-106]. Additionally, in the presence of copper ions, Aβ reduces Cu^{2+} to Cu^{+} and catalyses the formation of H_2O_2 that produces highly toxic hydroxyl radicals [107].

2.3.3. Complement Cascade

The complement cascade is a pathway that is largely involved in the recruitment and chemotaxis of inflammatory cells and direct induction of cell death [108]. Complement cascade is a sophisticated attack system, designed to provide protection against invaders

and to assist in the phagocytosis of waste materials [82]. This cascade is mediated by several soluble complement factors that are synthesized by the liver and circulate in the blood. These main complement factors are C1 through C9, whereas several additional factors also exist, including B and D. The complement cascade consists of three distinct pathways that are distinguished by their trigger mechanisms, known as the alternative pathway, the classical pathway and the mannose-binding lectin (MBL) pathway. These pathways converge with the final formation of the membrane attack complex (MAC) [108]. A key finding regarding the mechanism of complement activation suggested that Aβ is a strong complement activator. Hence, the plaques found in AD have a unique activator of complement [109].

2.3.4. Inflammation Pathophysiology in AD

There is increasing evidence that the accumulation of β-amyloid may not be sufficient to induce the cognitive findings of AD and that the immune system may play a critical role in the clinical symptoms. It has been speculated that the inflammatory response associated with the presence of neuritic plaques could be involved in neuronal damage and with the progression of the disease. Activated microglia and reactive astrocytes surrounding extracellular deposits of amyloid β-protein initiate an inflammatory response characterized by a local cytokine-mediated acute phase response, activation of the complement cascade and subsequent further cell damage [110]. This idea is supported by the findings that elevated levels of both pro- and anti-inflammatory cytokines, such as interleukins IL1A, IL6, tumor necrosis factor TNF, and C-reactive proteins in plasma and/or cerebrospinal fluid of patients with AD [111, 112]. Together with the complement activation in the AD brain, increased numbers of brain T-cells were detected implying the presence of inflammation [113]. Recent studies showed that Toll-like receptors (TLRs), having central roles in innate immune system, have also been linked to the pathogenesis of AD [114, 115]. Notably, the immune system consists of two networks, the innate and the adaptive. Initiation of the innate network response occurs through the recognition of “pathogen-associated molecular patterns” that are produced by microorganisms. In general, recognition relies on interaction with TLRs and inflammation on downstream signaling to activate the transcriptional factor nuclear factor-kB that leads to increased transcription of proinflammatory genes [113].

2.4. Network Biology

The analysis of networks received considerable attention in last decade following the introduction of scale-free network model [116]. Many real networks, including www [13], scientific collaborations [117, 118], social networks [12] have been shown to exhibit scale free property, where the connectivity distribution is far from random. Networks of biological systems, such as metabolic interaction [119] and protein interaction networks [120], also share this property.

2.4.1. The Essence of Biological Networks

Interaction data collected through individual studies and large-scale analyses can be incorporated in network representation. To date, according to the definition of the interaction type, a variety of biological networks were constructed whose topological structures contain significant biological properties, hence eligible to deliver insights about the cellular processes. The networks based on the data retrieved from high-throughput experiments have been studies in detail including (i) protein-protein interaction networks composed of direct physical links among proteins, (ii) co-expression networks based on the similar expression profiles of proteins under comparable states or mutants, protein (iii) transcriptional networks that contain regulatory elements of a cellular mechanism, (iv) phosphorylation networks combination of active/inactive forms of proteins, (v) metabolic networks, and (vi) genetic networks.

Among those, protein-protein interaction networks represent the most diverse and largest datasets available. Initial studies incorporate the results obtained from yeast-2-hybrid screens, where the individual interactions are delivered from yeast using a transcriptional readout. Large scale two-hybrid screens were available for various organisms, including *Saccharomyces cerevisiae*, *Drosophila, Caenorhabditis elegans* and humans. To date, methods have been developed to predict direct protein interactions, resulting in an enourmous amount of data. The high throughput methods developed to infer protein interactions are presented in Table 2.2.

Table 2.2. High-throughput methods for detecting protein interactions [121]

Yeast-2-hybrid assay	Pairs of proteins to be tested for interaction are expressed as fusion proteins (‘hybrids’) in yeast: one protein is fused to a DNA-binding domain, the other to a transcriptional activator domain. Any interaction between them is detected by the formation of a functional transcription factor
Mass spectrometry of purified complexes	Individual proteins are tagged and used as ‘hooks’ to biochemically purify whole protein complexes. These are then separated and their components identified by mass spectrometry. Two protocols exist: tandem affinity purification (TAP)10,43, and high-throughput mass-spectrometric protein complex identification (HMSPCI)
Correlated mRNA expression (synexpression)	mRNA levels are systematically measured under a variety of different cellular conditions, and genes are grouped if they show a similar transcriptional response to these conditions. These groups are enriched in genes encoding physically interacting proteins
Genetic interactions (synthetic lethality)	Two nonessential genes that cause lethality when mutated at the same time form a synthetic lethal interaction. Such genes are often functionally associated and their encoded proteins may also interact physically. This type of genetic interaction is currently being studied in an all-versus-all approach in yeast
In silico predictions through genome analysis	Whole genomes can be screened for three types of interaction evidence: (1) in prokaryotic genomes, interacting proteins are often encoded by conserved operons; (2) interacting proteins have a similar ‘phylogenetic profile’; and (3) seemingly unrelated proteins are sometimes found fused into one polypeptide chain. This is an indication for a physical interaction.

Yeast-2-hybrid screens are performed *in vivo*; offers the detection of transient and unstable interactions; and it is independent of endogenous protein expression; and it has fine resolution, enabling interaction mapping within proteins. However, at each test, only two proteins are tested; since it takes place in the nucleus, so many proteins are not in their native compartment; and it also predicts possible interactions, but is unrelated to the physiological setting. Mass spectrometry of purified complexes takes its power from tagging simultaneously several members of a complex and detection of real complexes in physiological settings. However, it might miss some complexes that are not present under the given conditions; as well as tagging may disturb complex formation; and loosely associated components may be washed off during purification. Correlated mRNA expression is also an *in vivo* technique, and it has much broader coverage of cellular conditions than other methods. Despite its power on discriminating cell states or disease

outcomes, it is relatively an inaccurate predictor of direct physical interaction; and it is very sensitive to parameter choices and clustering methods during analysis. Genetic interactions technique is another indirect *in vivo* technique; and it is applicable to unbiased genome-wide screens. *In silico* protein interaction prediction techniques are fast and inexpensive and depending on the genomes screened, its coverage expands. However, a structured orthology is required to be defined between proteins [121].

2.4.2. Global Topological Properties

Network, often called as graphs in mathematics, is a representation for connections among interacting entities. The members of the organizations are called as nodes (*N*), where the links connecting the nodes are called as edges (*l*). The network theory was first introduced as a simple mathematical model by Erdos and Renyi, where the complex systems are considered to be connected randomly [122].

The basic metrics used to characterize the structure of a large network is its 'degree distribution', $P(i)$, which is the probability that a randomly chosen node will 'have degree *i*', that is be linked to *i* other nodes. For the Erdos– Renyi network, the degree distribution is given exactly by the binomial distribution or, in the limit of large *n*, by the Poisson distribution: $P(i)= m^i e^{-m}/i!$, where *m* is the average number of links.

However, as introduced first by Milgram's play "Six Degrees of Separation", Watts and Strogatz develops the idea of small world networks [123], and further advances in the network theory showed that the entities in the real networks have preferential connections [116]. The application of this hypothesis to various fields of research resulted in a network model, where the probability of finding *k* links is represented by Power law:

$$P(k) \alpha k^{-\gamma} \tag{2.1}$$

where $2 \leq \gamma \leq 3$. The network model following Power law was called as scale-free.

The general measures that are used to characterize a network include diameter, degree, clustering coefficient and betweenness. The diameter, *d*, of a network is calculated

by first finding the shortest path (smallest number of links) between each pair of nodes; *d* is then the maximum of all shortest paths.

Degree (*k*) is the number of the interactions that one node has [124]. Karimpour-Fard *et. al.* showed that the connectivity of proteins is a strong indicator for essentiality in *E. coli* co-conserved protein interaction network, and is a useful tool for protein function prediction, hence can be used to infer function for uncharacterized proteins [16]. The average degree <*k*> for a network of *N* nodes and *l* edges is defined as:

$$< k > = \frac{2l}{N} \tag{2.2}$$

The clustering coefficient (C_i) indicates the degree to which *k* neighbors of a particular node are connected to each other. This measure provides information on how the neighbors are interconnected [124]. The clustering coefficient C_i is:

$$C_i = \frac{2l_i}{k_i(k_i - 1)} \tag{2.3}$$

where k_i is the number of neighbors connected to a particular node *i* and l_i is the number of existing linkages between neighbors.

Girvan and Newman first proposed the concept of edge betweenness in the context of network communities. The idea is that inter-community edges are more likely to be on some shortest paths than intra-community edges. By computing the all-against-all shortest paths of a graph and calculating the number of times each edge is traveled, one could identify the linkers between communities [11].

Betweenness (b_m) accounts for direct and indirect influences of proteins at distant network sites and hence it allows one to relate local network structure to global network topology. Specifically, betweenness is defined for the m^{th} protein in the network as:

$$b_m = \sum_{i \neq j} \frac{\Gamma(i, m, j)}{\Gamma(i, j)} \tag{2.4}$$

where $\Gamma(i,m,j)$ is the number of shortest paths between i^{th} and j^{th} nodes that passes through m^{th} node and $\Gamma(i,j)$ is the total number of shortest paths among them.

Joy *et. al.* reported the prevalence of low connectivity-degree nodes with high-betweenness values and the existence of such proteins points to the presence of modularity in the network, and suggests that these proteins may represent important connectors that link these putative modules [125].

2.4.3. Hub Proteins in Biological Networks

The protein–protein interaction (PPI) network has a small number of highly connected protein nodes, which are also known as hubs. These proteins participate in significant numbers of protein interactions and play critical roles in the organization and function of cellular protein interaction networks. Due to their pivotal roles in the biological networks, these highly connected proteins are shown to be evolutionary conserved, essential and attractive drug targets [119, 126-128]. For instance, computational analysis shows that removing hubs increases the proportion of unreachable pairs of nodes and the mean shortest path length between all pairs of reachable nodes in the network (i.e., network diameter) more than removing non-hubs [129]. Hence, hubs are more important than non-hubs to the maintenance of the global network structure.

He *et. al.* discussed the essentiality of hub proteins. Their study indicated that, the essentiality of a hub protein is strongly associated with the essentiality of the interaction and revealed that essential protein interactions are likely to occupy central locations in the network. It was concluded that although gene essentiality is an important phenomenon due to its determining capability of the organism's survival and reproduction, the significance of the network architecture may lie in other aspects of the cellular life, such as the interaction itself [126].

Hsing *et. al.* developed a hub protein classifier based on interaction data and GO annotations. The proposed algorithm in this study predicts the "hubness" of a protein in a protein interaction network. They have tested their algorithm on four different species. The

study demonstrates that highly-connected proteins in the protein interaction networks share certain GO annotations, hence the successful hub classifier was achieved [127].

Generally, high degree proteins are considered as hub proteins in the network [120], however selection criterion solely on connectivity of a protein might be inadequate to emphasize their abilities in maintaining robustness. Betweenness measures the centrality of a protein with respect to other, high betweenness proteins are more central and maintain the communication within the network; hence proteins with high betweenness are also accepted as hubs [125]. Since only a small fraction of genes are responsible for survival and reproduction, the removal of these proteins has more impact on the network structure. These proteins are known as essential proteins, and deletion of the genes encoding these proteins are more likely to be lethal. These essential proteins are accepted as hubs in proteins interaction networks [126]. According to their central roles, removal of a hub protein leads to dismantling of the network. Subsequent removal of hub proteins in the network has been shown to dismantle the network into fragments more than non-hub proteins [129].

2.4.4. Functional Module Enumeration Algorithms

The search for functional module derivation rooted back to the study by Spirin and Mirny with the introduction of modularity term. Modularity, *Q*, is defined as the ratio of the existing interactions within a cluster to the maximum allowable links [130]. The definition of modularity has directed the research from being focused on individual genes to the search for complex sets of molecules. Considering the fact that biological function is a complicated consequence of the action of a large number of different molecules that interact in many different ways; the derivation for complex set of proteins in a biological system is inevitable. These entities can be called as modules, where the members contribute and participate to specific function. The interactions and interconnections of these building blocks can be investigated with integrative computational tools providing an analytic approach to biological systems. The identification of the functional modules is based on a fundamental idea that the function of these discrete entities is separable from each other [131]. This notion introduces a critical concept that can be used by *in silico* studies focusing on the elucidation of functional modules from complex biological

systems. Until recently tremendous amounts of algorithms were proposed that seek functional modules hindered in complex networks by integrating various types of omics data, where the integration of expression profiles are in abundance.

One of the hallmarks of research addressing the modular architecture of communities is the introduction of global network topology term “edge betweenness” to the search algorithms. Girvan and Newman (GN) have introduced this divisive algorithm where the selection of the edges to be cut is based on the value of their “edge betweenness”. Consider the shortest paths between all pairs of nodes in a network. The betweenness of an edge is the number of these paths running through it. It is clear that, when a graph is made of tightly bound clusters, loosely interconnected; all shortest paths between nodes in different clusters have to go through the few intercluster connections, which therefore have a large betweenness value. The single step of the GN detection algorithm consists in the computation of the edge betweenness for all edges in the graph and in the removal of those with the highest score. The iteration of this procedure leads to the splitting of the network into disconnected subgraphs that in their turn undergo the same procedure, until the whole graph is divided in a set of isolated nodes. In this way the dendogram is built from the root to the leaves [11].

Co-clustering algorithm uses a distance function that utilizes similarity in the expression patterns for scoring the interaction between proteins. The distance metric is defined as one minus the Pearson correlation coefficient between the expression profiles. Hence, the more similarity in the expression profiles, the least is the distance between to interacting profiles. The algorithm presented in this study utilizes standard hierarchical clustering methods where the proteins are assigned to clusters according to the distance metric [132].

Another well-known algorithm in the field is presented by Rives and Galitski, which is based on one of the common network property; shortest path length. The methodology proposed in their study is based on (i) the shortest path between any two vertices is likely to be the most relevant one for functional associations and information transmission; (ii) each vertex in a network has a unique profile of shortest-path distances through the network to every other vertex; and (iii) module co-members are likely to have similar

(clustered) shortest-path-distance profiles. Each distance in the all-pairs-shortest-path matrix was transformed into an "association," defined as $1/d^2$, where d is the shortest-path distance. This transformation emphasizes local associations (short paths) in the subsequent clustering. The application of the method on yeast signaling network composed of 4079 proteins and 6761 interactions. The clustering algorithm yielded well known signaling pathways, as well as the interconnection of the modules was discovered for yeast filamentation through interacting components [133].

MCODE, Molecular Complex Research seeks for protein complexes that are densely connected in an interaction network based on a vertex weighting by local neighborhood density. [134]. The algorithm offers a fast network clustering method, which operates in three stages, vertex weighting, complex prediction and optionally post-processing to filter or add proteins in the resulting complexes by certain connectivity criteria. It uses the clustering coefficient as the basis to enumerate the modules.

SPIN, Search for Pairwise Interactions seeks for the functional modules that are co-regulated through transcriptional mechanisms, since the coexistence of biochemical components necessary for coordination by other mechanisms. The proposed algorithm examines genes from a pair of functional modules, and uses expression data to assess the possibility that a given TF(s) transcriptionally co-regulates genes in both modules. Their results indicated that processes related to storage and transmission of genetic information, such as mRNA, Cell cycle, tRNA, Nucleotides, Differentiation, and DNA are most highly connected, suggesting that the regulation of these mechanisms is controlled through orchestrated cellular mechanisms [135].

The CFinder algorithm allows elucidating the overlapping groups of densely interconnected nodes in graphs using the clique percolation method. Unlike the other clustering algorithms [136], the algorithm assigns the nodes into different clusters according to the edge weights and generates *k*-cliques. Among the functional enumeration algorithms, CFinder can be used up to a million nodes. The implementation of the method was carried out on the network of gene associations in the yeast genome [137].

Semantic Weights for Module Elucidation algorithm is based on the ranking of proteins according to their weighted neighborhood cohesiveness, i.e. clustering coefficient. The highest ranked nodes are considered as seeds for candidate modules. The algorithm then iterates through the neighborhood of each seed protein, to identify densely connected proteins with high functional similarity. The definition of clustering coefficient is translated into a weighted clustering coefficient, which accounts for the functional similarity for interacting proteins. The method takes advantage of two aspects of functional annotation encoded in Gene Ontology; molecular function and biological process, and combines these with topological properties of the protein network [138].

The weighted yeast protein interaction network was analyzed to derive functional modules with a betweenness-based partitioning algorithm. The proposed algorithm in this study utilizes the edge weights while assigning a node to a module based on the notion that not all interactions are equally important in a network, rather some interactions are used more frequently than others. To overcome the unbalanced partitioning due to betweenness definition in the previous algorithms [11], the “non-redundancy” was accounted on the computation for edge betweenness. The betweenness is the ratio for a number of travels through a particular edge to the shortest paths of all nodes in the network. The proposed algorithm only calculates betweenness when the end nodes are distinct. Therefore, the authors defined the betweenness as the maximum number of non-redundant all nodes shortest paths. The alteration in the calculation kept the original form of the betweenness definition, while preventing unbalanced partitioning. The proposed algorithm was tested on a weighted yeast protein interaction network and the modules enumerated were evaluated for their biological significance and upon the detailed analysis of the modules, the authors claimed to predict protein function as well [15].

Semantic similarity, which uses GO annotations, is another property that was introduced to derive functional modules in a protein interaction network. The proposed algorithm enumerates overlapping modules, where a protein can be assigned to various modules. The complexity of protein interaction networks caused by cross-talk between modules also makes functional module detection challenging. The protein interaction network was integrated with the GO annotations to enhance the modularization of the network. The unweighted protein interaction network was converted to a weighted

interaction network by assigning a similarity score according the similarity of the GO annotation that is shared by interacting partners. The flow-based algorithm first selects a small number of informative proteins, which are representative for modules. Next, it simulates information flow starting from each informative protein through the whole weighted interaction network. The flow then reveals a set of proteins under the influence of the informative protein as a potential functional module. The modules may overlap with each other if two or more informative proteins influence the same proteins. The proposed algorithm was examined on *Saccharomyces cerevisiae* protein interaction network. The interactions between the proteins were weighted with the similarity measure based on GO annotations and localizations extracted from MIPS database. The overlapping modules were enumerated through the proposed algorithm; however the lack of biological significance of the derived modules is a drawback for this algorithm [139].

SCAN, Structural Clustering Algorithm for Networks, is proposed to efficiently extract clusters in a biological network and hubs. The algorithm utilizes a similarity metric that accounts for the shared number of neighbors between two particular nodes. The resulting modules were then validated with GO annotations and manual literature survey [140].

DETMOD, Detect Modules, extracts overlapping modules by starting with a seed node and selects a subset of its promising neighbors, and subsequently expands its kernel, which is a transient state of local environment, to accept more proteins. The neighboring nodes of the seed protein are ranked according to their significance. The selection criteria for module proteins is the density of the kernel and the external links of the module [141].

Dense Module Enumeration Approach is a density-based mining approach is derived from the idea that the node sets with higher density in an interaction network are more likely to represent functional protein complexes. Therefore, by the satisfaction of a density threshold within a cluster, the algorithm extracts modules with high density [142].

Bron-Kerbosch algorithm [22] is a rigorous clique partitioning algorithm that aims to enumerate maximal cliques within a network. Branch-and-bound technique has been used to make the algorithm more efficient by cutting off branches of the search tree that will not

lead to new cliques. The algorithm starts with three disjoint sets of nodes *compsub*, *candidates* and *not*. Set *compsub* represents the growing module, set *candidates* involves the prospective nodes that are connected to all nodes in *compsub*, using which *compsub* can be expanded, and set *not* contains the nodes that are processed. All nodes which are connected to every node of *compsub* are kept either in *candidates* or *not*. The aim of keeping set *not* is to avoid repetition in the calculation of modules. The algorithm checks whether set *not* if it is empty to avoid reporting modules which are not maximal. If *not* is non-empty, the nodes in *not* may be added to *compsub*. The implementation is easy compared to some other clique enumeration algorithms [23]. The algorithm was already applied to various networks ranging from social networks to large scale proteomic networks to find overlapping cliques [24-27].

2.5. Network Approach for Disease Associated Studies

Systems approaches to disease arise from a simple hypothesis- disease emerges from the functioning of one or more disease-perturbed networks (genetic and/or environmental perturbations) that alter the levels of proteins and other informational molecules controlled by these networks. The dynamically changing levels of disease-perturbed proteins (networks) across disease progression presumably explain the mechanisms of the disease. Systems approaches to biology or medicine have two cardinal features: (1) global analyses to generate comprehensive data sets in the disease-relevant organ or cells across the dynamically changing disease process (e.g. all mRNA, miRNA, or protein levels) and (2) the integration of different levels of biological information (DNA, mRNA, miRNA, protein, interactions, metabolites, networks, tissues, organs, and phenotypes) to generate hypotheses about the fundamental principles of the disease [143].

A large number of gene variants are known to cause phenotypic disorders in humans. These disease-causing genes have been identified by linkage studies of the affected families and mutational screening. The identification of disease-causing genes not only facilitates the understanding of the protein function that provide direct insight into the progression of the disease, but also points out potential drug targets for further research [144].

Advances in the field suggest that genetic heterogeneity has its roots at the protein interaction level, suggesting that the genes associated with a similar phenotype have a similar role. Therefore, the phenotypic overlap between the diseases may reflect the relationships and functional properties shared underlying mechanisms.

Accumulation of the high throughput information in publicly available databases enables researchers to study complex problems using computational tools. While there is wealth of protein-disease relationships in literature and a number of predicted protein-protein interaction resources, there has been a growing interest of using protein interactions to decipher the underlying shared biochemical pathways between diseases. Network based approaches utilizing these resources have been widely used for predicting protein function, identifying functional modules and predicting protein interactions, and selection of target proteins within a particular pathway, identification of sequence of protein interaction processes [145-147].

The disease associated studies employing network approach are concentrated on two various fields; disease specific studies focus on a particular disease to identify related proteins and to decipher putative drug targets using network analysis and global search for disease associations through integration of various levels of information.

2.5.1. Network Approach for a Specific Disease

A protein–protein interaction network for the 20 or more different inherited cerebellar ataxias characterized by Purkinje cell (PC) degeneration was developed based on the phenotypic variations. The analysis of the network indicated the high connectivity between different ataxia-causing proteins, revealed common cellular pathways that might lead to PC dysfunction and degeneration and pointed out possible candidate genes for cerebellar ataxias. The ataxia network was built first using yeast-2-hybrid screens and then expanded by incorporating literature-curated and evolutionarily conserved interactions. The analysis of the network in terms of mean path lengths suggested that the mean path length between the disease causing proteins are shorter than non-disease proteins. The proteins involved in RNA binding and splicing have shown to be linked to several disease

causing proteins, suggesting that a subset of inherited ataxias might represent disorders of RNA splicing [8].

The modular architecture of protein interaction network of the human brain was investigated to reveal the significantly altered modular mechanisms during aging. The dynamic approach presented in this study utilizes gene expression profiles to construct protein interaction network during aging where the transcriptome and interactome data were integrated. The suggested algorithm finds a sub-protein interaction network that is active during specific processes. The resulting modules were classified as positively or negatively regulated subgroups depending only on the variations during the process. The network representations suggest that aging attack preferably key regulatory nodes that are important for network stability [6].

The organization of DNA variations related with obesity, diabetes and atherosclerosis-traits in a network representation was resulted in the identification of three genes encoding lipoprotein lipase, lactamase b and protein phosphatase-1 like proteins for their possible roles in obesity. The co-expression network from liver and adipose tissues collected from a mouse population was partitioned in sub-networks that exhibit significant associations with a complex of linked genetic loci related to obesity, diabetes and atherosclerosis traits [148].

Watkinson *et. al.* analyzed two sets of microarray data, one in the presence and one in the absence of a disease, identifying gene pairs whose correlation with disease is due to cooperative, rather than independent, contributions of genes, using the recently developed information theoretic measure of synergy. High levels of synergy in gene pairs indicates possible membership of the two genes in a shared pathway and leads to a graphical representation of inferred gene-gene interactions associated with disease, in the form of a "synergy network." They have applied this approach to prostate cancer expression data and revealed the significant association of oxidative damage to disease. Retinol binding protein 1 (RBP1) and prostaglandin D-synthase (PTGDS) were suggested to play central role in the pathways combining oxidative stress and prostate cancer [149].

Hwang *et. al.* constructed a protein interaction network associated with asthma and analyzed the network in terms of global properties to determine the hub proteins, which are considered to play important roles in the pathogenesis of the disease. The network approach and hub protein identification in their study revealed numerous target genes involved in cellular signaling, such as guanine nucleotide binding protein (GNB2L1) [5].

The construction of a protein interaction network was accomplished using co-expression profiles of genes involved in Alzheimer's disease, which is considered as a polygenic neurodegenerative disease affecting brain and the primary cause of dementia. The modules elucidated from the co-expressed set of genes indicated the association of Alzheimer's disease and cardiovascular disease and provided additional evidence that these diseases are linked through shared biochemical pathways [37].

Hwang *et. al.* studied the prion disease, which is a transmissible neurodegenerative disorder, with network approach by constructing a dynamic co-expression networks. These networks were segmented into four clusters representing the fundamental pathological processes evoked during the disease progress. Depending on the dynamic alterations in the clusters, they have identified novel functional modules with concerted changes of expression for closely related genes. These novel modules represent the androgen and arachidonate/prostaglandin metabolisms and iron homeostasis, which were not previously associated with prion disease [150].

Potential drug targets in a disease related protein interaction network was identified according to the topological position of this protein in the network. This was achieved by assigning a significance score to each protein on its role in providing the shortest path connectivity among proteins. The proposed methodology was tested on psoriasis to identify the key mediators. The differentially expressed genes were assembled and scored according to their relevance to the disease phenotype [151].

There are also numerous examples of cancer related protein interaction network pointing several candidate genes involved in various types of the disease. One of the interesting studies analyzes the dynamic modular structure of a protein interaction network related with breast cancer and predicts the outcome of the disease. The changes in the

organization of the interactome were monitored to estimate the patient outcome. The interaction network related with breast cancer was generated and analyzed in terms of global network properties to identify hub proteins. The hub proteins were then classified as intermodular, which displayed low co-expression profiles with their interacting partners and intramodular, which are highly co-expressed with their interacting proteins. The interactome network in the presence of intermodular and intramodular hub proteins was analyzed in terms of betweenness and characteristic path length. The removal of intermodular hub proteins had a substantial effect on the betweenness of the network, indicating that the proteins are involved mostly in information flow through the network. On the other hand, the characteristic path length, which is the shortest path between all nodes in the network, altered upon the removal of intermodular hub proteins and dismantled the network into fragments. The alterations in the network structure due to removal of intermodular hub proteins suggested that contribution of these proteins in the modular structure of the network. Upon the identification and classification of the hub proteins, the modules enumerated were incorporated with the patient outcome as "good" indicating the disappearance of breast cancer and "bad" indicating the death due to disease. The changes in the expression profiles linked to the hub proteins indicated several genes that can be used to predict the cancer outcome [9].

Recent study on psoriasis lesions revealed the similarities and differences between transcriptomics- and proteomics-level perturbations. The network topology was analyzed to identify common regulators for two datasets such as the most influential transcription factors and receptors [152].

2.5.2. Global Search for Disease Associations

As the availability of the data on human increases, the research in this field directed towards global analysis of disease associations to reveal cross-talk between the underlying pathways and mechanisms to reach a comprehensive perspective.

One of the preliminary studies presenting a global perspective is proposed by Sam *et. al.* In their study, human protein interaction network was incorporated with the OMIM records and PhenoGO database and they have calculated the hypergeometric distance

between the diseases in terms of shared proteins and obtained a significant correlation between xeroderma pigmentosum and Cockayne syndrome [153].

Lage *et. al.* investigated the protein complexes to construct a phenome-interactome network. Their strategy is based on the assumption that variations among different members of a protein complex lead to comparable phenotypes. They have ranked human protein complexes using a phenotype similarity score to identify novel candidates implicated in retinitis pigmentosa, inflammatory bowel disease amyotrophic lateral sclerosis, Alzheimer's disease, Type 2 diabetes and coronary artery disease [154].

Lee *et. al.* constructed a metabolic a metabolic disease network in which two disorders are connected if they are linked to potentially correlated reactions. We also explore to what degree the predicted relationships between often distinct phenotypes result in detectable co-morbidity patterns in patients. Our results demonstrate that the predicted links among diseases are frequently observed in patients and that the underlying disease network offers insights into the factors contributing to the mortality rate of diseases. Their results indicated that highly connected disease exhibit more co-morbidity and more prevalence in the population [155].

Goh *et. al.* constructed a bipartite network of disorders and disease genes by known disease-gene associations. They have generated a human disease network where the diseases are linked each other through shared genes and disease gene network where the genes are linked to each other if they are associated with the same disorder. The implications of diseases and genes are recruited from OMIM Morbid Map, although it offers a comprehensive resource, there are still missing associations, such as the missing link between insulin and Type 2 diabetes. Their study demonstrates that vast majority of disease genes are not essential and exhibit no tendency to act as hubs [1].

Park *et. al* converted medical records to network entities by integrating morbidity information with co-expression patterns and protein interaction networks and allowing to uncover the correlation in the occurrence of various diseases. Their study presents a promising implementation of network biology in health care. In their study Alzheimer's - myocardial infarction and autonomous nervous system disorder and carpal tunnel

syndrome exhibit significant co-morbidity that are connected through shared genes or protein interactions [156].

Li and Agarwal presented a computational approach to studying disease relationships through i) systematic identification of disease associated genes by literature mining, ii) associating diseases to biological pathways where disease genes are enriched, and iii) linking diseases together based on shared pathways. The disease network consists of 591 diseases and 6931 disease relationships. The diseases are linked to each other through common pathways. Their study revealed FOSB pathway is shared by drug-induced dyskinesia and amyotrophic lateral sclerosis [157].

3. METHODOLOGY

3. 1. Construction of Functional Linkage Networks

Construction of functional linkage network for a specific disease was started with the initial number of proteins encoded by the genes which were reported to be associated with the disease. The source of the proteins is mostly literature, the genes reported to be involved with the disease through genome-wide association studies. Addition to the publications, National Center for Biotechology Information (NCBI) database was searched for the specific disease term, such as “cardiovascular disease”, and the resulting genes are included. For the Alzheimer’s disease, due to the abundance of the resources, various disease associated databases were also considered. The initial collection of proteins is referred as core proteins, or c-proteins, which have previously associated with the disease of interest.

To establish the functional links between the proteins, STRING database v7.1 for cardiovascular disease and v8.1 for Type 2 diabetes and Alzheimer’s disease were utilized. Rather than using physical evidence of protein interactions, which could be obtained through records deposited for yeast-2-hybrid experiments, the preference of linkages was to use functionality, since modular approaches based solely on physical protein interactions generally yielded protein complexes. Hence, establishing functional linkages between proteins was anticipated to achieve more biologically relevant structures. STRING combines available information on protein–protein interactions and assigns a confidence score according to variety of the supporting data, including physical interactions, curated biological pathway knowledge, functional linkage, co-expression profiles, as well as the co-occurrences of protein pairs in database text fields and conservation across species [158]. While a confidence score greater than 700 was accepted to be high [159], in this study only the interactions having a confidence score over 900 were assumed to be significant.

To capture other putative proteins that have potential links with a disease, the first neighbors of the core proteins were also extracted from STRING database. All of the

proteins that interact with the core proteins, regardless of the confidence score, were accepted as “candidate neighbors”. Among these candidates, only the ones with a qualified interaction (having a confidence score > 900) were included in the network. The nodes that are not connected to the giant component and their interconnections were excluded, and finally the construction of the network of interest was completed.

To select the confidence threshold for functional linkages to achieve a comprehensive representation of the system under investigation, various disease-related functional linkage networks were constructed with changing confidence score for interactions. Starting from a lenient criteria (900) to a stringent score (990), each network was analyzed in terms of two measures: coverage and constitution. Coverage measure is defined as the fraction of number of core proteins included in the network to the number of proteins initially collected. Constitution measure was also defined as the number of core proteins included in the network to the number of total proteins. The aim was to keep maximum number of core proteins in the network, while considering the fraction of the number of core proteins in the network. Upon the selection of a suitable confidence score, the nodes that are not connected to the giant component of the network were eliminated for computational purposes.

3.2. Network Topological Properties

For all networks under investigation, the average degree (<*k*>), clustering coefficient (C_i) and betweenness (b_i) were calculated as previously described in the literature to provide the global characteristics of the network [125]. The interaction networks are defined as sets of N nodes, representing the proteins and l edges, representing the interactions among them.

(i) Degree (k) is the number of the interactions that one node has [124]. The average degree <*k*> for a network is defined as:

$$< k > = \frac{2l}{N} \tag{2.2}$$

(ii) Clustering coefficient (C_i) is the fraction of the number of existing interactions among the neighbors of a particular node, l_i, to the maximum allowable interactions among them. C_i ranges from 0 to 1, where 0 indicates that the neighbors of a particular node are not connected. This measure provides information on how the neighbors are interconnected [124]. The average clustering coefficient is:

$$< C > = \frac{1}{N} \sum_{i=1}^{N} \frac{2l_i}{k_i(k_i - 1)} \tag{3.1}$$

(iii) Betweenness (b_i) for a node is fraction of the number of shortest paths between two nodes that passes through a particular node to the total number of shortest paths between these two. When betweenness of a node is higher, this particular node is accepted to act as a contact point, hence responsible for the communication within the network [124]. Specifically, betweenness is defined for the m^{th} protein in the network as:

$$b_m = \sum_{i \neq j} \frac{\Gamma(i, m, j)}{\Gamma(i, j)} \tag{2.4}$$

where $\Gamma(i,m,j)$ is the number of shortest paths between i^{th} and j^{th} nodes that passes through m^{th} node and $\Gamma(i,j)$ is the total number of shortest paths among them.

Real biological networks exhibit a scale-free behavior, indicating that many proteins have low number of interactions, whereas fewer proteins have higher degrees of interaction. It is known that the degree distribution for a scale-free network follows the Power law, i.e. $f(k) = Ak^{-\gamma}$, where $f(k)$ is the frequency of nodes, $2 < \gamma < 3$ and A is a constant. Power law degree exponent, γ_c, of cumulative degree distribution, $n(k)$, is related with γ as $\gamma = 1 - \gamma_c$ [124]. For each degree, cumulative degree $n(k)$, clustering coefficient $C(k)$, and betweenness $b(k)$ distributions were calculated for each degree and Power law fitting was done for each of the distribution. To determine the significance of our analysis, 10^4 random networks were generated while keeping the number of nodes and edges constant and cumulative degree $n(k)$, clustering coefficient $C(k)$ distributions were also determined.

The hub proteins which are described as highly connected nodes in a network were selected on the basis of degree and betweenness in the present study. The proteins at the intersection of the top 1–2 per cent highest degree and top 1–2 per cent highest betweenness were considered to be hub proteins. Due to the central roles of hub proteins in the networks, the shared functions exerted by the hub proteins were analyzed using Gene Ontology terms. AmiGO Gene Ontology term enrichment tool [160] was employed to identify the significance of annotation for the hub proteins in the whole genome. UniProt-Swiss keys and EntrezGene IDs were used as identifiers in AmiGO.

3.3. Identification of Functional Modules

A functional module is the highly condensed subgraphs within a network, measured by modularity, *Q*, which is the fraction of existing interactions to the maximum possible interactions and is expressed by, $Q(m,n) = 2m/(n(n\text{-}1))$, where *m* represents the existing links within *n* nodes. For a functional module, all nodes are connected to each other, hence $Q = 1$ [130].

Bron-Kerbosch (BK) algorithm [22] implemented in Python scripting language was used to identify the functional modules in the network. The algorithm allows overlapping modules, where nodes can be allocated in multiple modules. The algorithm produces functional modules starting from two members, i.e. interaction partners. Therefore, according to the average size of the functional modules enumerated, and based on the literature suggesting the biologically relevant modules contain four or more members, the functional modules enumerated were restricted to size. For each of the 10^4 random networks, functional modules were computed upon which a module size distribution was completed. The functional modules that contain four or more members were considered for further investigation. The proteins included in these modules represent the modular structures involved in the disease. For each of the proteins in this modular representation, frequency of proteins in functional modules was calculated as:

$$f_i = \frac{m(i)}{M} \tag{3.2}$$

where f_i is the frequency for the i^{th} protein. $m(i)$ denotes the number of functional modules in which the i^{th} protein is present, M is the total number of functional modules.

3.4. Incorporation of Functional Modules with OMIM Database

A string-match algorithm based on exact match of disease terms with the corresponding disease text field was created to analyze the associated Online Mendelian Inheritance Database (OMIM) disease records and Medical Subject Heading disease classes.

The disease terms was extracted from the Medical Subject Headings (MeSH) database [161]. In this database, the diseases are organized in a tree structure and categorized according to the system that is affected. However, due to the enormous number of disease descriptions and multiple assignments to the classes, the disease terms were eliminated manually and the number of disease terms was reduced to 3630. To assign an appropriate class to a particular disease, on the tree structure, the highest disease class was accepted.

To integrate the genetic information with the diseases, the curated disease terms were linked to the genetic records in OMIM, which is the most updated and complete repository available on human diseases. OMIM database contains record-based textual information, one disorder or one gene deficiency per record. Each record also includes population studies, animal models, allelic variants and references. However, OMIM was originally designed to be read by humans, not by computer, which limits any kind of classification. Therefore, each record was parsed into categories such as, disease ID, disease name, disease text and references. The disease terms and classes that were generated previously were taken as the basis of our analysis.

The description of a disease is a limitation in such text mining studies. There are many words and phrases that characterize a disease; for example Crohn disease was also described as inflammatory bowel disease Type I, or inflammatory bowel disease I. An exact match between the disease terms in our list and the disease text in OMIM was

assumed to be associated. Also, the capitalization of all terms in a disease and the text was removed to overcome mismatches.

There are sentences in OMIM that contain phrases such as “is not related to cardiovascular disease”. Although, the appearance of a disease term may not mean that the gene is related with the disease; it still indicates a persuasive relation. Indeed, an algorithm that recognizes such a negative relation is not available; any kind of association that connects a gene to a disease term was accepted as valid. Also, in some of the records, both negative and positive relations appear simultaneously, these were accepted as valid associations.

3.5. Functional Module-Disease Term Associations

The computational framework used for incorporation of semantic disease information with CFN is presented in Figure 3.1. It consists of four major stages: (i) assembly of a core set of proteins manually extracted from literature and its extension with neighboring proteins to construct a functional linkage network, (ii) formation of protein-disease term linkages by a search algorithm that combines literature and OMIM database, (iii) derivation of functional modules from the functional linkage network, incorporation with the semantic information gathered from OMIM database and deciphering statistically significant disease associations with the functional modules, (iv) assembly of overlapping functional modules that were enriched with disease terms to achieve quantitative valid associations between disorders. In this layout, the databases used in the study were shown in blue boxes. The network representation shows the nodes (proteins) and edges (interactions). The protein-disease connectivity matrix is constructed by combining disease terms with their corresponding OMIM record. Module-protein connectivity matrix depicts the members in a functional module. The functional modules and statistically significant associated disease terms were incorporated in a global disease map.

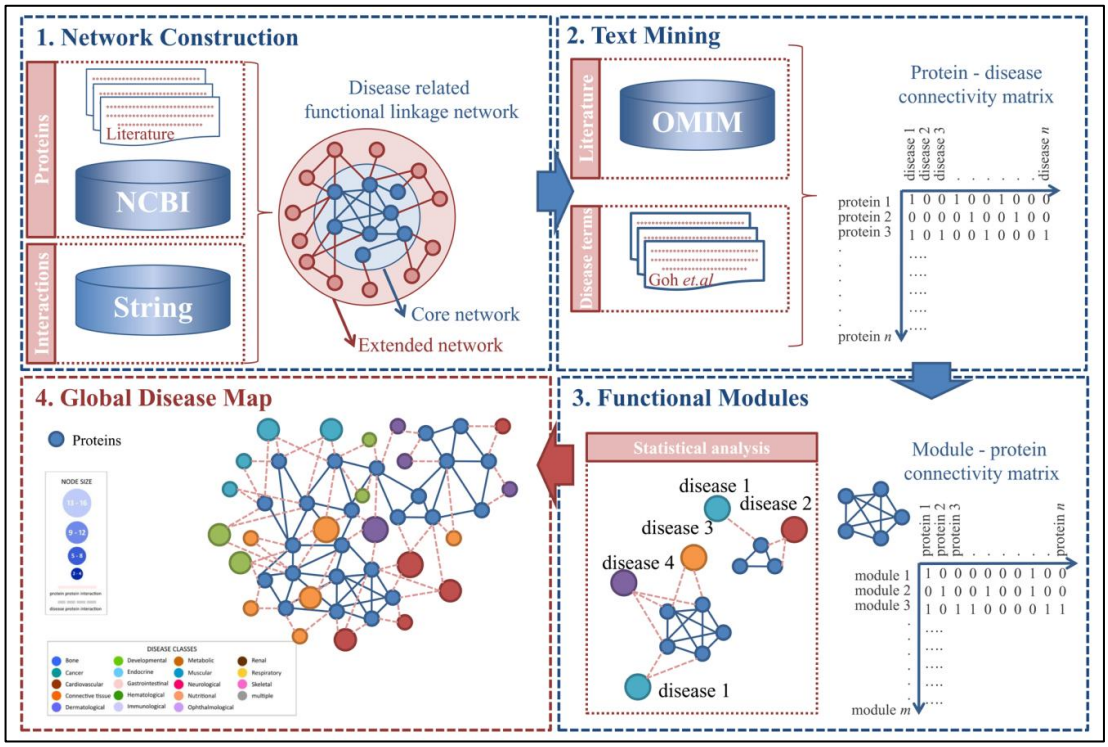

Figure 3.1. Computational framework of the study integrating CFN with bibliomics

The functional module-disease term associations were calculated by simple matrix multiplication of functional module-gene and gene-disease term matrices. This resulting matrix contains information about the number of disease terms associated with a functional module. Since, the algorithm is solely based on a basic string match, to determine the significant disease terms enriched within functional modules, gene-disease term pairwise relations were shuffled 10^4 times and for each randomization step, the functional module-disease term association matrix calculation was repeated. This provides information that the relation of a disease term with a functional module is different from an association that is found by randomly. If the probability of finding a functional module-disease term pair is smaller than 5.00E-04, this association was accepted as significant.

The overlapping members of the functional modules, which have significant associations with disease terms, were visualized by Cytoscape v2.5.1 [162]. The disease classes housing the disease terms were colored according to the classification in Medical Subject Headings. The significant associations for the modules were combined in a stratified graph showing protein-protein and protein-disease linkages. To provide a

quantitative measure between disease classes in this stratified graph, a connectivity measure is introduced:

$$cf_i = \frac{1}{K}\sum_i k_i \tag{3.3}$$

where cf_i is the connectivity frequency for the i^{th} disease class. k_i denotes the individual number of linkages for each of the diseases in the i^{th} disease class, and K represents the total number of disease associations present in map.

3.6. Evaluation and Assessment of Functional Modules

The computational framework to evaluate the functional modules enumerated from disease related functional linkage networks constitutes three major stages: (i) construction of a disease related protein interaction network and its extension with neighboring proteins (ii) enumeration of functional modules, scoring these modules for co-occurring KEGG pathway terms, localization information, an integrated disease ontology composed of MeSH terms and OMIM database, co-expression patterns and evaluation of these modules with Genetic Algorithm. (iii) Assembly of the high scoring modules and calculation of disease overlapping scores. The computational framework of this study is presented in Figure 3.2. The databases used in the study were shown in gray boxes. The network representation shows the nodes (proteins) and edges (interactions).

For each protein in the functional modules metabolic pathways (KEGG) [163], mammalian protein localization (LOCATE) [164] and a manually derived classification generated by merging Medical Subject Headings (MeSH) [161] and Online Mendelian Inheritance of Man (OMIM) [165] information was obtained. KEGG pathway database was used to associate biological pathways, involving 338 diverse pathways including disease pathways. LOCATE database provides the cellular localization of the proteins in 30 different compartments. Manual curation of MeSH disease terms yielded 3630 disease terms in 23 different disease categories. The disease terms were then matched with corresponding gene record in OMIM. To determine the consistency in a module, a score to

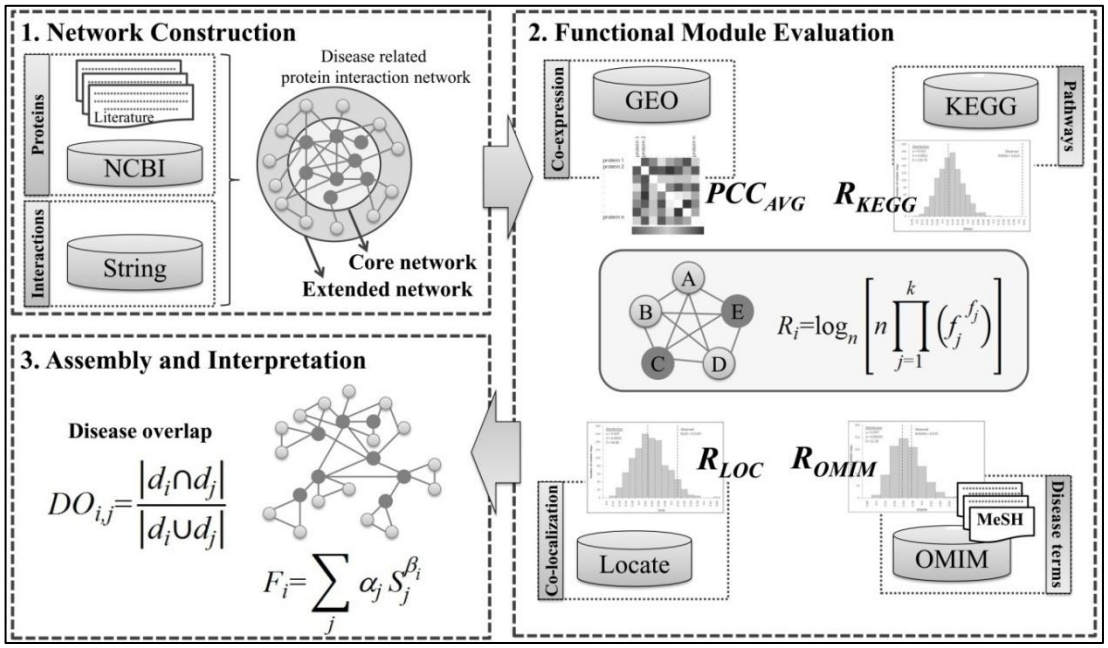

Figure 3.2. Computational framework of the study for evaluation of scoring functional modules

each of the functional modules reflecting the homogeneity of the cluster by calculating the redundancy, R_i:

$$R_i=\log_n\left[n\prod_{k=1}^{n}\left(f_k^{f_k}\right)\right] \tag{3.4}$$

where f_k represents the relative frequency of the class in cluster i and n is the total number of classes in the classification scheme. These scores range between 0 and 1, where 1 indicates all the members in the functional module belong to the same classification. To assess the significance of our results, 10^3 randomized classification schemes were generated.

The co-expression patterns of the functional modules were investigated by calculating the Pearson correlation coefficient ($PCC_{X,Y}$) for each interacting pairs in a functional module. The gene expression datasets were obtained from Gene Expression Omnibus [166]. For each functional module, PCC_{AVG} was calculated over all members of a functional module.

$$PCC_{X,Y}=\frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_i-\mu_X}{\sigma_X}\right)\left(\frac{Y_i-\mu_Y}{\sigma_Y}\right) \quad (3.5)$$

where the sums are over all the observed (X,Y) points. X_i and Y_i denote the observation, μ_x and μ_y are the average of the arrays to be compared and σ_x and σ_y are the standard deviation.

A non-linear model was proposed to evaluate the functional modules with a single resulting score.

$$F_i=\sum_{j}\alpha_j S_j^{\beta_j} \quad (3.6)$$

where $S_j \in (\log N, R_{KEGG}, R_{LOC}, R_{OMIM}, PCC_{AVG})$, all of which ranges between 0 and 1, 1 indicating consistency in the module; except for N, which denotes the size of the module. α_j and β_j are the nonlinear model coefficients. Genetic Algorithm (GA) was employed to estimate the nonlinear model parameters. GA approach is a population based optimization technique that is designed to search optimum values in a complex space. The nonlinear model parameters were predicted by evolving the population of tentative solutions of the model in the search space. The ten artificially generated functional modules, five of which have the highest score in each scoring scheme, were planted in the population representing the best achievable entity. The population of the modules was evolved for 100 generations. Upon the prediction of the model parameters, these model parameters were used to evaluate the functional modules and the high scoring functional modules were investigated for biological significance.

3.7. Disease Overlaps in Functional Linkage Network

According to the model parameters estimated by GE, the functional modules were ranked according to their score. To determine a cut-off value to select the functional modules providing the best representation of the systems, the members of the functional modules were assembled with various cut-off selections. For instance, following the ranking of the functional modules, descending from highest score to lowest, the proteins

included in the top-25 functional modules and their interactions are assembled together to achieve a condensed representation of the disease. This tentative condensed network was then analyzed in terms of the number of core proteins captured, the number of hub proteins included and the number of clusters observed. The number of core proteins in this condensed network indicates the effect of the reduction in the network. The number of hub proteins is another parameter that should be considered due to their central roles in the networks and their activity in diverse biological processes. The number of clusters formed accounts for the conjunction of the biological processes in the disease. In fact, the presence of many clusters in a condensed network indicate diverse biological functions, however, does not provide information on how these biological processes are connected to each other. Thereby, the proteins in top scoring (25, 50, 75, 100, 150 and 250) functional modules were assembled, and by considering the parameters explained above a cut-off value was determined. These proteins collected in a condensed map were anticipated to represent the fundamental biological processes involved in a disease.

To associate complex diseases with each other, the proteins assembled in a condensed map and their linkages with other proteins were used. The links between the proteins and disease were established previously while MeSH-OMIM disease classification scheme was generated. For instance, the proteins in the condensed network were linked to a disease term. Hence the diseases can be associated with each other through shared proteins. Each pairwise disease association was evaluated in terms of overlapping partners. A score representing the disease overlap (*DO*) was assigned to each pair of disease terms appear in the condensed network using:

$$DO_{i,j} = \frac{\left| d_i \cap d_j \right|}{\left| d_i \cup d_j \right|} \tag{3.7}$$

where d_i and d_j represent the proteins associated with disease term pair. To determine the significance of our analysis, we have performed 10^3 random control runs. The proteins and randomly annotated disease terms were shuffled and overlapping scores were recalculated. The random distribution obtained for each disease term and pathway overlaps were compared with the current score.

4. CARDIOVASCULAR DISEASE RELATED FUNCTIONAL LINKAGE NETWORK

4.1. Construction of Cardiovascular Disease Related Functional Linkage Network

The construction of cardiovascular disease functional linkage network was started with the proteins encoded by 234 genes reported to be associated with CVD. The linkages between the proteins were established by using STRING database v7.1, which combines available information on functional linkages and assigns a confidence score according to the variety of supporting data, including physical interactions, curated biological pathway knowledge, co-expression profiles, as well as the co-occurrences of protein pairs in database text fields and conservation across species [43]. The core set of proteins were incorporated with the first neighbors to capture other proteins that have potential links with the disease. Selection of a stringent threshold confidence score for the linkages enabled us to construct a coherent functional linkage network. The core network was extended with the interacting partners for each of the confidence scores, and the threshold was set according to the coverage of the core proteins. When a very stringent criteria for interactions (confidence score 990) was set, only 133 (35.85 per cent) proteins listed in the core proteins were encapsulated, which might lead to loss of information. When a lenient criteria (confidence score 800) was used as the threshold; 228 (97.44 per cent) core proteins were constituted, addition to 2145 other interacting proteins, which might result in coverage of irrelevant proteins.

Table 4.1. Selection of confidence score for the construction of CFN

Confidence score	**800**	**840**	**860**	**900**	**940**	**960**	**980**	**990**
N	2373	1975	1835	1572	802	687	447	371
l	6065	4631	4159	3374	1043	840	493	358
c-proteins	228	225	225	220	196	181	146	133
<*k*>	2.56	2.34	2.27	2.15	1.30	1.22	1.10	0.96
% constitution	9.61	11.39	12.26	13.99	24.44	26.35	32.66	35.85
% coverage	97.44	96.15	96.15	94.02	83.76	77.35	62.39	56.84

Therefore, the core network was extended with the candidate neighbors having qualified interactions with the core proteins, as well as the interactions among these neighbors. This resulted in a network consisting of 1572 nodes (proteins) and 3374 edges (qualified interactions). Visual inspection of this network revealed a giant component and some other singletons that have no connection with the giant component. Therefore, these orphan nodes (36) and their interconnections (29) were excluded from the network. Finally, the giant component comprising 1536 nodes and 3345 interactions was entitled as cardiovascular disease functional linkage network. Among these 1536 nodes, 206 out of 234 (88.1 per cent) nodes were the members of the core, constituting 13.41 per cent of the extended network.

4.1.1. Analysis of CFN Topological Properties

The topological features of this extended network were analyzed by investigating the degree, *k*, clustering coefficient, *C*, and betweenness, *b*, for each node. The average degree $<k>$, average clustering coefficient $<C>$ and average betweenness $<b>$ for the network were calculated to be 4.35, 0.197, 4569, respectively. For a random network with the same number of nodes and edges, the $<C>$ was calculated as 0.0023. A higher average clustering coefficient than that of a random network is considered to be an indicator of a scale-free network; suggesting that our extended network is scale free. The cumulative degree distribution $n(k)$ (Figure 4.1) and clustering coefficient distribution $C(k)$ (Figure 4.2) with respect to degree distribution, followed the Power law, indicating that many proteins are linked to a few other proteins but only a few of them have many interactions, hence providing additional evidence that the network of interest exhibits a scale-free behavior with degree exponent, $\gamma = 2.28$ ($R^2 = 0.9692$).

The betweenness distribution $b(k)$ was also well-characterized with Power law scaling (Figure 4.3), meaning that many proteins are located at the periphery while a few proteins are located at the center of the network, and hence responsible for communication within the network. The effect of removing the nodes with higher betweenness is similar to that of excluding a hub protein from the network and proteins displaying high betweenness are reported to be likely essential and evolutionary conserved.

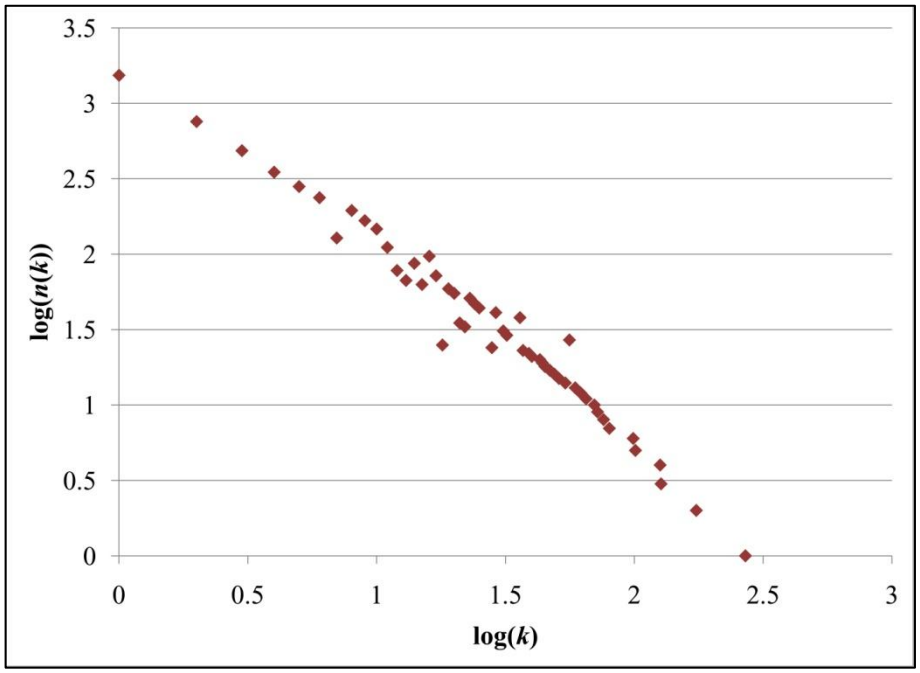

Figure 4.1. The cumulative degree distribution $n(k)$ of CFN with respect to degree distribution

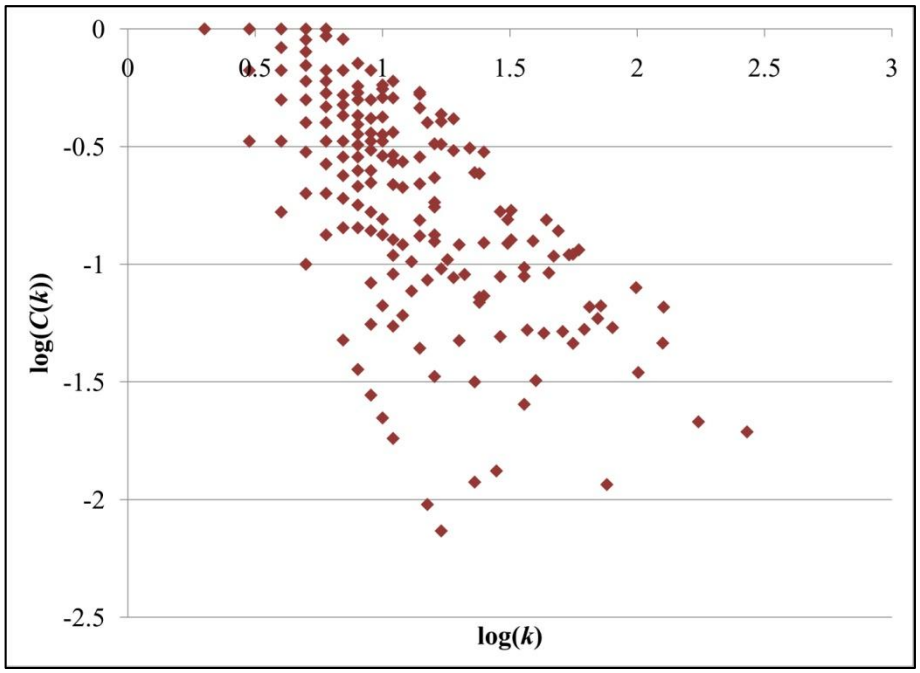

Figure 4.2. Clustering coefficient distribution $C(k)$ with respect to degree distribution in CFN

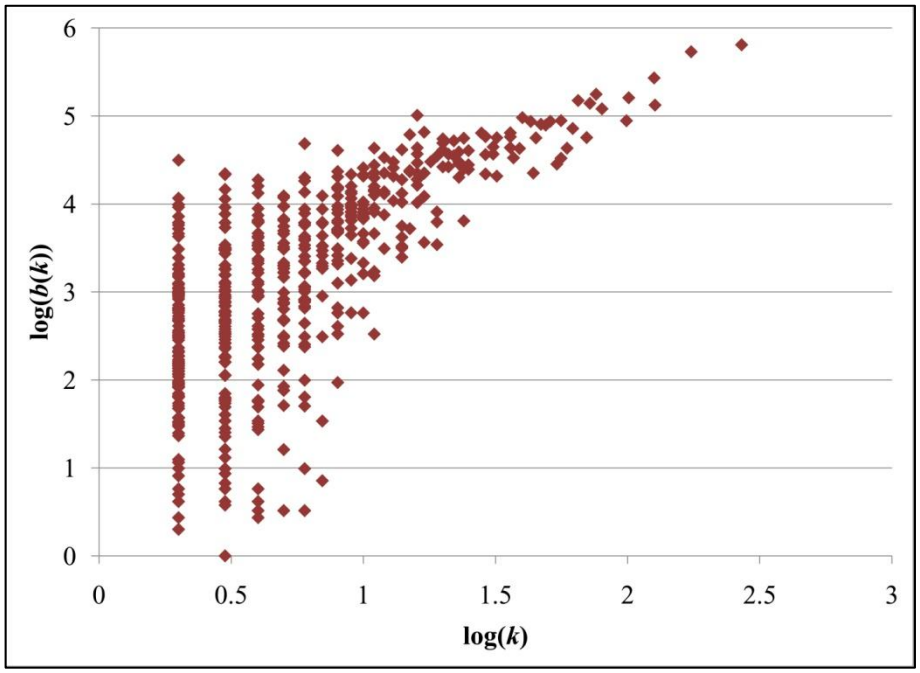

Figure 4.3. Betweenness distribution $b(k)$ with respect to degree distribution in CFN

4.1.2. Selection of Hub Proteins in CFN

Hub proteins corresponding to key nodes in the network are considered to have important roles in biological systems [126, 167]. Therefore, we first determined these key nodes in CFN to identify the underlying molecular mechanism of CVD and its linkage with other complex disorders. Although hub proteins are mostly distinguished by their degree [5, 127], but there is no consensus on how many interactions a hub protein should have. The nodes that are above the position of the sharp turn on the accumulative protein degree distribution or the nodes with half of the maximum degrees are accepted as the criteria for the selection of hub proteins. A hub protein acts like a bridge in the network, since it should have large number of neighbors, but these neighbors should have fewer interactions in between.

In this study, both degree and betweenness measures were taken as the basis in the selection of hubs. The accumulative degree distribution graph shows that 50.65 per cent of the proteins in CFN have single linkage (Figure 4.4), whereas 62.04 per cent of the proteins were located at the periphery (Figure 4.5). Each protein in the network was ranked according to their degree and betweenness, and 15 proteins that were listed in both top 20 highest degree and 20 highest betweenness nodes were selected as hub proteins, all of

which were core proteins. The degree (k) of hub proteins ranges from 47 to 270, with an average of 97, whereas the betweenness (log(b)) ranges from 4.8581 to 5.8096.

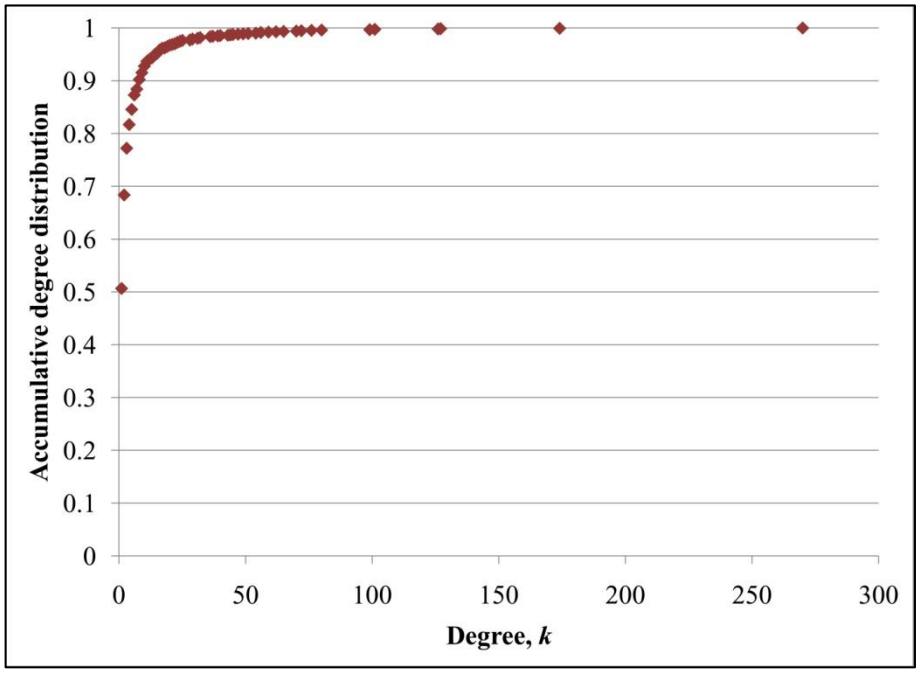

Figure 4.4. Accumulative degree distribution of CFN with respect to degree distribution

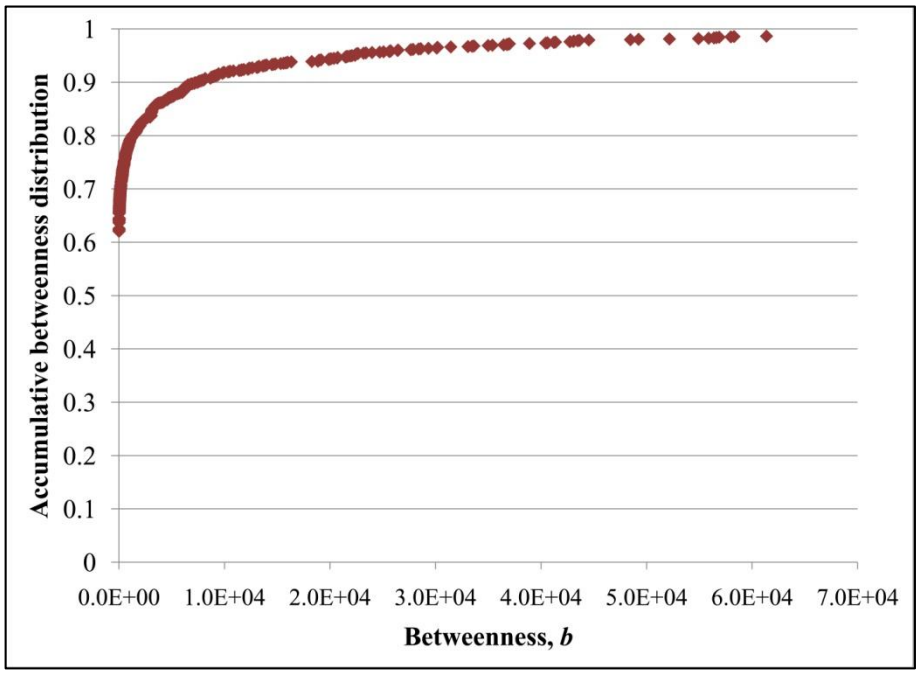

Figure 4.5. Accumulative betweenness distribution of CFN with respect to degree distribution

Table 4.2. Topological properties of hub proteins in CFN

Symbol	**Name**	**Degree, *k***	**Betweenness, log(*b*)**	**Clustering coefficient, *C***
TNFa	tumor necrosis factor	270	5.8096	0.0194
INS	Insulin	174	5.7306	0.0214
IL6	interleukin 6	127	5.1238	0.0657
VEGFA	vascular endothelial growth factor A	126	5.4323	0.0462
STAT3	signal transducer and activator of transcription 3	101	5.2071	0.0347
IL1A	interleukin 1, alpha	99	4.9461	0.0796
PTGS2	prostaglandin-endoperoxide synthase 2	80	5.0818	0.0538
ESR1	estrogen receptor 1	76	5.2472	0.0116
TGFB1	transforming growth factor, beta 1	72	5.1441	0.0665
AGT	Angiotensinogen	65	5.1761	0.0659
IRS1	insulin receptor substrate 1	62	4.8581	0.0529
INSR	insulin receptor	56	4.9472	0.0461
EGR1	early growth response 1	51	4.9396	0.0518
NOS2A	nitric oxide synthase 2A	49	4.8967	0.1386
LEP	Leptin	47	4.9057	0.1082

The 15 hub proteins in CFN have degrees, *k*, ranging from 47 (LEP – leptin) to 270 (TNFa – tumor necrosis factor α), with an average of 97 interactions, whereas the betweenness, log(*b*), ranges from 4.8581 (IRS1 – Insulin receptor 1) to 5.8096 (TNFa) (Table 4.2). Gene Ontology (GO) terms enriched in these 15 hub proteins were determined by AmiGO [160]. The biological process GO categories with significantly enhanced (*p-val* < 5.50E-08) representation in the hub proteins include regulation of response to stimulus, regulation of localization, response to external stimulus, cell surface receptor linked signal transduction and cell communication.

KEGG pathway database was searched to elucidate the function of hub proteins in the biological pathways. The entire network of proteins display an average presence in two biological pathways, however, the hub proteins function in 4.8 biological pathways, on the average. The KEGG pathways were found to be related with several aspects of signaling pathways: MAPK, JAK-STAT, TGF-β, calcium, insulin, adipocytokine signaling pathways, and as well as disease pathways: Asthma, Alzheimer's disease, Prion disease, Type 1 and Type 2 diabetes mellitus, and some cancer variants. The presence of TNFa in 15 pathways ranging from the immune system to signal transduction pathways implies that

this protein is responsible for crosstalk between signaling pathways, in concordance with previous reports.

AmiGO Gene Ontology term enrichment tool [160] was employed to identify the significance of annotation for the hub proteins in the whole genome. UniProt-Swiss keys and EntrezGene IDs were used as identifiers in AmiGO.

Table 4.3. Gene Ontology terms enriched in the hub proteins in CFN

GOTerm	***p-val***	**Proteins**
GO:0048583 regulation of response to stimulus	1.03E-12	TNFa INS IL6 VEGFA PTGS2 TGFB1 AGT NOS2A LEP
GO:0032879 regulation of localization	3.22E-12	TNFa INS IL6 VEGFA IL1A TGFB1 AGT INSR NOS2A
GO:0032101 regulation of response to external stimulus	5.81E-12	INS IL6 VEGFA PTGS2 TGFB1 AGT LEP
GO:0006928 cell motion	1.44E-10	TNFa IL6 VEGFA STAT3 PTGS2 TGFB1 AGT INSR
GO:0051674 localization of cell	1.47E-10	TNFa IL6 VEGFA PTGS2 TGFB1 AGT INSR
GO:0009605 response to external stimulus	2.60E-10	TNFa INS IL6 VEGFA IL1A PTGS2 AGT LEP
GO:0007166 cell surface receptor linked signal transduction	8.17E-10	TNFa INS IL6 VEGFA STAT3 IL1A TGFB1 AGT IRS1 INSR
GO:0007243 protein kinase cascade	1.26E-09	TNFa INS IL6 STAT3 TGFB1 AGT IRS1 INSR
GO:0010646 regulation of cell communication	1.56E-09	TNFa INS IL6 VEGFA ESR1 TGFB1 AGT NOS2A
GO:0007154 cell communication	5.50E-08	TNFa INS IL6 VEGFA STAT3 IL1A ESR1 TGFB1 AGT IRS1 INSR NOS2A

4.1.3. Enumeration of Functional Modules in CFN

Genes that contribute to a common disorder show an increased tendency for their products to interact with each other through protein-protein interactions, tend to participate in similar pathways and share GO terms [5, 37]. Therefore, functional modules/cliques, which are regarded as highly condensed subgraphs, were derived using modularity measure. The modularity, *Q*, is defined as the fraction of existing linkages to the maximum possible interactions in a clique. 2448 functional modules with $Q = 1$ were identified in CFN using the Bron-Kerbosch (BK) algorithm [22]. The algorithm allows the presence of

proteins in overlapping modules, which is a realistic requirement, considering the fact that one protein may participate in many biological processes. Out of 1536 nodes in the network, only 257 (16.73 per cent) were represented in the functional modules. The largest functional module in the dataset comprised eight members and the average number of members in a functional module was calculated as 4.9. Analysis of the module size using 10^4 random networks revealed that modules of size four or greater are statistically significant, in concordance with previous findings [130]; 566 functional modules with four or more members, were considered for further investigation.

The presence of one protein in many functional modules indicates that this particular node highly interacts with its neighbors; therefore loss of any functional activity of this protein will affect a wide range of functional modules. The distribution of the frequency of occurrence of hub proteins in functional modules and the self distribution of each hub protein according to the size of the modules are presented in Figure 4.6 and Figure 4.7. All hub proteins appeared in functional modules. TNFa was found to be present in 343 out of 566 functional modules with maximum frequency of occurrence (f_{TNFa} = 0.606) as expected from its high degree of connectivity. TNFa exists in 107 four member functional modules, 81 five member modules, 102 six member modules, and the rest exists in seven and eight member functional modules.

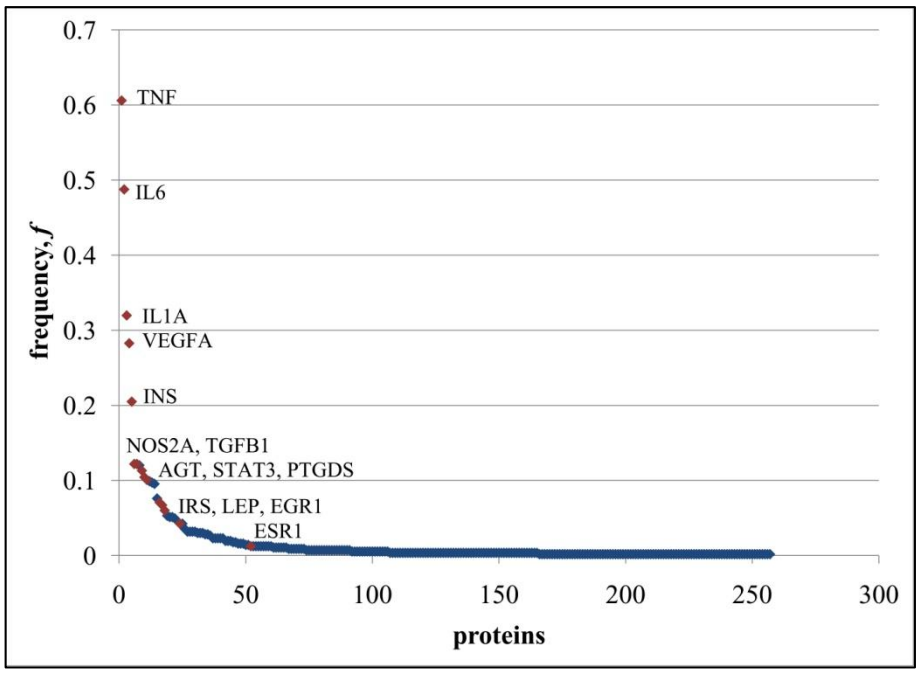

Figure 4.6. Frequency of the presence of the proteins in CFN in functional modules

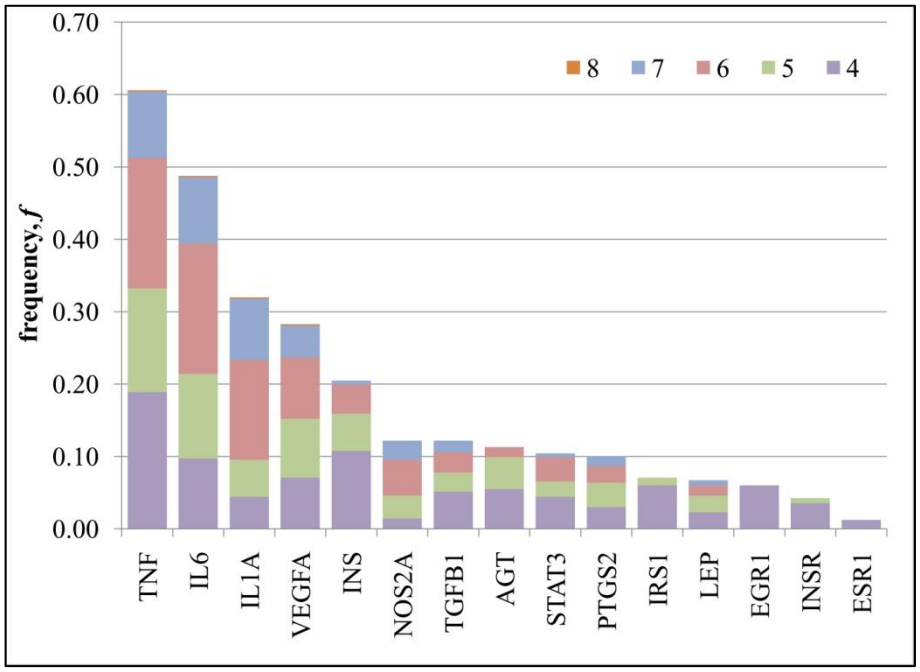

Figure 4.7. Self distribution of hub proteins in CFN according to the size of the functional modules

4.2. Integration of CFN with Semantic Disease Information

The integration of bibliomics with CFN includes the generation of gene-disease association matrix and its conversion into module-disease associations. The general framework of this section is explained in Methodology with a representative schema explaining each step (Figure 3.1).

4.2.1. Disease Associated Functional Modules

In order to delineate common molecular mechanisms and shared genes involved in CVD as well as in other complex disorders, bibliomics data was integrated with functional modules. The 257 proteins present in functional modules were matched with their corresponding disease terms within the OMIM database using a string-match algorithm which was developed in the present study.

The associations between disease terms and genes were identified using 1287 disease terms as described by Goh *et. al.* [1]. Diseases are categorized according to Medical Subject Headings (MeSH) [161]. Only 301 out of 1287 disease terms were found to be

associated with 257 proteins that are present in functional modules. The functional module-disease term associations were identified by simple matrix multiplication of functional module-gene and gene-disease term matrices. In order to investigate the significance of these linkages, the links between genes and disease terms were shuffled randomly 10^4 times while keeping the number of genes in each disease term unchanged. If the probability of finding a functional module-disease term pair is smaller than 5.00E-04, this association was accepted as significant. Finally, among the 566 modules, 227 were found to be significantly associated with one or more disease terms.

19 modules were identified to have significant associations with cardiovascular disease class. 11 proteins out of 41 (IFNG, IL10, IL1B, IL4, JUN, MAPK3, MAPK8, SP1, HIF1A, CXCL12, PROC) identified within these modules were not present in the core network and except PROC they are not reported previously to be associated with CVDs. They are captured by the extension of the core network and present potential targets for further studies.

IFNG, IL10, IL1B, IL4 are the members of T-cell receptor signaling pathway. JUN, MAPK8 (JNK1), MAPK3 (ERK1) are members of MAPK signaling pathway and have important roles in the induction of proliferation, differentiation and inflammation as a response to stress.

Among these proteins which are predicted to have involved in CVD, a hypoxia induced transcription factor 1 (HIF1A) induces vascular endothelial growth factor (VEGF) pathway, as a response to growth factors and hormones such as insulin, in endothelial cells and plays an important role in mammalian cells cultured under reduced oxygen tension and systemic response to hypoxia. A number of complex disorders such as myocardial, cerebral and retinal ischemia, preeclampsia and cancer were reported to be closely associated to the deprivation of oxygen and glucose [168]. Although a number of studies were undertaken to elucidate the function of HIF1A, there is still need to investigate the detailed effect of HIF1A in CVD and other related conditions [169]. Therefore further detailed studies on oxygen homeostasis and HIF1A are required to understand the underlying mechanisms of these pathophysiological processes and its potential therapeutic role.

SP1, which is a transcription factor that is known to play an important role in hemopoiesis, was also predicted to be associated with CVD in the present study. Over-expression of SP1 was reported to mediate inhibition of cell cycle progression that precedes apoptosis [170]. Its expression is induced by hypoxia and suggested as a target for anti-invasion therapies [171]. It is also known that SP1 plays a critical role in the transcriptional regulation of glycolytic enzymes [172, 173] and copper homeostasis [174]. Therefore, modulation of its activity may also be considered as potential therapeutic tool for CVD and other related disorders.

Stromal cell-derived factor 1 (CXCL12), which is member of chemokine signaling pathway was also predicted to be associated with CVD in this study. This pathway is known to be induced by proinflammatory stimuli such as lipopolysaccarides, TNFa or IL1A and results in the NO induction. The key role of CXCL12 in several types of cancer includes the stimulation of CXCR4 leading to invasion of cells through extracellular matrix and attenuation of TNFa and stimulated cell growth [175, 176]. CXCL12 is also involved in platelet aggregation and were reported to be expressed in macrophages in human atherosclerotic plaques [177]. These results indicate the possible role of CXCL12 between chronic inflammation, CVD and cancer.

4.2.2. Functional Modules Associated with More than One Disorders

Two modules (CVD-m501 and CVD-m565) were matched significantly with five disease terms; five modules were significantly associated with four disease terms, 17 were linked with three disease terms, 61 with two disease terms and the rest of these 227 functional modules were associated with single disease term. 85 functional modules, having associated with at least two disease terms was searched for the enriched GO terms, the results indicated that all proteins in these modules are functionally linked (*p-val* < 9.80E-03).

Among the 227 functional modules that are significantly associated disease terms, seven of them were investigated in detail. Disease terms significantly associated with functional modules (*p-val* < 0.50E-03) are presented in Table 4.4. The proteins in a module were separated according to the source as “core” or “extended (extnd)”. The number of

members that are associated with the corresponding disease term was indicated with "fraction" term. Figure 4.8 displays these selected functional modules associated more than one disorders. The protein and disease nodes are depicted with boxes and circles, respectively. Disease nodes are colored according to the class. CVD-m565, CVD-m206, CVD-m77 and CVD-m460 have significant associations with CVD disease class. CVD-m129 is a conjunction connecting Alzheimer's disease, DiGeorge syndrome and rheumatoid arthritis. CVD-m501 and CVDm-289 connects diabetes, hypertriglyceridemia and hypoglycemia, exemplifying the links between the metabolic disorders. It should be mentioned that, SOCS3 present in CVDm-289 has no significant association with these disease terms.

CVD-m565 is a four member module which is found to be associated with five disease terms and consists of Factor 2 (F2), Protein C (PROC), Protein S (PROS1) and Factor V (F5). These proteins are the members of complement-coagulation cascade thus functionally linked. In this module, PROC was not included into the core network and captured upon the extension of the network. This protein is a key component of anticoagulation system which is activated by thrombin (F2) on the surface of epithelial cells. There is accumulating evidence that this protein is an important regulator of microvascular inflammation and elements of the Protein C pathway were found to be down-regulated during inflammation [50]. Protein S is a non-enzymatic cofactor of the activated protein C, which inactivates the procoagulant Factor V and Factor VIII. The proteins encoded by PROC, F2 and F5 were found to be linked with an increased risk for thrombophilia. Thrombophilia is a complex disorder which is recognized by the susceptibility to develop blood clots due to abnormality in the coagulation system [48]. The proteins encoded by F2, PROC and PROS1 in this module were found to be significantly associated with venous thrombosis; as already reported by several investigators [48, 178, 179]. This finding also implies that the obstruction of a coronary artery by a thrombus potentially leads to myocardial infarction. Therefore, the presence of these proteins in CVD-m565 suggests and provides additional evidence that an impaired coagulation system leading to development of thrombus is one of the possible causes of cardiovascular diseases.

Table 4.4. Disease terms significantly associated with functional modules (*p-val* < 5.00E-04)

CVD-m	functional module members	Core	Extnd	Hub	Disease Term	Disease Class	fraction
565	PROC*, F2, PROS1, F5	3	1	0	Angioedema	Cardiovascular	2 / 4
					Epidermolysis bullosa	Congenital	2 / 4
					Protein S deficiency	Hemic and Lymphatic	2 / 4
					Thrombophilia	Hemic and Lymphatic	3 / 4
					Venous thrombosis	Cardiovascular	3 / 4
206	APOB, APOE, LPL, INS, LDLR	5	0	1	Atherosclerosis	Cardiovascular	5 / 5
					Hypercholesterolemia	Nutritional and Metabolic	3 / 5
					Hyperlipoproteinemia	Nutritional and Metabolic	3 / 5
					Hypertriglyceridemia	Nutritional and Metabolic	3 / 5
77	TNF, VCAM1, NOS2A, IL1A, TNFRSF1A, IFNG*	5	1	3	Listeria monocytogenes	Bacterial Infections	3 / 6
					Stroke	Cardiovascular	4 / 6
					Tuberculosis	Bacterial Infections	4 / 6
460	ACE, AGT, AGTR1, NOS3	4	0	1	Hypertension	Cardiovascular	4 / 4
					Renal tubular dysgenesis	Urogenital	3 / 4
129	TNF, IL6, IL1A, VEGFA, TGFB1, FOS*	5	1	5	Alzheimer disease	Nervous System	4 / 6
					DiGeorge syndrome	Congenital	2 / 6
					Rheumatoid arthritis	Musculoskeletal	5 / 6
501	INSR, INS, SLC2A4, LEP	4	0	3	Diabetes mellitus	Nutritional and Metabolic	4 / 4
					Growth retardation	Congenital	3 / 4
					Hypertriglyceridemia	Nutritional and Metabolic	3 / 4
					Hypoglycemia	Nutritional and Metabolic	3 / 4
					Insulin resistance	Nutritional and Metabolic	4 / 4
289	INSR, INS, IRS1, LEP, SOCS3*	4	1	4	Diabetes mellitus	Nutritional and Metabolic	4 / 5
					Hypertriglyceridemia	Nutritional and Metabolic	3 / 5
					Hypoglycemia	Nutritional and Metabolic	3 / 5
	* these proteins are added to the network through the extension with first level of interactors						

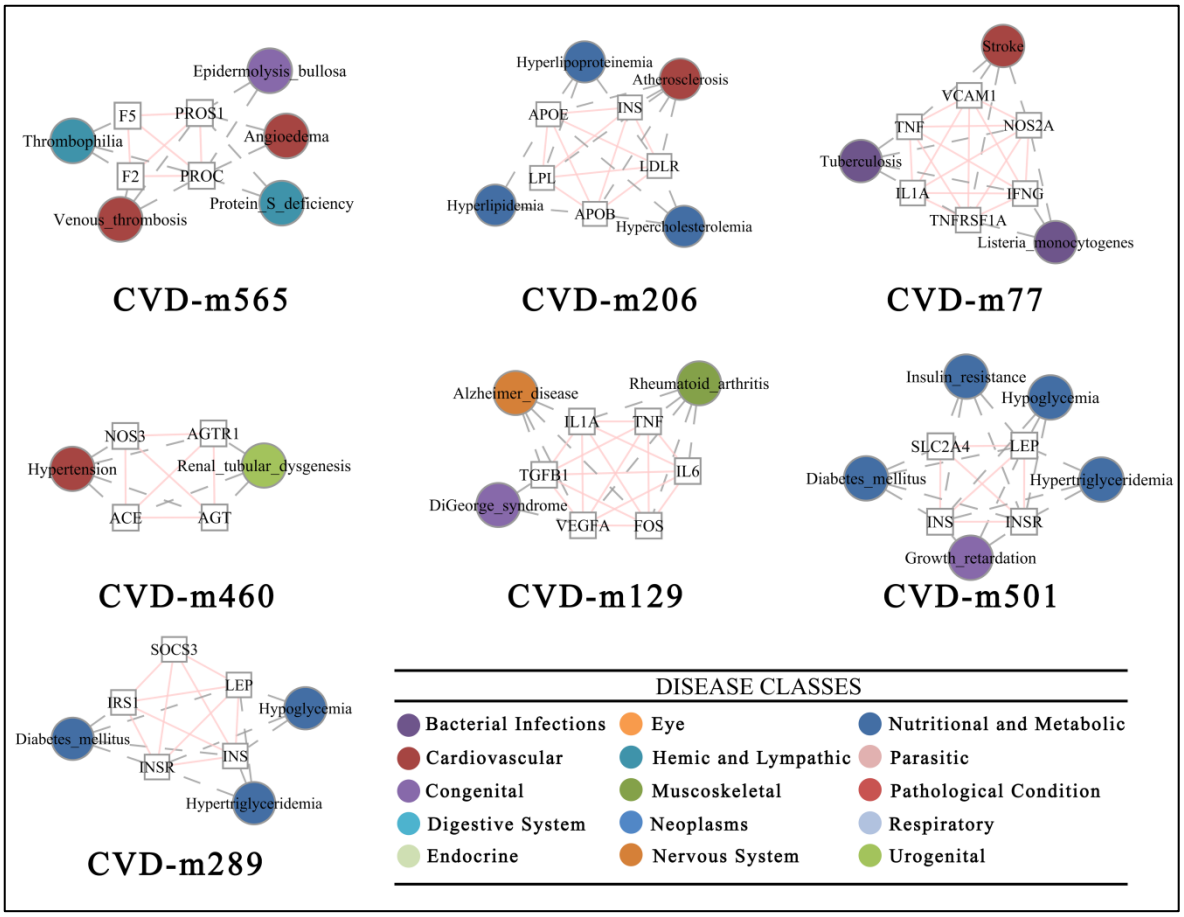

Figure 4.8. Schematic representation of modules associated with disorders

Another disease that is significantly enriched in this module is epidermolysis bullosa, which refers to a group of genetically distinct disorders characterized by the blistering of skin and mucosa. An abnormal increase in fibrinogen, thrombin–antithrombin complex, factor VII, factor VIII, and prothrombin was reported for this disease [180], and development of venous thrombosis was observed in one epidermolysis bullosa patient[181]. Significant association of PROC and PROS1 within this functional module to both epidermolysis bullosa and venous thrombosis is attributed to the presence of common genetic factors shared between these two different groups of disorders.

Angioedema is another disease that is significantly associated with this functional module linked to the members of complement-coagulation cascade. A deficiency in Inhibitor of Complement Component1 Inhibitor (C1-INH also known as SERPING1), which is a member of complement- coagulation cascade, is known to be the underlying cause of the either familial or acquired angioedema. SERPING1 is reported to be important

in the control of vascular homeostasis including inflammation, blood pressure and coagulation [182]. Although SERPING1 was present in CFN, it could not be captured in any functional module due to the low confidence score of its interactions with partners.

All five members of CVD-m206 are associated with atherosclerosis. Any deficiency in lipoprotein lipase (LPL) is reported to be linked with impaired lipolysis. Low density lipoprotein receptor (LDLR) is involved in cholesterol homeostasis [183]. APOE and APOB are the main proteins of the chylomicron and associated with lipoprotein metabolism [70]. Insulin (INS) has an important role in glucose as well as in lipid homeostasis [64]. These reports also provide additional evidence that all proteins in this module are functionally linked to lipoprotein metabolism. Plasma cholesterol and triglycerides are well documented predictors of atherosclerosis, which is the most common cause of myocardial infarction and stroke. The development of atherosclerosis is initiated by the accumulation and subsequent oxidation of low density lipoproteins in arterial cell walls [42]. APOE is linked to all members in this module, indicating that it may be a potential drug target for hypercholesterolemia, hyperlipoproteinemia and hypertriglyceridemia. APOE, which serves as a ligand for LDLR, is one of the principal cholesterol carriers. Studies indicate that statins reduce the expression of APOE, and they are effective at lowering low-density lipoprotein (LDL) cholesterol, leading to a reduced risk for CVD [184, 185].

Tumor necrosis factor (TNFa), nitric oxide synthase (NOS2A) and vascular cell adhesion molecule (VCAM1) in CVD-m77 were found to be significantly associated with stroke. Ischemic stroke is a complex disease consisting of a group of heterogeneous disorders related to multiple genetic and environmental risk groups. It occurs as a result of a disturbance of blood supply to brain due to ischemia, thrombus, embolus or hemorrhage. TNFa, which is secreted by activated monocytes and macrophages, is a multifunctional cytokine with an important role in inflammatory response [55]. It is reported that it is also involved in lipid metabolism, coagulation, insulin resistance and endothelial function [186, 187]. Overproduction of TNFa was observed in autoimmune diseases, osteoporosis, tuberculosis and septic shock. NOS is involved in the synthesis of nitric oxide (NO) which is a messenger molecule that displays many properties of a neurotransmitter. NO is implicated in neurotoxicity associated with stroke and neurodegenerative diseases and

neural regulation of smooth muscle. It is also responsible for the activity of endothelium-derived relaxing factor (EDRF) regulating blood pressure. The role of NOS2A was reported in the development of hypertension [188], which is considered to be one of the important risk factors in stroke. NOS2A is a cytokine inducible form of iNOS which functions as an anti-inflammatory effector molecule in atherosclerosis and vasculopathy [189]. Myocardial iNOS activity is observed after TNFa or IL6 exposure [56, 190]. Increased iNOS expression contributes to myocardial dysfunction and higher mortality after myocardial infarction in mice [191]. VCAM1 is a cell surface glycoprotein expressed by cytokine-activated endothelium and it is involved in the leukocyte-endothelial adhesion and signaling [192]. Increased serum levels of VCAM1 are associated with the recurrent ischemic stroke [193]. Tumor necrosis factor receptor super family 1A (TNFRSF1A) serves as major receptor for TNFa associated with epithelial cells [194]. Poirier *et. al.* reported a predisposition to atherosclerosis and coronary artery complications in individuals with polymorphism R92Q of the tumor necrosis factor receptor 1 gene [195]. Interferon-gamma (IFNG) is a cytokine critical for innate and adaptive immunity against viral and intracellular bacterial infections and for tumor control. Aberrant IFNG expression is found to be associated with several autoinflammatory and autoimmune diseases [2, 42]. Listeria monocytogenes and tuberculosis, which are classified as respiratory disorders, were also found to be significantly associated with this module possibly due to the presence of three hub proteins. NOS2A in this module might play a key role, since macrophage production of nitric oxides was reported to be crucial in host defense against a wide variety of pathogens including *Mycobacterium tuberculosis*, and Listeria monocytogenes [196].

The proteins located in the CVD-m460, namely ACE, AGT, AGTR1 and NOS3 are the members of renin-angiotensin system. Angiotensinogen (AGT) is cleaved by renin, producing angiotensin I, which in turn is converted to angiotensin II by angiotensin converting enzyme (ACE). The activation of Angiotensin II is achieved upon its binding to the angiotensin receptor (AGTR1). The finding that hypertension is linked to the members in this module is anticipated, since ACE, AGT and AGTR1 function together in blood pressure regulation, as also suggested [197]. On the other hand, endothelial nitric oxide, NOS3, is classified as a vasodilator, which is involved in the maintenance of vascular homeostasis, whereas Angiotensin II is a vasoconstrictor. The balance between

Angiotensin II and NOS3 is crucial to healthy endothelium. The elevated levels of Angiotensin II or decreased levels of NOS3 lead to dysfunction of endothelial cells. The imbalance can cause many of the changes in the endothelium that stimulates atherosclerotic process [198]. Another disease that is enriched in this module is renal tubular dysgenesis (RTD). RTD is a congenital disorder of renal tubular development in the neonatal period. The existence of RTD in this module is attributed to the expression of renin-angiotensin system components during early stages of kidney development. This connection is also supported by the studies signifying the inactive or suppressed renin angiotensin system leading to low perfusion pressure, resulted in autosomal recessive RTD [51]. Additionally, unilateral RTD leading to hypertension was recently observed in one neonate with cardiac failure [199].

Four members of CVD-m129 were found to be significantly associated with Alzheimer's disease, five members with rheumatoid arthritis and two with DiGeorge syndrome, the diversity of which is explained by the presence of five hub proteins in this functional module. All proteins within this module are functionally linked by the term signal to external stimulus and cell surface receptor linked signal transduction (*p-val* < 5.30E-05). FOS genes encode a leucine zipper protein which can dimerize with JUN family forming a transcription factor complex AP-1. These proteins have regulatory roles in cell proliferation, differentiation, and transformation. Expression of FOS was found to be associated with apoptosis [200]. VEGFA is an endothelial mitogen which is involved in vascular permeability, angiogenesis, vasculogenesis and also in the inhibition of apoptosis [201]. Transforming growth factor 1 (TGFB1) is a multifunctional peptide which is involved in the control of cell proliferation, differentiation as well as in apoptosis [202]. In the presence of IL6, TGFB1was reported to induce differentiation of T-cells into proinflammatory cytokine producing T-cells which promote inflammation and autoimmunity [203]. TNFa gene encodes a multifunctional proinflammatory cytokine that belongs to the tumor necrosis factor (TNF) superfamily. It is involved in the regulation of a wide spectrum of biological processes including cell proliferation, differentiation, apoptosis, lipid metabolism, and coagulation. This well studied cytokine has been implicated in a variety of diseases, including autoimmune diseases, insulin resistance, and cancer [204]. Knockout studies in mice also suggested the neuroprotective function of this cytokine [205]. IL1A operates through a mechanism involving the RNA processing

apparatus and the alternate splicing of apoptosis regulatory proteins, hence induces apoptosis [206]

Other functional modules determined in the present study displayed relations among other complex disorders which may also be associated with cardiovascular diseases. Members in CVD-m501, namely Insulin Receptor (INSR), Insulin (INS), Solute Carrier Family 2, Member 4 (SLC2A4) and Leptin (LEP) were found to be significantly related with diabetes mellitus, growth retardation, hypertriglyceridemia, hypoglycemia and insulin resistance. Three of these four proteins were also identified as hub proteins. The members of this module are functionally linked to glucose homeostasis. The role of INS and its receptor INSR in glucose homeostasis is well documented [207]. Mutations in SLC2A4, which is also known as glucose transporter 4 (GLUT4), were reported to be associated with noninsulin-dependent diabetes mellitus (NDDM) [208], and the growth retardation in Type 1 diabetes [209]. LEP is known to be pleitropic molecule involved in several physiologic and pathological processes such as T-cell mediated liver toxicity in association with regulatory effects on thymus and peripheral blood cellularity as well as on the production of two proinflammatory cytokines TNFa and IL8 [210]. C-reactive protein (CRP), which is considered to be a good predictor of CVD, is an inhibitor of leptin binding to its receptor, hence blocks signaling [211]. On the other hand, insulin is a potent activator of leptin mRNA expression and protein secretion, possible crosstalk between insulin and leptin signaling pathways was proposed [212].

In a five member functional module CVD-m289, three members, namely, INS, LEP, INSR, which are also members of the CVD-m501, and insulin receptor substrate 1 (IRS1) are involved in glucose homeostasis. These four members are significantly associated with disease terms hypoglycemia, hypertriglyceridemia and diabetes mellitus. The presence of SOCS3 in this module was attributed to the formation of a link between cardiovascular and diabetes. Although no significant disease term was matched for Suppressor of Cytokine Signaling 3 (SOCS3), this protein is involved in feedback inhibition of a range of cytokine signals. It has been reported that SOCS3, which is expressed by terminal transcription factors, such as STAT3 and c-fos, inhibits the function of leptin and downstream steps in insulin pathway [213]. The genes that are up-regulated by JAK-STAT pathway through STAT3 following cardiovascular injury, are involved in cardioprotection, hypertrophy, and

angiogenesis [214]. SOCS proteins are natural inhibitors of STATs and SOCS3 was proposed to be one of the therapeutic targets for obesity and insulin resistance [65]. Further studies are required to determine the therapeutic potential of SOCS3 in decreasing the CVD risk.

4.2.3. Connectivity Map

In order to elucidate the shared mechanisms and genes by co-occurring complex disorders, 227 functional modules comprising shared proteins and significantly associated disease terms were assembled in a map (Figure 4.9) depicting a comprehensive connectivity map between genes and diseases (CFN-G graph). In this connectivity graph, 152 proteins and 58 disease terms in 15 disease classes are represented as nodes; 276 disease-protein associations and 703 functional linkages were indicated. Among those 152 proteins, 67 of them were already present in the core network and the remaining 85 proteins were found to be functionally linked as first neighbors and represent potential target genes encoding proteins shared by CVD and other disorders by common mechanisms. All hubs proteins identified in CFN were present in this CFN-G graph. In this map, disease node sizes were adjusted proportional to the number of proteins that was found to be related with and nodes are colored according to the disease class. Pink lines represent the protein-protein interactions, gray dashed lines were the associations between protein and disease terms. Nodes were located at the closest proximity to neighboring nodes. The minor cluster, which does not interfere with the main component, represents the complement-coagulation cascade. All members except SERPINA5 are functionally enriched in blood coagulation (*p-val* = 3.55E-20). Diseases observed in this cluster mostly belong to hematological diseases, as well as venous thrombosis, epidermolysis bullosa, angioedema and arthropathy. The major cluster clearly separated by lipid metabolic process (*p-val* = 2.31E-05). The close proximity of atherosclerosis to the lipoprotein mechanism is explained by the accumulation of low-density lipoproteins during early stages of disease. The bridge formed by the hub proteins TNFa, INS, INSR and LEP connects the lipoprotein metabolism to the rest of the map. Obesity, diabetes mellitus and insulin resistance are clustered in a region. The proteins mostly involved in renin-angiotensin system are clustered around hypertension. Co-existence of AGT, IL6, TNFa and VEGFA in six functional modules establishes links between inflammation,

hypertension and atherosclerosis. Proinflammatory cytokines and the members of Toll-like receptor signaling pathway are located in the vicinity of immunological disorders.

Among those 58 diseases present in this graph; several cancer types (adenocarcinoma, adrenocortical carcinoma, cervical carcinoma, hemangioma, Kaposi sarcoma), three diseases caused by bacterial infections (meningococcal disease, leprosy, listeria monocytogenes); two hemic disorders (hemorrhagic diathesis, Protein S deficiency); one congenital (epidermolysis bullosa); one eye disease (macular dystrophy) and one nervous system disease (myasthenia) were not previously reported to be linked with CVD .

Connectivity frequency for each disease class was calculated as a ratio of proteins connected to a specific disease class to the total number of proteins linked to a disease term in CFN-G. Figure 4.10 shows the connection frequency of 15 disease classes hosting 58 diseases in CFN-G. Connectivity frequency provides a quantitative measure on the interconnection of disease classes within CFN-G. The nutritional and metabolic disease class, with a connectivity frequency of 0.27, including diabetes mellitus, obesity, insulin resistance, hypertriglyceridemia, hypoglycemia, hyperlipidemia, hyperlipoproteinemia and hypercholesterolemia, is a highly connected disease class in CFN-G map. This result signifies the fact that any imbalance in lipid metabolism is the underlying cause of many complex disorders including cardiovascular disorders.

The third class represents the muscoskeletal diseases with a connectivity frequency of 0.09. Arthropathy, osteoporosis, rheumatoid arthritis and short stature belong to muscoskeletal diseases that affect joints, internal organs and connective tissue. Immunological disorders display a connectivity frequency of 0.079 in the CFN-G. The disorders caused by bacterial infections, such as sepsis, leprosy, Crohn disease indicate the contribution of inflammation to the disease.

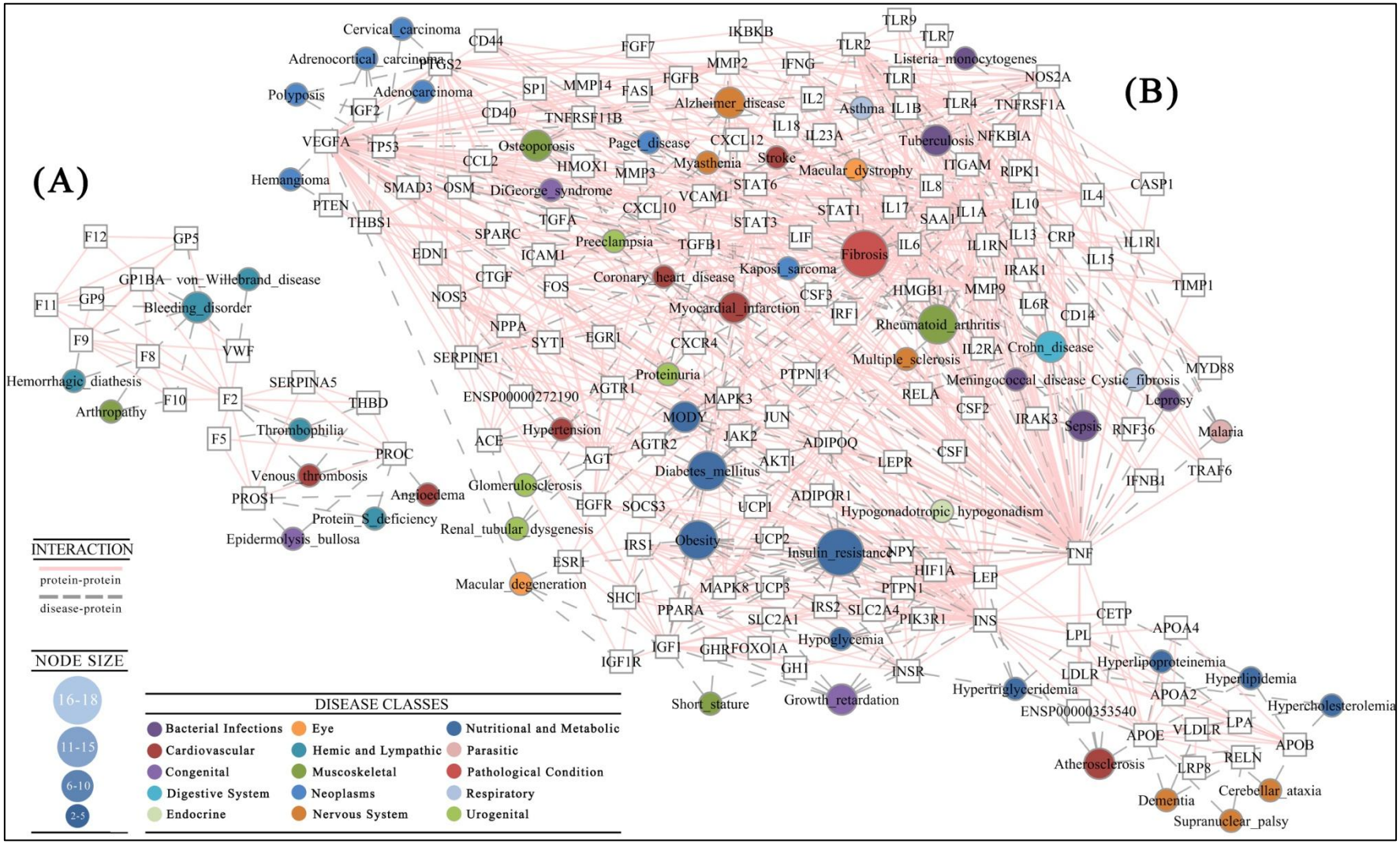

Figure 4.9. Complete CFN-G graph as the combination of 227 functional modules that are found to be associated with disease terms

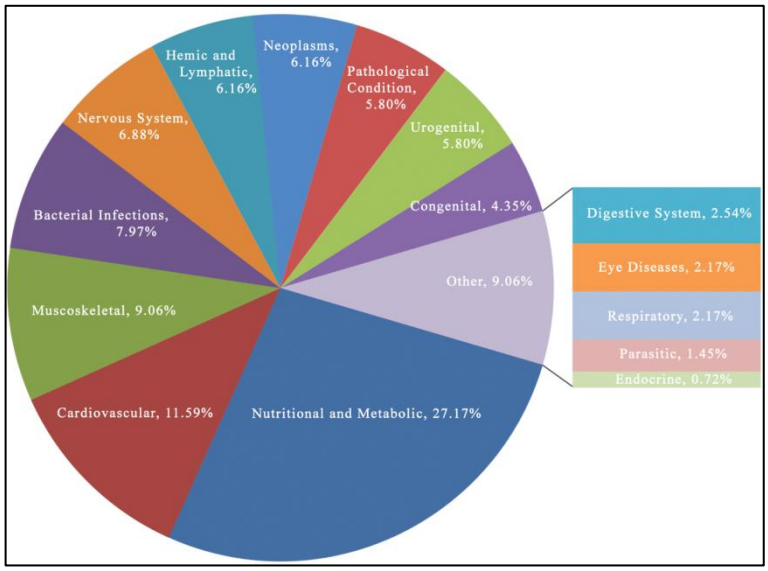

Figure 4.10. Distribution of disease terms and disease classes in CFN-G graph

4.3. Evaluation of CFN Functional Modules

The computational framework to evaluate the functional modules enumerated from CFN constitutes the scoring the functional modules for co-occurring KEGG pathway terms, localization information, and an integrated disease ontology composed of MeSH terms and OMIM database, co-expression patterns and evaluation of these modules. The steps that are followed were explained in Methodology and general layout of the computations is depicted in Figure 3.2.

4.3.1. Evaluation and Scoring of CFN Modules

The functional modules enumerated in CFN were evaluated in terms of the shared pathways, cellular localization and shared disease terms. There is accumulating evidence that proteins participate in the same/similar pathways are often co-localized and tend to share GO terms. In addition, genes contributing to a common disorder have an increased tendency for their products to interact through protein-protein interactions. Therefore, taking the pathways and localization information together has the potential to generate a reliable measure to select the informative functional modules.

Each of these functional modules was scored for the frequency of appearance of pathway, localization and disease term. KEGG pathway classification scheme involves 338 pathways [163], LOCATE scheme contains 30 cellular localizations [164] and MeSH disease scheme involves 3630 disease terms [161]. Co-expression is another indicator to evaluate the relation between the proteins. The gene expression data was extracted from Gene Expression Omnibus [166]. To understand the co-expression profile Pearson correlation coefficient in assigned to each of the interactions in the network. Later for each of the 566 functional modules the average Pearson correlation coefficient is calculated and assigned to the corresponding module (PCC_{AVG}).

To achieve a reliable measure to analyze and evaluate the functional modules a non-linear model, which is a function of the corresponding scores of R_{KEGG}, R_{LOC} R_{OMIM} and PCC_{AVG}, was proposed. Genetic Algorithm is applied to estimate the model parameters where the population was evolved for 100 generations. The estimated model parameters were then used to evaluate the functional models enumerated from CFN. The estimated model parameters are presented in Table 4.5.

Table 4.5. Estimated nonlinear model parameters for CFN

Classification	α	β
N	2.0155	0.2159
R_{KEGG}	2.0998	0.0017
R_{LOC}	1.9674	0.0177
R_{OMIM}	2.0016	0.0002
PCC_{AVG}	1.8023	0.2109

4.3.2. Assembly of Top Scoring CFN Modules

The top scoring functional module is a four member module constituting F7, F9, F10 and FURIN and has well documented roles in extrinsic pathway of blood coagulation. F7, F9 and F10 are core proteins, however FURIN is a ubiquitously expressed endoprotease [215], which responsible for the maturation of key platelet aggregation/coagulation mediators synthesized by the platelet producers [216]. FURIN has been found to improve the processing efficiency and activity of F9 [217].

Table 4.6. Top scoring functional module in CFN

Module431	R_{KEGG}	R_{LOC}	R_{OMIM}	PCC_{AVG}
Score	1.000	0.7206	0.6568	0.4931
p-val	0.00E+00	6.68E-07	6.87E-06	-
F7, F9, F10, FURIN	Complement and coagulation cascade	Extracellular region, endoplasmic reticulum	Atherosclerosis, myocardial infarction, coronary artery disease, hematuria	

The selection criterion was based on the clusters formed by the assembly. Since the major aim is to elucidate the shared mechanisms, the number of clusters formed was the basis. According to this, when the members of 25 top scoring functional modules are assembled in a network, five distinct clusters were observed. These individual clusters correspond to the major mechanisms that play important role in the disease, such as coagulation cascade, renin-angiotensin system, lipoprotein mechanism and inflammatory response. However, the assembly of top scoring 100 functional modules formed a single cluster; thereby potentially indicated shared mechanisms. In addition, the number of hub proteins included in the network was maximized for the top-100 network, hence sufficient to link the shared pathways. The basis of the selection criteria is presented in Table 4.7.

Table 4.7. Selection criteria of the top scoring functional modules in CFN according to the results obtained from Genetic Algorithm

Modules	***Top-25***	***Top-50***	***Top-75***	***Top-100***	***Top-150***	***Top-200***
N	50	83	98	108	130	147
l	121	253	375	412	587	709
c-proteins	34	52	58	62	84	101
hub proteins	1	7	9	10	10	10
clusters	5	4	3	2	2	2

Among 566 functional modules evaluated, the 108 proteins involved in top scoring 100 functional modules were assembled in a condensed linkage network indicating the proteins involved in the biological processes of CVD. This condensed linkage network is presented in Figure 4.11. The proteins appear in this network suggest the possible intersections between the biological processes. Among those 108 proteins, 62 of them are the core proteins, i.e. it is well documented that these proteins are associated with CVD.

Among the core proteins, 10 proteins were identified as hub proteins, which play central roles in CFN. 46 of these proteins were not reported at the time this study started.

To elucidate the shared pathways the proteins assembled in this network were analyzed with AmiGO to identify distinct biological processes. Seven GO terms with non-overlapping 59 members indicated the major processes involved in the progress of the disease. The GO terms enriched in the condensed map are presented in Table 4.8.

Table 4.8. GO terms enriched in the condensed network. Seven distinct GO terms corresponding to separate processes were elucidated

GO Term	***p-val***	**Genes**
GO:0007598 blood coagulation, extrinsic pathway	7.19E-10	F3 F9 F10 F7 TFPI
GO:0042730 fibrinolysis	2.47E-08	F12 GP1BA F2 SERPINE1 PLG F11 PLAU
GO:0010883 regulation of lipid storage	5.39E-06	PPARA APOB CD36 LPL APOC4 ITGB3
GO:0003018 vascular process in circulatory system	1.15E-10	VEGFA NOS1 AGTR2 NOS3 NPPA AGT AGTR1 PTGS2 APOE INS ACE EDN1
GO:0030299 intestinal cholesterol absorption	4.32E-06	APOA1 APOA2 LEP APOA4 LDLR
GO:0046427 positive regulation of JAK-STAT cascade	2.15E-07	IL6 IL2 IL4 IL3 CSF2 IFNG IL5
GO:0007249 I-kappaB kinase/NF-kappaB cascade	1.10E-15	TLR8 MYD88 IRAK2 LY96 TLR10 TIRAP TNFRSF1A TLR7 CD40 TNF ADIPOQ TLR6 IL10 TLR9 IRAK4 TLR2 TLR4

The members of the top scoring functional modules in CFN were assembled in a condensed functional linkage network (Figure 4.11). Circle nodes represent the proteins that were previously reported to be associated with the disease. Diamond nodes are also core proteins, but these proteins serve as hubs in the CFN. Square nodes represent the proteins that are captured through the extension of the core network. The nodes are colored according to the GO terms enriched in the network. In this condensed network, as expected, the presence of the members present in the complement and coagulation cascade, fibroblast growth factors, the proteins involved in the lipoprotein mechanism, renin-angiotensin system and the processes involved in inflammation was clearly observed through visual inspection.

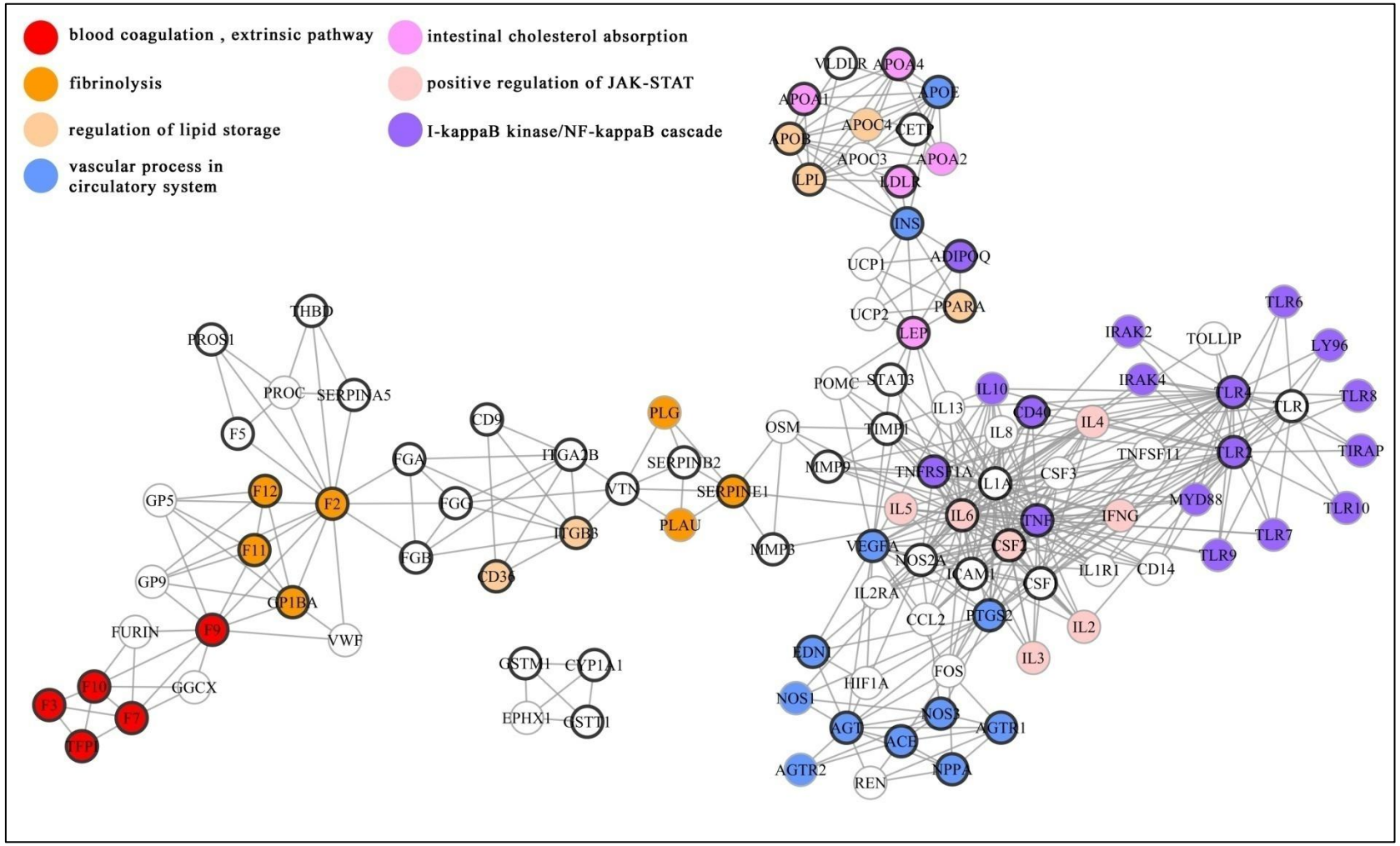

Figure 4.11. Condensed functional linkage network constructed from the top scoring CFN functional modules

F3, F9, F10, F7, TFPI represent the members of the extrinsic pathway of coagulation cascade (*p-val* = 7.19E-10) in this condensed network. Blood coagulation and platelet-mediated primary haemostasis are well characterized defense mechanisms against bleeding, and is triggered in response to rupture of endothelium, which allows exposure of blood to the extravascular tissue. At sites of vascular injury, activation of blood coagulation results in the generation of high concentrations of thrombin that activate platelets and coagulate blood. Depending on the factors affecting the initiation of the coagulation process, the cascade is mainly composed of two paths: extrinsic and intrinsic pathway. Quantitatively, intrinsic pathway possesses a slower action and extrinsic pathway provides a rapid response to tissue injury. The instant response of extrinsic pathway is to augment the activation of intrinsic pathway, which takes several minutes. The initiation of the extrinsic pathway starts with the exposure of tissue factor (F3) to blood, serving as a cofactor to the enzyme F7, where subsequently the F3-F7 complex efficiently activates factor IX (F9) and factor X (F10) [48]. In fact, a careful control of coagulation is required for anticoagulant mechanisms and, under normal conditions; they prevail over the procoagulant forces [49]. An efficient coagulation system is accomplished by several anticoagulant mechanisms, ensuring that the clotting process remains a local process [218]. One of the ways of controlling blood coagulation is achieved by tissue factor pathway inhibitor (TFPI) where it regulates the very initiation of coagulation by binding and inhibiting active form of factor V that is associated with the F3-F7 complex [219, 220].

Another cluster associated with fibrinolysis through intrinsic blood coagulation constitutes F12, GP1BA, F2, SERPINE1, PLG, F11, PLAU, which ultimately possess fibrinolysis (*p-val* = 2.47E-08), where F2, F11 and F12 are well characterized proteins involved in the intrinsic pathway of blood coagulation. The initiation of intrinsic pathway starts with Factor XII (F12) binding of the protein to the endothelial surface that is exposed to tissue injury. Upon the activation of F12, further amplification is triggered by the activation of Factor XI (F11). The extrinsic and intrinsic pathways coalesce at the production of thrombin (F2) [49, 221]. Thrombin results in thrombus formation either by conversion of fibrinogen to fibrin or by platelet activation. Plasmin is the major fibrinolytic protease. Plasminogen (PLG), a circulating plasma zymogen, can be converted into plasmin by urokinase (PLAU). Through a positive feedback mechanism, plasmin cleaves PLAU, transforming into a more active peptide. Fibrin, the major plasmin substrate,

regulates its own degradation by binding plasminogen activator (PLA) on its surface, thereby localizing and enhancing plasmin generation. Once formed, plasmin cleaves fibrin, generating soluble degradation products. Fibrin dissolution is also regulated by inhibitors of PLG activation, such as PLG activator inhibitor (SERPINE1) [222]. Although not presented with the fibrinolysis GO annotation; Von Willebrand Factor (VWF), which is a glycoprotein, serves as a mediator for the adhesion of platelets to the collagen exposed on endothelial cell surfaces. VWF has an important function in primary homeostasis and has a pivotal role in thrombogenesis. Glycoproteins, GP1BA, GP5 and GP9 are the major receptors for VWF, promoting a more stable interaction with circulating platelets [223].

Notably, in the condensed network the blood coagulation pathways (both intrinsic and extrinsic) are connected to fibrinolysis through fibrinogens (FGA, FBG, FGG), which are cleaved by thrombin to form fibrin, where the cleavage products of fibrinogen and fibrin regulate cell adhesion displaying vasoconstrictor activities; antigens (CD9, CD36) and integrins (ITGA2B, ITGB3). Addition to these, the modularization of fibrinogens, antigens and integrins are completed with vitronectin (VTN), which functions as a cell adhesion factor. VTN is a circulatory protein that regulates coagulation and fibrinolysis, cell lysis via the complement cascade, and cell binding and migration. VTN forms a bridge in the condensed network; it might have a dual effect on fibrinolysis while interacting both with ITGB3 and SERPINE1. Reports suggesting that inhibition of ITGB3 are influenced by the alternations in the VTN expression [224], and the lifetime extension of SERPINE1 upon binding of VTN during fibrinolyis inhibition [225] provides additional evidence for the possible dual role of VTN in fibrinolysis. SERPINE1-VTN binding interaction also affects cell adhesion and motility. Elevated SERPINE1 activities are associated both with coronary thrombosis and with a poor prognosis in many cancers [225]. Therefore, in the condensed network the link relating fibrinolysis to tumor progression is proposed to be established through the ITGB3-VTN-SERPINE1 connection. VTN is suggested to play a pivotal role in the process; hence further assessments are required to evaluate its potential as a drug target.

In the condensed network, two proteins encoded by the genes FURIN and GGCX were not present in the core proteins, i.e. were not previously associated with CVD. FURIN is a ubiquitously expressed endoprotease [215], which responsible for the

maturation of key platelet aggregation/coagulation mediators synthesized by the platelet producers [216], and involved in tumor growth and progression [226]. The association of FURIN with the coagulation cascade was reported in Chinese hamster ovary cells, where over-expression of FURIN improved the processing efficiency and resultant specific activity of F9 [217]. GGCX is a carboxylase which was reported to be involved in the familial combined deficiency of vitamin K-dependent coagulation factors. All vitamin K-dependent coagulation factors, F2, F7, F9, and F10 are carboxylated to attain binding of calcium and attachment to phospholipid membranes. These evidences suggest that, although previously reported, FURIN and GGCX are involved also in the platelet activation and coagulation process, hence further investigation is required to assess their function as mediators in the coagulation cascade. In fact, the pivotal role of FURIN in the malignant tumor progression and mediator in coagulation cascade may provide the link between the neoplasms and CVD.

APOB, LPL, APOC4, PPARA, CD36, ITGB3 are functionally associated with regulation of lipid storage (*p-val* = 5.39E-06). Apolipoprotein B (APOB) plays a vital role in transportation of lipids in the form of chylomicrons, which subsequently undergo lipolysis by lipoprotein lipase (LPL). LPL is the principal enzyme responsible for hydrolysis of circulating triglycerides and is active in differentiated macrophages. Although its role in lipid metabolism is known, reports suggest that interleukin-6 (IL6) [69] and tumor necrosis factor (TNF) [204] inhibit LPL activity, thereby these proinflammatory cytokines have pathophysiological roles in altering human adipocytokine metabolism. Although statins have a moderate effect on increasing high-density-lipoproteins (HDL) concentrations with 5-10 per cent, the action of transcription factor peroxisome proliferator-activated receptor alpha (PPARA) under fibrate influence has a profound effect on the therapy to raise HDL concentration and decrease triglyceride levels without making a major effect on low-density-lipoprotein (LDL) levels [185]. The involvement of CD36 in lipid metabolism is through its role in the prevalence of modified LDL uptake in macrophages [183]. In fact, the binding of modified LDL to CD36 triggers the release of proinflammatory cytokines.

As a continuation of the regulation of lipid storage, APOA1, APOA2, APOA4, LDLR and LEP have roles in intestinal cholesterol absorption (*p-val* = 4.32E-06). APOA1,

APOA2 and APOA4 are the members of apolipoprotein family, where APOA1 and APOA2 are the main components of HDL. The HDL regulation is also controlled through PPARA, where its agonists have been shown to up-regulate APOA1 gene [227] and increase in the APOA2 concentration, thereby contributing to the increase in the HDL concentration [228], where low-density-lipoprotein receptor (LDLR) contributes to the uptake of HDL by the liver. In this cluster associated with the cholesterol absorption, although the function of APOA4 is not precisely determined, it participates in the regulation of many pathways, appears to be stimulated by fat absorption, which is mediated by chylomicrons [229], and involves in the body weight regulation in humans [230]. In fact, leptin (LEP), an adipocyte derived hormone, controls body weight and lipid storage and ensures body energy homeostasis through three different mechanisms: control of food intake, energy partitioning and energy expenditure [231]. However, recent report suggests the regulation of APOA4 is achieved by LEP, indicating the existence of a functional linkage leading to food intake suppression [232].

Along with lipid and cholesterol metabolism, vascular processes in the circulatory system (*p-val* = 1.15E-10) that controls the blood pressure is presented in the condensed network by the proteins APOE, INS, AGTR2, AGT, AGTR1, ACE, NOS1, NOS3, NPPA, VEGFA, PTGS2, EDN1. The unexpected entities in this cluster are apoliporotein E (APOE), which is involved in the lipid metabolism, and insulin (INS), which is associated with insulin metabolism. The presence of these proteins is attributed to the cross-talk between the lipid, insulin and renin-angiotensin mechanism (RAS), where the majority of the proteins belong to. RAS has been the center for intensive research for its pivotal roles in regulating blood pressure in hypertension, as well as optimization of blood glucose. Angiotensinogen (AGT), angiotensin II receptor, type 1 (AGTR1), angiotensin II receptor, type 2 (AGTR2), angiotensin I converting enzyme (ACE) have pronounced roles in the RAS [233, 234].

The blockage of RAS exerts potent antiatherosclerostic events, hence many inhibitors of the RAS are clinically approved for reducing blood pressure, including drugs that target ACE, AGTR1 and AGTR2 [41]. The contribution of NOS elements is due to their vasodilator role in the vascular homeostasis. NOS has three major isoforms, neuronal NOS (nNOS or NOS1), inducible NOS (iNOS or NOS2), the endothelial NOS (eNOS or

NOS3), all of which have the same mechanism of NO production. NOS1 and NOS3 constitutively produce small quantities of NO and are expressed by endothelium and perivascular nerves, respectively [56, 235]. Despite NOS1 has been identified in atherosclerotic lesions, its role is unclear. But experimental studies support the evidence that NOS1 also generates vascular responses [236]. On the other hand, NOS3 is activated by mechanical stresses such as blood flow-mediated shear stress and stimulation with vasoactive substances. NOS3 controls vascular tone, inhibits monocyte and leukocyte adhesion to the endothelium, inhibits platelet aggregation, and decreases endothelial permeability. However, the action of NOS3 catalytic activity can be uncoupled, leading to superoxide production. Reaction with superoxide reduces the bioavailability of NO, impairs vasomotor function, and increases platelet aggregation, monocyte and leukocyte adhesion to the endothelium. Superoxide and other reactive oxygen species are involved in modification of lipids, induce proinflammatory genes, and increase cellular proliferation. Superoxide can also activate matrix metalloproteinases and produce apoptosis, which may contribute to instability of atherosclerotic lesions. Therefore superoxide can be regarded in general as a pro-atherogenic molecule [236]. There is direct evidence that NOS3 mutations can cause hypertension [237, 238] and many of the local functions of NO counterbalance proatherogenic effects of angiotensin II, showing that NOS3 deficiency in mice have increased plasma renin activity, hence probably also have increased production of angiotensin II [239]. Similar to NOS, natriuretic peptide precursor alpha (NPPA) controls the extracellular fluid volume, plays a key role in the cardiovascular and electrolyte homeostasis. Endothelin 1 (EDN1) is a potent vasoconstrictor that maintains vascular tone, and associated with pulmonary hypertension [157]. It has been reported that AGTR1 exerts indirect contractile response by simulating the synthesis of EDN1 [233].

The involvement of signaling pathways in the pathophysiological processes of the cardiovascular diseases received much attention. JAK-STAT signaling (*p-val* = 2.15E-07) and NF-κB signaling (*p-val* = 1.10E-15) pathways are also represented in the condensed network. The former is one of the major pathways that cytokine receptors signal through. Five members of interleukin family (IL2, IL3, IL4, IL5, IL6), CSF2 and IFNG are annotated with positive regulation of JAK-STAT pathway. JAK kinases and STAT proteins are constitutively expressed in the vessel wall and transfer intracellular signaling events of various receptor families, such as cytokines, growth factors and vasoactive

peptides such as angiotensin II or EDN1. The activation of JAK-STAT pathway is mediated through variety of ligands and their receptors, where its activation leads to stimulation of cell proliferation, differentiation, migration and apoptosis. However, the inhibition of the pathway is achieved through negative regulators such as: suppressors of cytokine signaling (SOCS), protein inhibitors of STATS (PIAS) or protein tyrosine phosphatases [240]. In the last two decades, there was an intense research to understand the action of interleukins as inflammatory cytokines and their prominent roles in the activation of JAK-STAT pathway [55, 241-244]. The potential effect of the JAK-STAT signaling pathway on cardiovascular pathophysiology is possibly due to the fact that JAKs may bind directly to AGTR1 [245].

The latter, NF-κB pathway is one of the main signaling pathways activated in response to proinflammatory cytokines, including TNF, and IL1, as well as following activation of the Toll-like receptors (TLR). It is now a well documented phenomenon that inflammation is significantly associated with cardiovascular disease [2, 56, 246]. NF-κB signaling is stimulated by cytokines such as TNF, which activates the IκB kinase (IKK) complex that phosphorylates IκB proteins, leading to IκB degradation and NF-κB translocation into the nucleus. Activated NF-κB regulates transcription from promoter regions of approximately 300 genes, including those encoding cytokines and several NF-κB family members that can feedback to regulate the system. Signaling through NF-κB can regulate diverse cellular outcomes, including cell death or division [247].

The module which is not connected to the major component of the condensed network constitutes proteins GSTT1, GSTM1, CYP1A1 and EPHX1. GSTM1 and GSTT1 are members of Glutathione-S-transferase family of proteins, which play a major role in cellular antioxidant defense mechanisms by catalyzing the reduction of potentially harmful peroxides. The family of enzymes is known as detoxification enzymes responsible for the metabolism of a broad range of xenobiotics and carcinogens [248]. GSTM1 and GSTT1 enzymes detoxify not only products of oxidative stress but also carcinogenic compounds such as polycyclic aromatic hydrocarbons, a major constituent of tobacco smoke. Therefore, these proteins were suggested as susceptibility factor in smoking related CHD [249]. The variations in the GSTM1 and GSTT1 were shown to alter the susceptibility to atherosclerosis [248]. Also these proteins have protective roles for developing CVD [250].

Similar to GSTM1 and GSTT1, the genetic variability in CYP1A1 is suggested to play a role in coronary heart disease [251, 252].

Although GSTM1, GSTT1 and CYP1A1 have well documented roles in CVD, contradictory results were reported for EPHX1. Microsomal epoxide hydrolase (EPHX1, also referred as mEH) is a xenobiotic-metabolizing enzyme involved in the biotransformation of reactive intermediates and is expressed in several human tissues including the heart and blood vessels [253] and liver, lung, kidneys, intestine, brain, prostate, heart and testes [254]**.** The broad substrate selectivity and prominent expression in the liver and other metabolizing organs, ensures the widespread defense against potential genotoxic epoxides. Due to the special features described above, the mEH is involved in the efficient detoxification of many reactive epoxide intermediates [255]. Mutations in this gene cause preeclampsia, neurodegenerative disorders such as Alzheimer's and Parkinson's disease, early-onset lung cancer, and respiratory disorders [256-259]. Although, EPHX1 was not considered to play a significant role in the development of CVD [257], EPHX1 is capable of modulating the levels of DNA damage, which may contribute to the development of CVD. Since, there is growing evidence that DNA damage caused by mutagens found in tobacco smoke may accelerate the development of atherosclerosis [260], there is a need for further assessment in determining the roles of EPHX1 in cardiovascular disease.

In the condensed network, as mentioned above fibrinolysis action is proposed to be associated with blood coagulation through ITGB3-VTN-SERPINE1 connection. The cellular processes involved in fibrinolysis and blood coagulation is connected to the rest of the condensed network through oncostatin M (OSM), which is not previously reported to be associated with CVD. Oncostatin M (OSM) is a pleiotropic cytokine that is produced by activated T-cells and macrophages and has shown to influence the course of inflammatory responses [261]. The protein is structurally and functionally related to IL6 and IL11, proteins that also influence immune and inflammatory function [262]. Despite OSM is not previously associated with CVD, there is direct evidence suggesting that SERPINE1 present in the vessel wall through its antifibrinolytic and antiproteolytic properties is modulating key events in the pathophysiology of cardiovascular diseases. Increased amounts of SERPINE1 have been detected in atherosclerotic lesions [263]. Recently,

SERPINE1 has been also shown to increase smooth muscle cell proliferation and to decrease smooth muscle apoptosis [264]. Recent reports showed that OSM robustly up-regulates the expression of SERPINE1 in human vascular smooth muscle cells [265] and suggested to establish the link between coagulation and inflammation [266]. Hence, our finding in this work supporting the potential role of OSM in connecting coagulation to inflammation is evident and in concordance with the previous reports.

Another protein that is not previously associated with CVD is the hypoxia induced transcription factor 1 (HIF1A), might have roles in relating renin-angiotensin system with inflammation. HIF1A induces vascular endothelial growth factor (VEGF) pathway, as a response to growth factors and hormones such as insulin, in endothelial cells and plays an important role in mammalian cells cultured under reduced oxygen tension and systemic response to hypoxia. A number of complex disorders such as myocardial, cerebral and retinal ischemia, preeclampsia and cancer were reported to be closely associated to the deprivation of oxygen and glucose [168]. Although a number of studies were undertaken to elucidate the function of HIF1A, there is still need to investigate the detailed effect of HIF1A in CVD and other related conditions [169]. Therefore further detailed studies on oxygen homeostasis and HIF1A are required to understand the underlying mechanisms of these pathophysiological processes and its potential therapeutic role.

4.3.3. Disease Interventions Derived from CFN

The significant associations between diseases were generated through the shared genes among disease pairs. Initially the manual curation and elimination of MeSH terms yielded 3630 disease terms in the disease classification scheme. The 108 proteins in the condensed network were related to 445 disease terms. The associations between these 445 diseases were calculated for each pair in terms of shared proteins. The significant associations were determined through random experiments, where each disease is associated with randomly assigned proteins, where the number of proteins linked with a particular disease term is kept constant. The significantly associated 162 disease pairs sharing at least two proteins were selected (*p-val* < 2.00E-03). According to this selection criterion, the disease pairs having a significant partnership are assembled in a complex disease map, presented in Figure 4.12. In this map, these 162 diseases were colored

according to the disease class, derived from Medical Subject Headings classification. The dominant disease class belongs to cardiovascular disease class, as expected. Along with the cardiovascular diseases, various type of neoplasms and diseases affecting endocrine metabolism, connective tissues and respiratory system are present in this map. Notably, the urogenital disorders are localized around cardiovascular diseases, suggesting that several types of kidney diseases are significantly also associated. The selected disease pairs having significant associations were presented in Table 4.9. The corresponding disease classes were shown in parenthesis, *DO* represents the disease overlapping score and *p-val* indicates the significance of the linkage.

Proteinuria and left ventricular hypertrophy (LVH) is associated with a disease overlapping score 0.429 (*p-val* = 4.28E-13) where NOS3, TNF and ACE are shared among these diseases. Also, LVH is also linked to general kidney disease (*p-val* = 8.39E-08). Proteinuria is a kidney disease characterized by the presence of excessive protein in the urine and is accepted as an early manifestation of kidney dysfunction. It is regarded as a marker of loss of the normal selective barrier at the glomerular filtration slits. It also strongly correlated with markers of endothelial dysfunction [267]. On the other hand, left ventricular hypertrophy is the thickening and enlargement of the muscle in left ventricle of the heart, which belongs to a cardiovascular disease class. These two disorders have been found to be associated through TNF, NOS3 and ACE, where TNF is a multifunctional cytokine with an important role in inflammatory response and involved in lipid metabolism, coagulation, insulin resistance and endothelial function [55, 187]. NOS3 is activated by blood flow-mediated stress and has a role in controlling vascular tone and inhibits platelet aggregation and related with decreased endothelial permeability [236]. ACE has a prominent role in RAS involved in the regulation of blood pressure and RAS pathway inhibition facilitates the reduction of blood pressure [41]. The fact that ACE is a vasoconstrictor and NOS3 is a vasodilator indicates the importance of vascular homeostasis in both proteinuria and LVH. Statins are a class of drugs that are used to lower blood cholesterol levels and suggested to improve the prognosis of LVH [268]. Inhibition of ACE with statins arrests proteinuria and reduce modestly urinary protein excretion particularly in populations with cardiovascular disease [269-271]. Hence, our finding showing that proteinuria and LVH are strongly associated through partnering proteins can be attributed to the mechanisms maintaining vascular homeostasis, in concordance with a

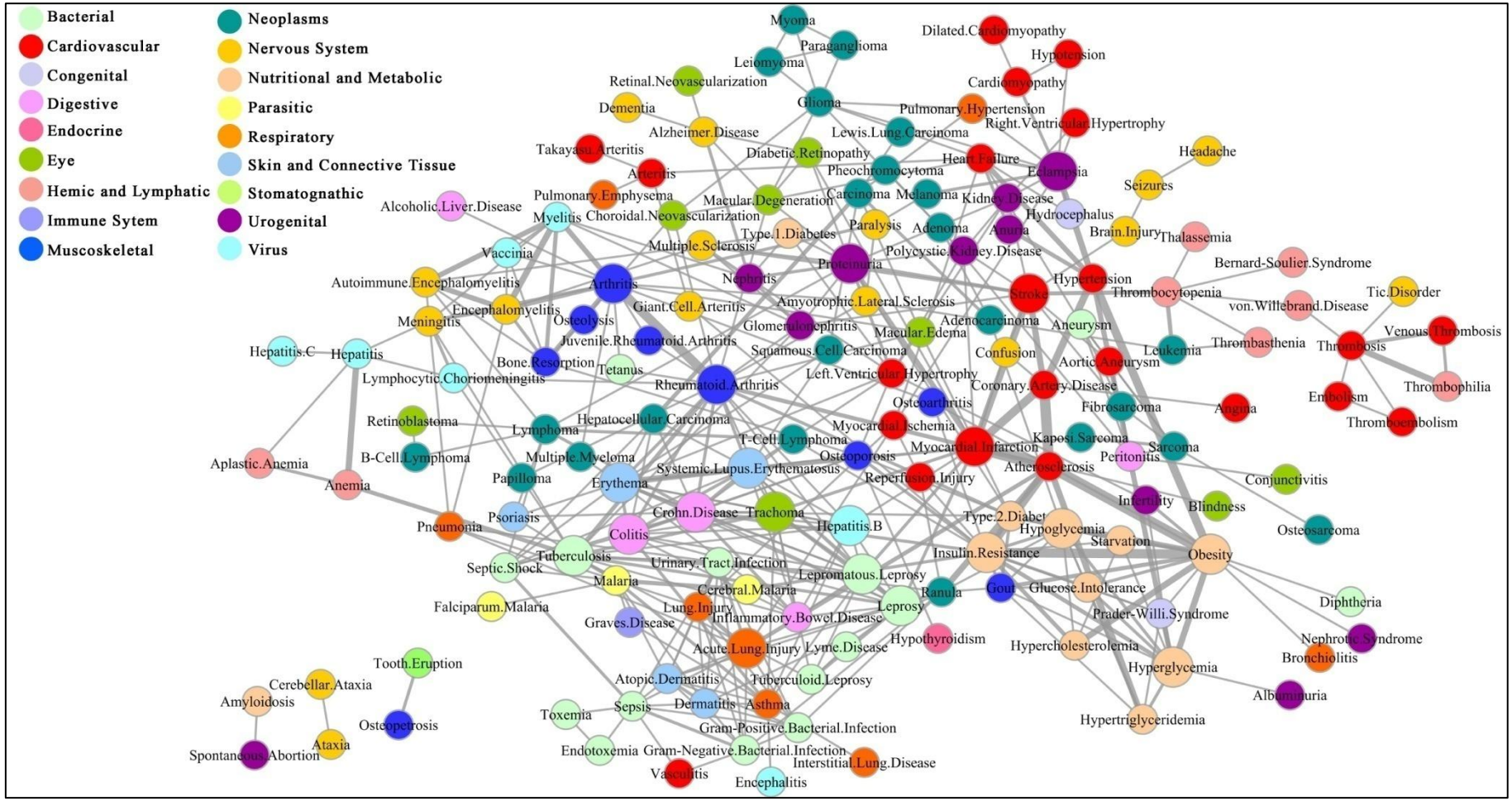

Figure 4.12. Disease associations derived through CFN

Table 4.9. Selected pairs of disease terms derived from CFN displaying significant associations in terms of shared genes

DiseaseTerm1	**DiseaseTerm2**	***DO***	***p-val***	**Shared genes**
Proteinuria (Urogenital)	Left Ventricular Hypertrophy (Cardiovascular)	0.429	4.28E-13	NOS3, ACE, TNF
Anemia (Hemic and Lymphatic)	Hepatitis (Virus)	0.429	5.49E-13	LDLR, PLG, GSTM1, IFNG, EPHX1, IL10
Proteinuria (Urogenital)	Myocardial Infarction (Cardiovascular)	0.294	4.63E-10	NOS3, ACE, TNF, VEGFA, TNFRSF1A
Atherosclerosis (Cardiovascular)	Rheumatoid Arthritis (Musculoskeletal)	0.114	1.57E-09	CCL2, IL10, IL6, VEGFA
Left Ventricular Hypertrophy (Cardiovascular)	Kidney Disease (Urogenital)	0.333	8.39E-08	NOS3, TNF
Aortic Aneurysm (Cardiovascular)	Sarcoma (Neoplasms)	0.250	5.68E-07	MMP2, PLG
Asthma (Respiratory Tract)	Colitis (Digestive System)	0.231	5.58E-05	IL10, TNF, IL13
Tuberculosis (Bacterial Infections)	Rheumatoid Arthritis (Musculoskeletal)	0.267	6.22E-05	CCL2, IFNG, IL10, TNF
Systemic Lupus Erythematosus (Skin and Connective Tissue)	Rheumatoid Arthritis (Musculoskeletal)	0.231	2.76E-04	IL10, IL6, TNF
Hypertension (Cardiovascular)	Obesity (Nutritional and Metabolic)	0.276	1.66E-03	ACE, CD36, CETP, LEP, POMC, PPARA, SERPINE1, ADIPOQ
Atherosclerosis (Cardiovascular)	Obesity (Nutritional and Metabolic)	0.278	1.73E-03	ADIPOQ, APOB, CCL2, CD36, CETP, FGA, INS, LEP, TLR4, TNF, UCP1

recent cohort study predicting the strength and nature of the association between proteinuria and the subsequent risk of CVD [272]. Therefore, the use of therapeutic strategies targeting to reduce proteinuria coupled with cardioprotection might be also beneficial the overall burden of CVD.

In addition, a significant association of proteinuria and myocardial infarction was also observed with a *DO* = 0.249 (*p-val* = 4.63E-10) through partnering proteins NOS3, ACE, TNF, VEGFA and TNFRSF1A. The presence of TNF, VEGFA and TNFRSF1A as shared proteins suggests that inflammation is one of the key players in these disorders. TNF is a famous multifunctional cytokine, VEGFA is an endothelial mitogen which is involved in vascular permeability, angiogenesis, vasculogenesis and also in the inhibition of apoptosis [201]. Despite the blockage of RAS is suggested to reduce the risk of myocardial infarction with kidney dysfunction, a recent study employed the markers of inflammation and homeostasis in predicting the outcome of myocardial infarction in terms of platelet aggregation and effects on kidney disease. Taken together, our finding indicating the association of proteiuria with LVH and myocardial infarction suggest that these diseases are linked not only by RAS but also cytokines with an implication of inflammatory pathways, hence development of therapeutic strategies targeting both RAS and inflammatory mechanisms require further assessment to prevent the progress of CVD and the presence of proteinuria might be one of the markers in the disease.

Atherosclerosis is linked with rheumatoid arthritis (RA) with *DO* = 0.114 (*p-val* = 1.57E-09), although a low disease overlapping score was calculated for this association, the high confidence obtained through randomized experiments indicated its significance (Table 4.9). The link among these disorders was established through CCL2, IL10, IL6 and VEGFA. RA is characterized with the chronic inflammation of the joints. The co-existence of IL6 and IL10, which are well known inflammatory cytokines, indicates the possible link of between inflammatory pathways and atherosclerosis. VEGFA maintains vascular permeability, angiogenesis and vasculogenesis. CCL2 as a partnering protein establishing the link between atherosclerosis and RA was not previously reported with its association with CVD. In fact, CCL2 is one of the best studied chemokines, also known as monocyte chemotactic protein-1. Its binding with CC-chemokine receptor 2 initiates a series of signaling events recruit monocytes to inflammation sites in a variety of chronic

inflammatory diseases, hence regulates the gene expression that coordinate homeostatic response [273]. Increased cardiovascular morbidity and mortality is observed with RA patients, where the life expectancy is reduced by 10-15 years [274, 275]. The finding that endothelial dysfunction is a marker of early atherosclerosis is also valid for RA, and hence the therapies targeting anti-inflammatory mechanisms are suggested to improve the endothelial function in RA [276]. Supported with the evidence with previous reports, inflammatory mechanisms responsible for lesions in RA and involvement of endothelial dysfunction may contribute to the development of atherosclerotic plaques. Notably, RA is also found to be linked with tuberculosis (*p-val* = 6.22E-05) and systemic lupus erythematosus (SLE) (*p-val* = 2.76E-04) through inflammatory cytokines. The presence of CCL2 as a partnering protein in the link between RA and tuberculosis also supports our finding that CCL2 might have a vital role in maintaining homeostatic response.

5. TYPE 2 DIABETES RELATED FUNCTIONAL LINKAGE NETWORK

5.1. Construction of Type 2 Diabetes Related Functional Linkage Network

The construction of Type 2 diabetes functional linkage network was started with the proteins encoded by 494 genes (c-proteins) collected from genome–wide associations reported in literature [277-282]. The linkages between the proteins were extracted from STRING database v8.1, which establishes interactions through functional linkages, such as physical interactions, curated biological pathway knowledge, co-expression profiles, as well as the co-occurrences of protein pairs in database text fields and conservation across species, and assigns a confidence score according to the reliability of supporting data [158]. The core set of proteins were incorporated with the first neighbors to achieve a comprehensive disease related network constituting putative proteins that have potential links with the disease. To select the suitable confidence score for interactions, confidence scores varying from 900 to 990 was set to construct networks and these networks were analyzed in terms of coverage of core set of proteins and constitution of core proteins in the network. According to these selection criteria, as shown in Table 5.1., the confidence score threshold was set as 940; selection of such a stringent threshold confidence score for the linkages enabled us to construct a coherent functional linkage network. Although, all c-proteins are covered for a confidence score 900, such a lenient criteria leads to abundance in the linkages, therefore a confidence score 940 was set as the threshold to eliminate the linkages while keeping the sufficient amount of core proteins in the network. Following the removal of singletons, the giant component of network based on the confidence score 940 comprised 2734 nodes (proteins) and 14823 edges (functional linkages) and this network was entitled as Type 2 diabetes related functional linkage network (TDFN). In this network, among the 2734 proteins, 474 of them are c-proteins, which have previously defined associations with the disease. These 474 proteins form 17.33 per cent of the functional linkage network.

Table 5.1. Selection of confidence score for the construction of TDFN

Confidence score	**900**	**920**	**940**	**960**	**980**	**990**
N	3693	3217	2755	2201	1550	1221
l	31294	20788	14889	9966	5014	3222
c-proteins	494	489	477	458	406	365
<*k*>	8.47	6.46	5.45	4.53	3.23	2.64
% constitution	13.38	15.20	17.31	20.81	26.19	29.89
% coverage	100.00	98.99	96.56	92.71	82.19	73.89

5.1.1. Analysis of TDFN Topological Properties

The topological features of TDFN were analyzed in terms of the degree, k, clustering coefficient, *C*, and betweenness, *b*, for each node. The average degree <*k*>, average clustering coefficient <*C*> and average betweenness <*b*> for the network were calculated to be 10.84, 0.291 and 7581.6 respectively. The cumulative degree distribution $n(k)$ (Figure 5.1) and clustering coefficient distribution $C(k)$ (Figure 5.2) with respect to degree distribution, followed the Power law, indicating that many proteins are linked to a few other proteins but only a few of them have many interactions, hence providing additional evidence that the network of interest exhibits a scale-free behavior with a degree exponent, $\gamma = 2.54$ ($R^2 = 0.8995$).

The betweenness distribution $b(k)$ was also well characterized with Power law scaling (Figure 5.3), meaning that many proteins are located at the periphery while a few proteins are located at the center of the network, and hence responsible for communication within the network. The effect of removing the nodes with higher betweenness is similar to that of excluding a hub protein from the network and proteins displaying high betweenness are reported to be likely essential and evolutionary conserved.

5.1.2. Selection of Hub Proteins in TDFN

Hub protein in network play crucial roles in the networks while maintaining stability and providing intercommunication. Therefore, these key players in the TDFN were selected to

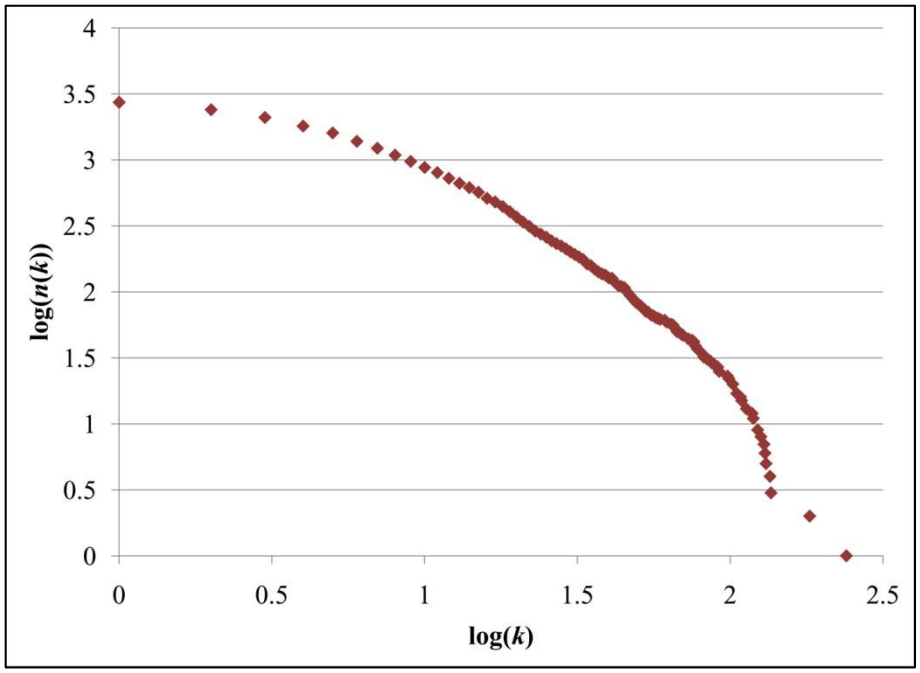

Figure 5.1. Cumulative degree distribution $n(k)$ of TDFN with respect to degree distribution in TDFN

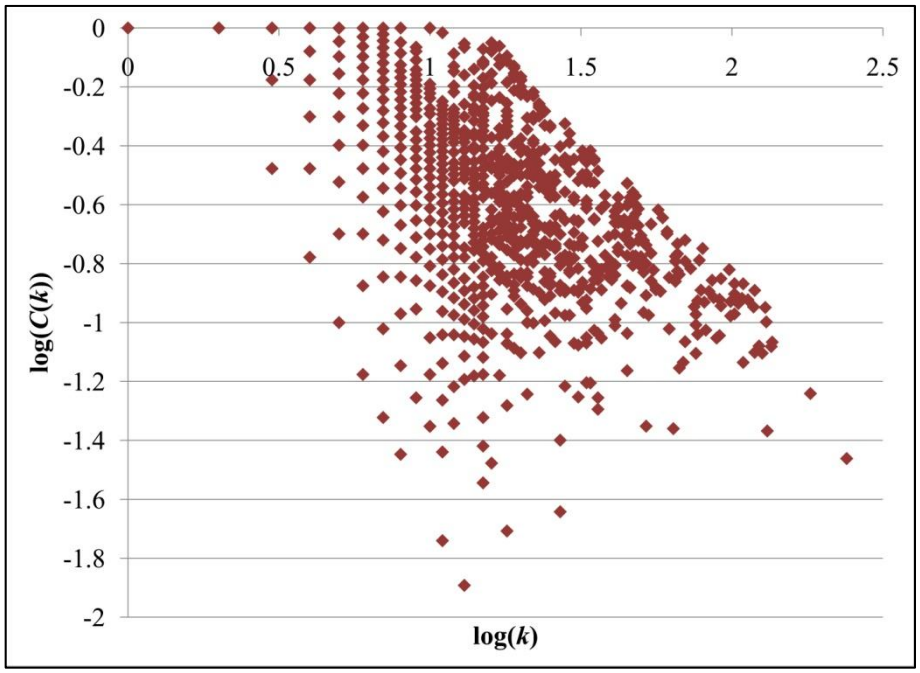

Figure 5.2. Clustering coefficient distribution $C(k)$ with respect to degree distribution in TDFN

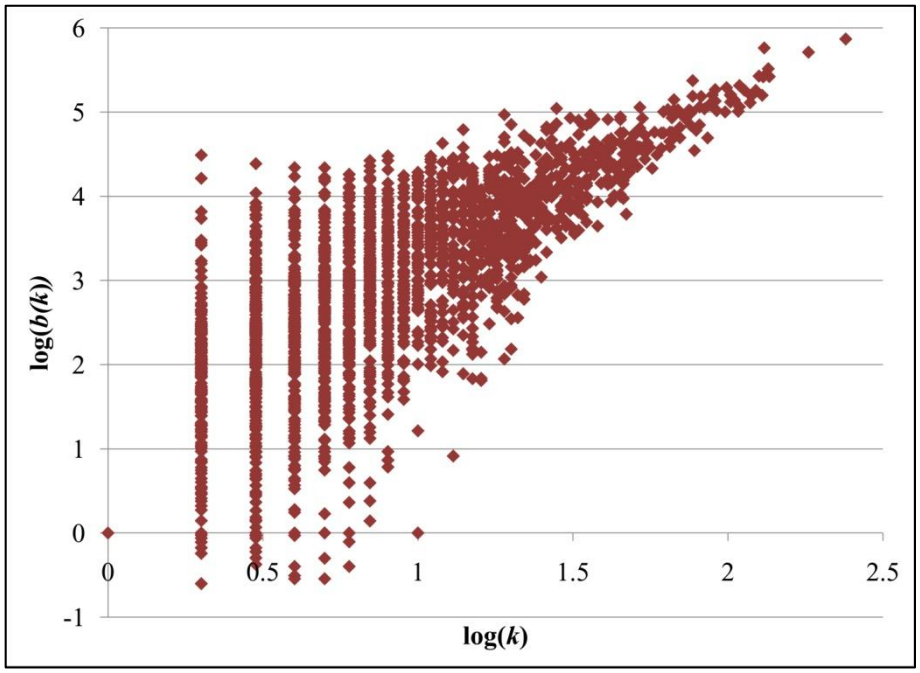

Figure 5.3. Betweenness distribution $b(k)$ with respect to degree distribution in TDFN

understand the governing major biological systems present in the network. The hub protein selection criterion is based on the accumulative degree and betweenness distributions with respect to degree. These two measures were incorporated to identify hub proteins. To select the most informative entities in the network, the proteins that are listed at the top one per cent of the accumulative degree and accumulative betweenness distributions were entitled as hub proteins. The accumulative degree distribution graph shows that 12.1 per cent of the proteins in TDFN have single linkage (Figure 5.4), whereas 18.4 per cent of the proteins were located at the periphery (Figure 5.5), hence no contribution to the communication in the network.

The intersection of the highest top one per cent accumulative degree and betweenness distribution yielded 23 proteins, which have vital roles in the network. The degree of hub proteins ranges between 89 to 240 with an average of 121 interactions, whereas the betweenness (log(b)) ranges between 5.113 to 5.869.

15 out of 23 hub proteins are previously associated with the disease, listed in c-proteins. The hub proteins identified in the network are presented in Table 5.2.

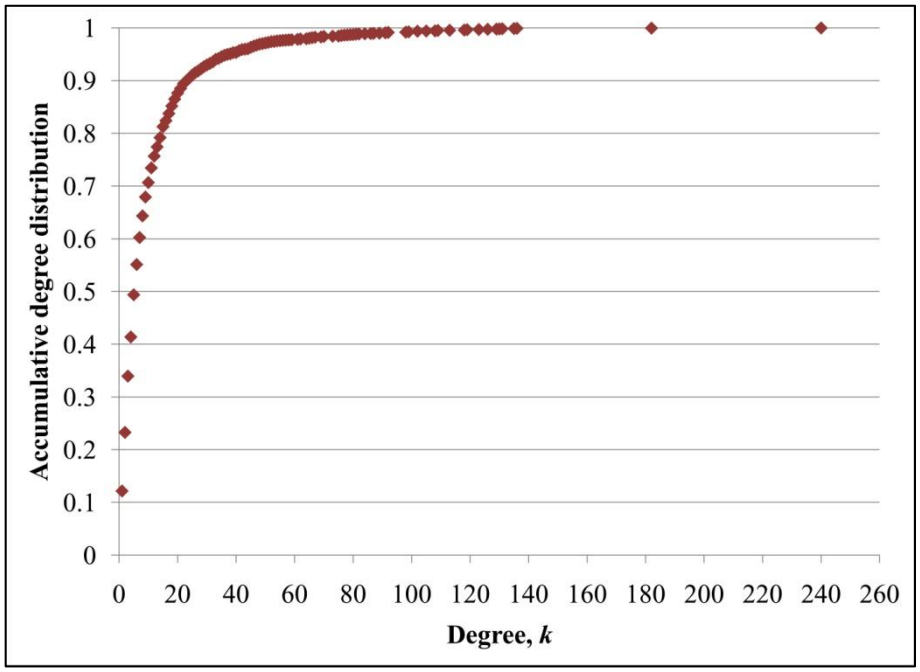

Figure 5.4. Accumulative degree distribution of TDFN with respect to degree distribution

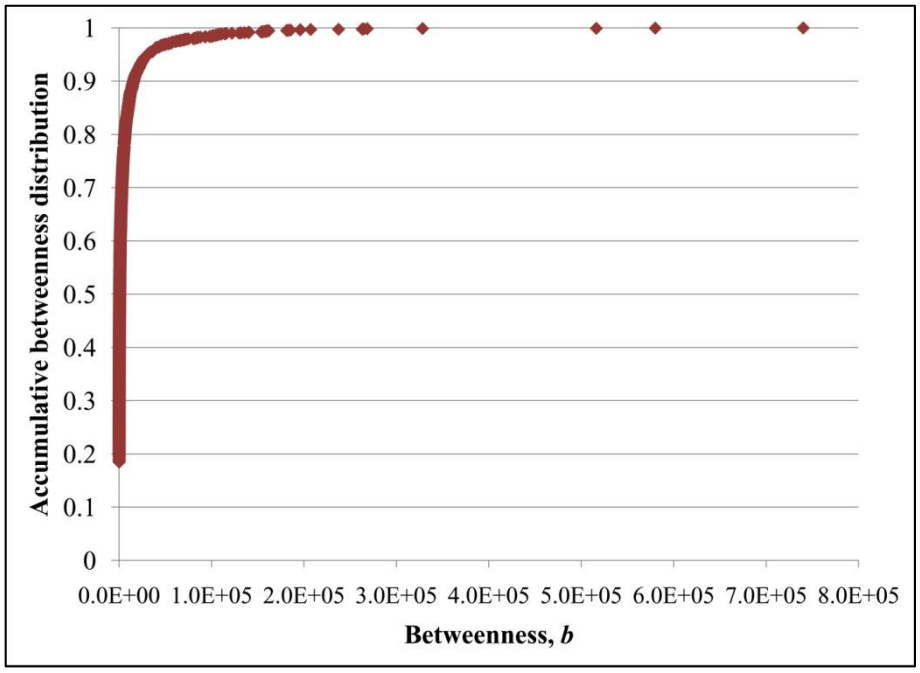

Figure 5.5. Accumulative betweenness distribution of TDFN with respect to degree distribution

Table 5.2. Topological properties of hub proteins in TDFN

Symbol	**Name**	**Degree, *k***	**Betweenness, log(*b*)**	**Clustering coefficient, *C***
TP53	tumor protein p53	240	5.869	0.0346
AKT1	v-akt murine thymoma viral oncogene homolog 1	182	5.713	0.0574
EGFR	epidermal growth factor receptor	136	5.423	0.0858
SRC	proto-oncogene tyrosine-protein kinase Src	135	5.516	0.0829
INS	insulin	131	5.764	0.0429
JUN	transcription factor AP-1	130	5.420	0.1005
IFNG	interferon, gamma	129	5.201	0.1125
MAPK8	mitogen-activated protein kinase 8	126	5.429	0.0787
MAPK3	mitogen-activated protein kinase 3	123	5.258	0.0833
IL2	interleukin 2	119	5.203	0.1286
MAPK1	mitogen-activated protein kinase 1	119	5.192	0.0792
GRB2	growth factor receptor-bound protein 2	118	5.113	0.1130
STAT3	signal transducer and activator of transcription 3	113	5.263	0.1195
EP300	E1A binding protein p300	109	5.317	0.0732
STAT1	signal transducer and activator of transcription 1	105	5.145	0.1161
IL6	interleukin 6	102	5.197	0.1355
CCND1	G1/S-specific cyclin-D1	102	5.209	0.1250
MYC	C-myc protein	102	5.121	0.1070
VEGFA	vascular endothelial growth factor A	99	5.292	0.1053
FOS	v-fos FBJ murine osteosarcoma viral oncogene homolog	99	5.292	0.1179
EGF	epidermal growth factor	91	5.132	0.1199
CTNNB1	catenin	91	5.268	0.0904
IGF1	insulin-like growth factor 1 (somatomedin C)	89	5.205	0.0886

The 23 hub proteins in TDFN have degrees, *k*, varying in the range of 89 (IGF1 – insulin like growth factor) to 240 (TP53 –tumor protein 53). TP53 is a tumor repressor and responds to a variety of cellular stresses that regulate cell cycle arrest, apoptosis and DNA repair. Gene Ontology (GO) terms enriched in the 23 hub proteins indicate the major cellular mechanisms that govern throughout the network. The biological processes GO categories significantly enriched in hub proteins (*p-val* < 1.48E-09) indicate that the selected hub proteins are involved in the regulation of response mechanisms to external stimulus, signal transduction and cell proliferation. The selected GO terms enriched in 23 hub proteins are listed in Table 5.3.

Table 5.3. Gene Ontology terms enriched in the hub proteins in TDFN

GOTerm	***p-val***	**Proteins**
GO:0007243 intracellular protein kinase cascade	4.08E-16	STAT1 IL6 INS IL2 MAPK8 GRB2 STAT3 IFNG AKT1 MAPK1 EP300 SRC EGF CTNNB1 EGFR
GO:0008284 positive regulation of cell proliferation	1.54E-15	STAT1 IL6 CCND1 INS MYC IL2 JUN IFNG IGF1 VEGFA MAPK1 EGF CTNNB1 EGFR
GO:0009605 response to external stimulus	1.56E-13	STAT1 IL6 CCND1 INS MYC IL2 FOS JUN IFNG AKT1 VEGFA MAPK1 EP300 TP53
GO:0070887 cellular response to chemical stimulus	6.20E-12	STAT1 IL6 CCND1 INS GRB2 FOS STAT3 IFNG AKT1 VEGFA EP300 CTNNB1
GO:0009725 response to hormone stimulus	2.17E-11	STAT1 IL6 CCND1 INS GRB2 FOS STAT3 AKT1 MAPK1 EP300 CTNNB1
GO:0045941 positive regulation of transcription	9.47E-10	IL6 MYC IL2 FOS STAT3 JUN IFNG VEGFA MAPK1 EP300 TP53 CTNNB1
GO:0010628 positive regulation of gene expression	1.48E-09	IL6 MYC IL2 FOS STAT3 JUN IFNG VEGFA MAPK1 EP300 TP53 CTNNB1

5.1.3. Enumeration of Functional Modules in TDFN

Genes participate in similar biological processes, share GO terms and operate in similar functions have a tendency to localize as dense groups in interaction networks [5]. These entities are considered as functional modules, where the members functionally linked to each other. The functional modules in TDFN were derived based on modularity measure, as explained in Methodology. The Python scripting language is implemented to decipher the functional modules, where the members have the maximum allowable interaction, hence $Q = 1$. The algorithm used in this study, rather than assigning proteins into distinct clusters, allows the presence of proteins in many functional modules. The algorithm produced 10015 functional modules, the size of the modules ranges from two to 14, with an average module size of 4.03, hence with the supporting information that modules consisting four or more members are biologically meaningful, the 5355 modules of size four and above were considered for further analyses. In the selected functional modules, out of 2734 proteins in TDFN, 1362 (49.8 per cent) were represented.

The hub modules play important roles in maintaining the communication in the network, therefore to understand the extent of the effect due to any loss of activity in these proteins, the distribution of hub proteins in the functional modules was calculated. As expected, because of their high numbers of interaction and presence in central network, all

of the hub proteins display the highest frequency in functional modules. Figure 5.6 shows the frequency of occurrence of the hub proteins in the functional modules with respect to the other proteins. The selected hub proteins are indicated with red diamonds. For instance, TP53 is present in 526 modules with a frequency of occurrence 0.098. The graph is presented in descending appearance of proteins from highest to lowest. The hub proteins are colored to facilitate visual inspection. The self distribution of each hub protein in functional modules according to the size of the modules is presented in Figure 5.7. TP53 is distributed among functional modules with varying sizes from four to eight. Notably, the largest modules in TDFN belong to protein complexes.

The two largest modules consist of 14 proteins, where the members are DNA directed polymerase II subunits (POLR2A, POLR2B, POLR2C, POLR2F, POLR2G, POLR2H, POLR2I, POLR2K, POLR2L), TFIIH basal transcription factor complex helicase subunits (ERCC2, ERCC3), cell division kinase protein 7 (CDK7), transcription initiation factor IIB (GTF2B), TATA-box-binding protein (TBP) and cyclin-H (CCNH), where TBP and CCNH are interchangeable in two modules. These proteins function in RNA elongation from RNA polymerase II promoter (*p-val* = 3.92E-41), where the extension of RNA is performed after the transcription initiation at an RNA polymerase II-specific promoter.

The six modules consisting of 12 proteins belongs to proteasome subunits (PSMA1, PSMA2, PSMA3, PSMA4, PSMA5, PSMA6, PSMA7, PSMB1, PSMB2, PSMB3, PSMB4, PSMB5, PSMB6, PSMB7, PSMD4) with interchangeable proteins PSMB6, PSMB7 and PSMD4. These proteins function in the negative regulation of ubiquitin-protein ligase activity (*p-val* = 7.60E-36).

On the other hand, other 12 member functional modules consist of the minichromosome maintenance complex (MCM2, MCM3, MCM4, MCM5, MCM6, MCM7, MCM10), origin recognition complex (ORC1L, ORC2L, ORC4L, ORC5L), cell division cycle 45-like (CDC45L), cell division cycle homolog 6 (CDC6) and activator of S phase kinase (DBF4), with varying proteins. The proteins in these modules play roles in DNA replication (*p-val* = 1.00E-25). Replication of DNA must be accurate and precisely regulated, since it should be completed only once in each cell cycle. The initiation of cell

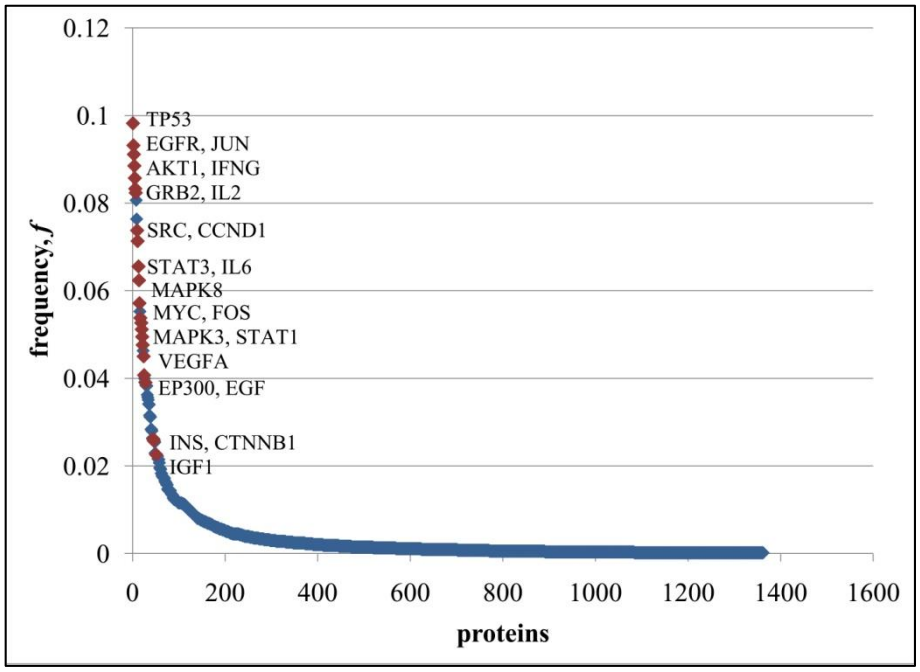

Figure 5.6. Frequency of presence of the proteins in TDFN in functional modules

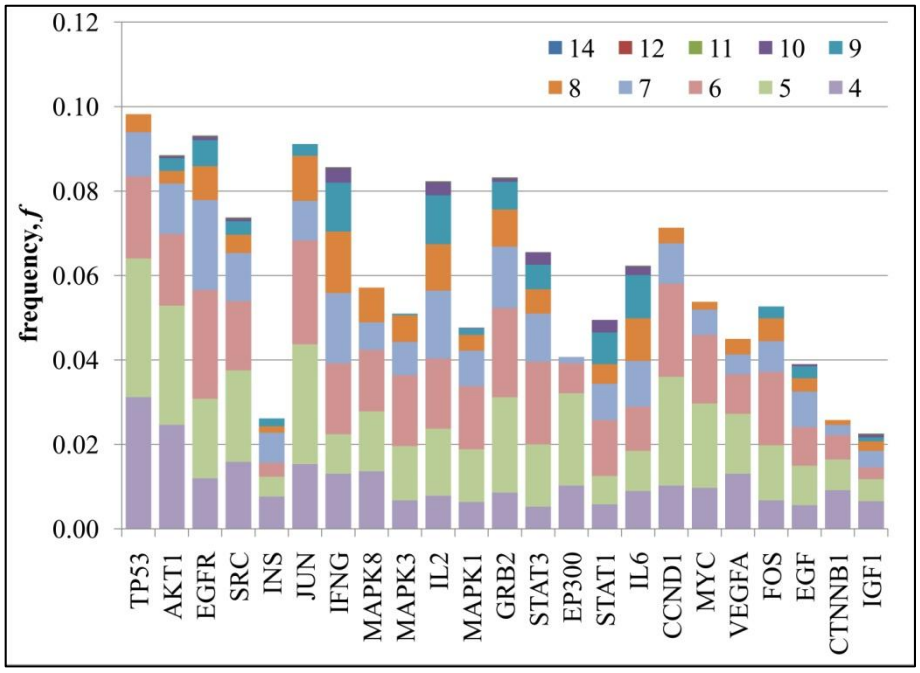

Figure 5.7. Self distribution of hub proteins in TDFN according to the size of the functional modules

cycle requires the assembly of pre-replication complex, which recruits many proteins and complexes. ORCL1-6 proteins are the members of origin recognition complex which binds to chromatin. The formation of origin recognition complex is a cascade of protein subunit

interactions, where the major component is ORCL2, which is responsible to bind the chromatin throughout the cell cycle [283]. CDC6 associates with ORCLs and the pre-replication complex is completed upon the interaction of MCM2-7 proteins. The initiation of DNA replication is triggered with a coordinated action of DBF4-CDC7 and CDC45L. The action of CDC45 is to regulate DNA replication by controlling only one replication of chromosomal DNA is completed per cell cycle, hence considered as the rate limiting step in the DNA replication process [284].

5.2. Evaluation of TDFN Functional Modules

The second phase includes the evaluation and scoring of the functional modules according to the member's co-occurrence in similar pathways, co-localized, shared similar disease terms and exhibit similar expression patterns. The computational layout explaining the steps considered is the same with CFN, presented in Figure 3.2.

5.2.1. Evaluation and Scoring of TDFN Modules

The module enumeration algorithms produce massive amounts of entities to be analyzed to elucidate the most informative and reliable components. Due to the mounting evidence that proteins function together to exhibit a single action often tend to participate in similar pathways, co-localized and share GO Terms. In fact, genes contributing to a disorder have increased tendency for their products to be functionally related. Hence, the functional modules enumerated from TDFN were evaluated in terms of sharing similar expression patterns, participation in pathways, co-localization and association with similar diseases. Combination of various resources provides a deliberate and consistent evaluation method.

The scoring of the functional modules was initiated with the assembly of the data that will be incorporated. KEGG Pathway database was used to associate biological pathways, where the classification scheme involves 338 pathways, including major metabolic processes, as well as disease pathways [163]. LOCATE database was utilized to determine the co-localization information, where the proteins are assigned to 30 different subcellular compartments [164]. Medical Subject Headings (MeSH) [161] was

incorporated with OMIM database to achieve the disease associations. Manual curation of the MeSH database yielded 3,630 disease terms and these disease terms were categorized into 23 different disease classes depending on the system that is exposed to disease. These 3,630 disease terms were then searched in OMIM database to associate genetic information with the diseases. To assess the reliability of the scores assigned to the modules, the associations in the classification schemes were randomly shuffled 10^3 times and calculated score was compared with the distribution of the random scores.

Co-expression patterns of 58 samples were analyzed through the incorporation of four microarray experiments, namely GSE7146 [285], GSE15653 [286] and GSE2510 [166, 287]. To evaluate the co-expression pattern in the functional modules, Pearson correlation coefficients (PCC) were calculated and assigned for each interaction in the network. The functional module average PCCs (PCC_{AVG}) were calculated from the interactions and this score is considered as the co-expression score of the modules.

Upon the scoring of the functional modules, four different scores were obtained: R_{KEGG}, R_{LOC}, R_{OMIM} and PCC_{AVG}. These scores were varied from zero to one, where one indicates the consistency in the module. To select the informative top scoring functional modules a non-linear model was proposed, where the parameters were estimated using Genetic Algorithm. The 10 artificially generated functional modules, five of which have the one for each scoring scheme, were planted in the population, and the population of the modules was evolved for 100 generations. The estimated model parameters were then used to evaluate the modules enumerated from TDFN and summarized in Table 5.4.

Table 5.4. Estimated nonlinear model parameters for CFN

Classification	α	β
N	1.8371	1.7865
R_{KEGG}	1.4732	0.0322
R_{LOC}	1.5706	1.7374
R_{OMIM}	0.9576	1.8596
PCC_{AVG}	0.9838	1.6176

5.2.2. Assembly of Top Scoring TDFN Modules

One of the top scoring functional modules belongs to DNA replication unit, involving minichromosome maintenance complex (MCMs), origin recognition complex member ORC5L, cell division cycle proteins (CDCs) and activator of S phase kinase (DBF4) protein. These well-studied proteins are involved in the initiation of DNA replication and regulation of cell cycle, present in nucleous, nucleolous and cytoplasm. The module might provide information that links various types of neoplasms, such as carcinoma and leukemia to lactose intolerance and polycystic kidney disease.

Table 5.5. Top scoring functional module in TDFN

Module205	R_{KEGG}	R_{LOC}	R_{OMIM}	PCC_{AVG}
Score	0.885	0.7043	0.6780	0.6166
p-val	0.00E+00	0.00E+00	2.06E-03	-
MCM2, MCM3, MCM4, MCM5, MCM6, MCM7, MCM10, CDC7, CDC6, CDC45L, ORC5L, DBF4	Cell cycle, DNA Replication	Nucleus, Nucleolus, Cytoplasm	Carcinoma, Leukemia, Adenocarcinoma, Lactose Intolerance, Polycystic Kidney Disease	

The selection criterion for the number of top scoring functional modules that will be assembled was based on the number of proteins, clusters and hub proteins. Since, the motivation of such as assemble is to derive the underlying shared mechanism; the number of clusters obtained through incorporation of the overlapping members of the selected functional modules is an important parameter. Addition to this, the number of core proteins in the network is another issue that should be considered to keep the association with the disease. Considering the decline in the number of clusters after top-50 modules, and the increase in the number of included hub proteins, the top scoring 75 functional modules were selected to be incorporated to construct a condensed functional linkage network. The core proteins, which have well defined associations with the disease, constitute 26.16 per cent of this condensed network. Additionally among the 23 hub proteins, 8 of them were present in this network. The basis of the criteria is presented in the Table 5.6.

Table 5.6. Selection criteria of the top scoring functional modules in TDFN according to the results obtained from Genetic Algorithm

Modules	***Top-25***	***Top-50***	***Top-75***	***Top-100***	***Top-150***	***Top-250***
N	56	89	172	235	312	420
l	352	507	957	1458	2061	3001
c-proteins	3	14	45	60	80	107
hub proteins	0	0	8	11	11	13
clusters	2	4	3	3	4	4

Among 5355 modules, the overlapping members of 75 top scoring functional modules were assembled reflecting the biological processes involved in the disease. This condensed network, presented in Figure 5.8., contains 172 proteins, 127 of which were not previously reported with Type 2 diabetes, and 957 interactions.

5.2.3. Indicators of Fundamental Cellular Processes in TDFN

To elucidate the fundamental biological processes involved in the disease and to determine the shared partners of the processes, the proteins in the condensed network was analyzed in terms of distinct Gene Ontology terms enriched in the map. 12 GO Terms with non-overlapping 111 members indicating the major cellular processes involved in the progress of the disease. The GO Terms enriched in this condensed map is represented in the Table 5.7. The core proteins in the assembly of proteins (Figure 5.8) are black bordered and the nodes are colored according to the distinct GO Term that is enriched in the network.

In this network, the largest group of proteins represented by the members of the proteasome unit (PSMs) function in positive regulation of ligase activity. PSMA6, PSMD9 and SMAD have previously known associations with the disease. PSMA6 gene encodes the proteasome subunit α type 6, a component of the 20S proteasome, the core particle for the 26S ubiquitin–proteasome system (UPS), which is responsible for the degradation of the majority of intracellular proteins. UPS is activated by various stimuli, including oxidative stress and plays a pivotal role in the regulation of many cellular processes, particularly in the activation of nuclear factor B (NF-κB) a transcription factors, which induces the transcription of proinflammatory cytokines. It was reported that, Type 2

diabetes up-regulates UPS in rat muscle [288]. The investigation of the myocardial infarction susceptibility in Type 2 diabetes showed that UPS plays an important role in arterial plaque formation [289], hence the up-regulation of the UPS was suggested to be potential mechanism that links myocardial infarction to the disease [290]. Another core protein, coactivator Bridge-1 (PSMD9), regulates the transcriptional activation of glucose-responsive enhancers in the insulin gene. The over-expression of PSMD9 in transgenic mice reduces insulin gene expression and results in insulin deficiency and severe diabetes. Dysregulation of PSMD9 signaling increases pancreatic apoptosis, hence glucose homeostasis and pancreatic ß-cell survival is regulated through the signals mediated through PSMD9 [291]. The expression of PSMD9 can be stimulated with activin [292]. One distinctive core protein that is found to be involved in proteasome activity is SMAD7. SMAD7 inhibits TGF-β and activin signaling by binding to the activated receptor complex. Thus, SMAD7 functions as an adaptor protein that mediates degradation of the TGF- β receptor complex [293]. PSMD9 expression is stimulated with activin simultaneously with SMAD2, SMAD3 and SMAD4 [292]. A recent study reported an increased proteasome activity in insulin-resistant muscle [294], where the muscle protein degradation is caused by the suppression of PI3K/Akt signaling leading to activation of caspase-3 and ubiquitin-proteasome proteolytic pathway [295]. Taken together, PSMD9 and SMAD7 may have regulatory roles in the proteasome activity through activin signaling, leading to the suppression of PI3K signaling and decreased insulin expression. Two proteins, that are not associated with diabetes in previous studies, cell division cycle 2 (CDC2; cyclin dependent kinase 1, CDK1) and polo-like kinase 1 (PLK1) are also found to be involved in ubiquitin-protein ligase activity. These two proteins have pronounced roles in DNA damage checkpoint. Recently, in *Saccharomyces cerevisiae*, PLK1-CDC2 forms a negative feedback loop that is suggested to have pivotal roles in preventing the activation of the DNA damage checkpoint during mitosis [296].

Another cluster that is linked with transcription initiation consists of 26 proteins include two core proteins, peroxisome proliferator-activated receptor gamma, coactivator 1 alpha (PPARGC1A) and polymerase (RNA) II (DNA directed) polypeptide D (POLR2D). The protein encoded by the gene PPARGC1A is a transcriptional regulator that is involved in energy metabolism and is an important factor regulating the expression of genes for oxidative phosphorylation and ATP production in target tissues through coactivation of

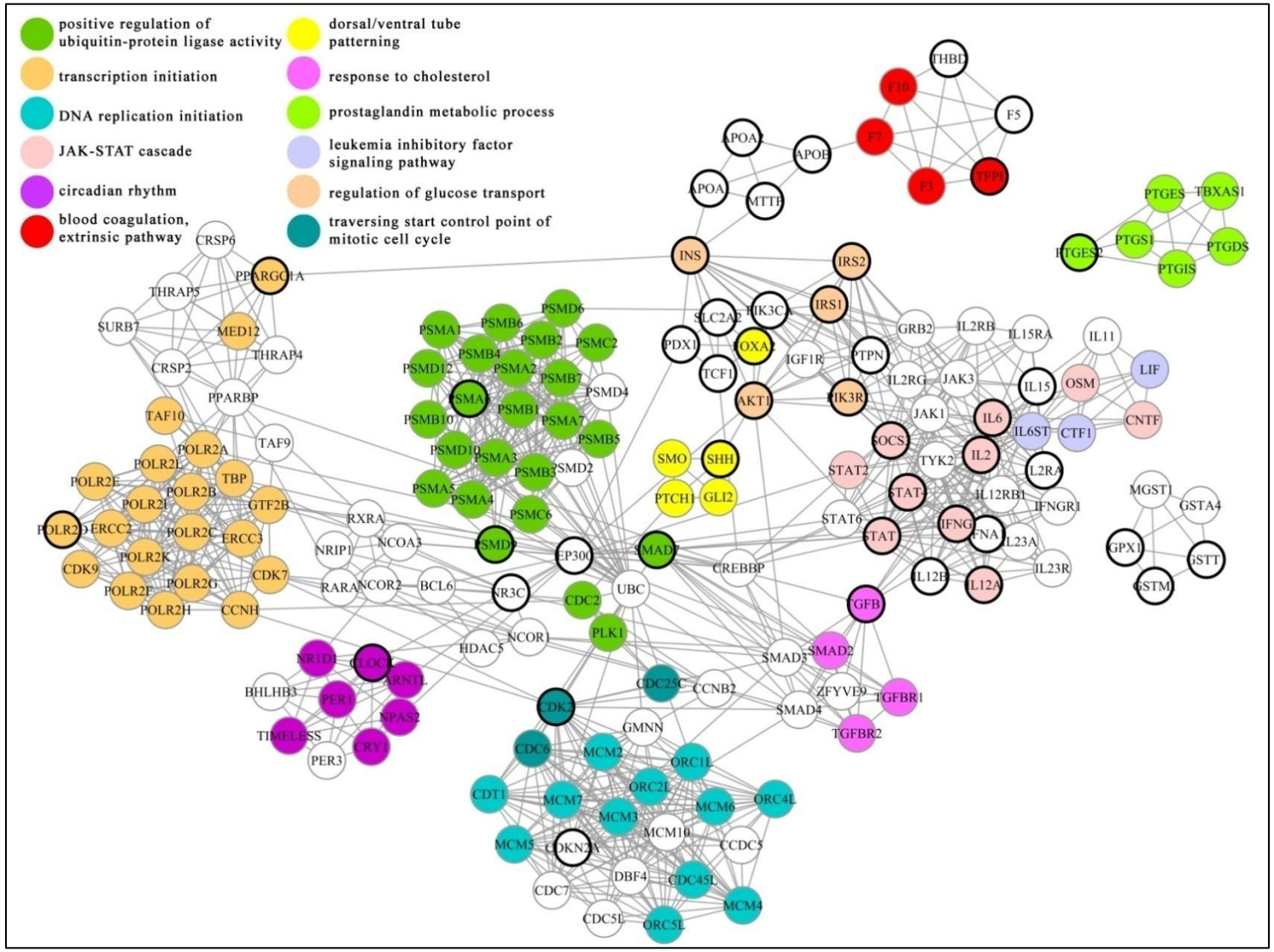

Figure 5.8. Condensed functional linkage network constructed from the top scoring functional modules in TDFN

Table 5.7. 12 distinct GO terms corresponding to separate cellular processes enriched in the TDFN condensed network

GO Term	***p-val***	**Genes**
GO:0051443 positive regulation of ubiquitin-protein ligase activity	9.83E-33	CDC2 PLK1 PSMA1 PSMA2 PSMB10 PSMD12 PSMB6 PSMA3 PSMB2 PSMA7 PSMD6 PSMA6 PSMD9 PSMC6 PSMC2 PSMB5 PSMB3 PSMD10 PSMB7 PSMA5 PSMA4 PSMB1 PSMB4 RPS27A SMAD7
GO:0006352 transcription initiation	1.44E-31	CDK9 MED12 POLR2C POLR2K POLR2F POLR2I TBP MED17 POLR2A ERCC2 TAF10 CDK7 CCNH POLR2L POLR2H PPARGC1A MED16 POLR2E MED14 POLR2G ERCC3 POLR2B MED24 POLR2D MED1 GTF2B
GO:0006270 DNA replication initiation	7.53E-15	MCM4 MCM3 MCM5 ORC1L MCM7 CDC45L MCM6 MCM2 ORC4L CDT1 ORC2L ORC5L
GO:0007259 JAK-STAT cascade	3.62E-10	IL6 IFNG OSM STAT4 IL12A SOCS3 IL2 STAT2 STAT1 CNTF
GO:0007623 circadian rhythm	3.62E-07	TIMELESS CLOCK CRY1 NPAS2 BHLHE41 ARNTL NR1D1 PER1
GO:0007598 blood coagulation, extrinsic pathway	8.72E-06	TFPI F7 F10 F3
GO:0021904 dorsal/ventral neural tube patterning	4.09E-04	SMO GLI2 PTCH1 FOXA2 SHH
GO:0070723 response to cholesterol	5.63E-04	TGFB1 TGFBR2 TGFBR1 SMAD2
GO:0006693 prostaglandin metabolic process	9.39E-04	PTGS1 PTGES PTGES2 PTGIS TBXAS1 PTGDS
GO:0048861 leukemia inhibitory factor signaling pathway	3.63E-03	IL6ST LIF CTF1
GO:0010827 regulation of glucose transport	3.67E-03	PIK3R1 AKT1 IRS2 INS IRS1
GO:0007089 traversing start control point of mitotic cell cycle	7.23E-03	CDC6 CDK2 CDC25C

nuclear receptors. PPARGC1A mRNA expression has been found to be correlated with glucose-stimulated insulin release, and its inhibition of expression was shown to be associated with a decline in INS mRNA expression [297]. On the other hand, the protein encoded by POLR2D is the fourth largest subunit of RNA polymerase II, the polymerase responsible for synthesizing messenger RNA in eukaryotes.

The third largest group of proteins in the TDFN condensed network belong to DNA replication unit consist of 12 proteins forming the minichromosome maintenance complex (MCM2, MCM3, MCM4, MCM5, MCM6, MCM7), origin recognition complex (ORC1L, ORC2L, ORC4L, ORC5L), cell division cycle 45-like (CDC45L), chromatin licencing and DNA replication factor (CDT1). These proteins are the members of pre-replication complex. Replication of DNA must be inherently accurate and precisely regulated, since

DNA replication should be completed only once in each cell cycle. The initiation of cell cycle requires the assembly of pre-replication complex, which recruits many proteins and complexes. During G1 phase of cell cycle, each origin is licensed by the formation of pre-replication complex. ORC proteins are the members of origin recognition complex which binds to chromatin. The formation of origin recognition complex is a cascade of protein subunit interactions, where the major component is ORC2L, which is responsible to bind the chromatin throughout the cell cycle. On the other hand, MCM2-7 proteins are the members of the microchromosomal maintenance complex, which is another essential component, associates with CDT1:CDC6:ORC to complete pre-replication complex [283]. The activation of pre-replicative complexes at origins of DNA replication at the beginning of S phase occurs due to the action of many CDKs [298]. This cluster is also linked with transversing start control of mitotic cycle (*p-val* = 7.23E-03), where CDC6, CDK2 and CDC25L are the members. The action of CDC45L is to regulate DNA replication by controlling only one replication of chromosomal DNA is completed per cell cycle, hence considered as the rate limiting step in the DNA replication process. In addition, compared to the normal human cells, cancer derived cells harbor more CDC45L protein, supporting the fact that DNA replication occur more frequently in cancer cell lines. Therefore, CDC45L is suggested to be one of the markers in cancer progression [284].

Seven proteins formed a cluster that is enriched with circadian rhythm GO biological process term (*p-val* = 3.62E-07). CLOCK is previously associated with Type 2 diabetes [299] and diabetes induced vessel injury in the eye [300]; these proteins involve in maintaining circadian rhythm to ensure the temporal organizations of pathophysiological processes, such as sleep, glucose homeostatis and fat metabolism, as well as the adaptation to environmental changes [301]. Transcriptional activator of the molecular clock consists of a heterodimer between either the CLOCK or the neuronal PAS domain protein 2 (NPAS2) and the aryl hydrocarbon receptor nuclear translocatorlike protein (ARNTL) that binds to E-box elements in the promoter of three period (PER) and two cryptochrome (CRY) genes, thereby activating their transcription [302]. A number of other genes, such as nuclear receptor subfamily 1, group D, member 1 (NR1D1), and timeless homolog (Drosophila) (TIMELESS), are involved in the feedback loops. Type 2 diabetes is associated with increased incidence of hypertension and disrupted blood pressure (BP) circadian rhythm [303].

The JAK-STAT signaling pathway is represented with CNTF, IFNG, IL2, IL6, IL12A, OSM, SOCS3, STAT1, STAT2 and STAT4, where seven of them have well characterized roles in the pathophysiological processes of the disease. JAK-STAT signaling pathway transmits extracellular signals from a variety of cytokines, lymphokines and growth factors to the nucleus. JAK-STAT activation stimulates cell proliferation, differentiation, migration and apoptosis. Besides its activation, suppressors of cytokine signaling (SOCS) proteins inhibit JAK-STAT pathway [240]. SOCS proteins are also known as JAK-binding protein, STAT-induced STAT inhibitor. The members of this family of proteins are responsible for establishing inducible negative regulations of cytokine signaling via inhibition of JAK-STAT pathway. SOCs proteins function in binding to specific phosphorylated tyrosine residues in cytokine receptors, hence inhibit their action or mediate their degradation. Cytokine-induced activation by STATs is a major mechanism of SOCS induction; however, there is increasing evidence that SOCS expression can also be induced by other stimuli, such as lipopolysaccharide and insulin [304]. Studies also suggest that SOCs proteins are able to inhibit leptin and insulin signaling pathways, and have impact on the energy and glucose homeostasis [65]. The IL6 cytokine family, including interleukin 6 (IL6), oncostatin M (OSM), ciliary neurotrophic factor (CNTF) use a common receptor subunit for signal transduction. The receptors for IL6, OSM and CNTF are dimerized, or heterodimerized, which brings corresponding JAK protein to close proximity, where autophosphorylation is attained. The activation of the JAK proteins creates docking sites for STAT proteins, where they are phosphorylated prior to translocation to nucleus to induce gene expression [240]. Due to the regulatory roles of IL6, OSM, CNTF and SOCS3 in JAK-STAT signaling pathway, the conjunction of this pathway with insulin resistance and glucose transport may modulate the biological processes that affect the pathophysiological conditions of the disease.

The glucose transport mechanism (*p-val* = 9.39E-04) is represented in the condensed network with the proteins, INS, IRS1, IRS2, PIK3R1 and AKT (Figure 5.8). Indeed, the JAK-STAT signaling pathway is connected to response to cholesterol (*p-val* = 5.63E-04), where TGFB1, TGFBR1, TGFBR2 and SMAD2 constituted in this process. Transforming growth factor β1 (TGFB1) is a ubiquitously expressed in humans, its levels are up-regulated in some cancers, and play important physiological roles in tissue regeneration,

cell differentiation, embryonic development, the regulation of the immune system and apoptosis [305]. It is known that hyperglycemia is one of the major factors for TGFB1 expression, and patients with diabetes have higher levels of TGFB1 than healthy people. TGFB1 induces the phosphorylation of the TGF-β receptor activated protein (SMAD2), and its responsiveness is modulated by cholesterol by binding TGFB receptors [306].

One distinctive cluster in the TDFN belongs to six proteins function in prostaglandin metabolic process (*p-val* = 9.39E-04), where PTGES2 was one of the core proteins, i.e. has been associated with Type 2 diabetes [307]. PTGES, PTGDS, PTGIS, PTGS1 are members of prostaglandins, which are generally considered to be potent pro-inflammatory mediators. Thromboxane A synthase 1 (TBXAS1) catalyzes the conversion of prostaglandin H2 to thromboxane A2, a potent vasoconstrictor and inducer of platelet aggregation. Thromboxane A2 is well-characterized for its role in modulating hemodynamics and cardiovascular function. It is a potent mediator of platelet shape change and aggregation. Increased thromboxane synthesis has been linked to cardiovascular diseases including acute myocardial ischemia and heart failure, and inhibition of platelet thromboxane production by aspirin is widely used to reduce the risk of myocardial infarction. PTGD2 and prostacyclin (PTGIS) also mediate a number of other effects including inhibition of platelet aggregation, smooth muscle relaxation and contraction, vasodilation and vasoconstriction. PTGIS functionally opposes the effects of thromboxane A2 and has been shown to specifically inhibit platelet activation [308]. Prostaglandin E2 synthases (PGES and PTGES2) are a group enzymes that are engaged in paracrine signaling system. Recent studies reported PTGES1 involvement in inflammatory pain and ischemic brain injury, and suggested as potential pharmacological targets for stress-related disorders [309, 310]. Although, a direct relation between prostaglandins and Type 2 diabetes has not yet been established, the report revealing the supplementary role of PTGDS in muscle and adipose glucose transport supports these findings that prostaglandins may have reminiscent function in the disease, where further assessments are required.

Four proteins involved in blood coagulation (*p-val* = 8.72E-06) are linked to glucose transport mechanism through apolipoproteins (APOB, APOA2 and APOA1), as expected. However, one module located at the conjunction of glucose metabolism, proteasome

activity, JAK-STAT signaling and DNA replication unit was enriched with dorsal/ventral neural tube patterning (*p-val* = 4.09E-04) constituting forkhead box A2 (FOXA2), smoothened homolog (SMO), sonic hedgehog homolog (SHH), patched homolog 1 (PTCH1) and glioma-associated oncogene family zinc finger 2 (GLI2) proteins imply the modulating function of Hedgehog (Hh) signaling in diabetes. Fat-body-specific activation of Hh signaling inhibits fly fat formation, which is uniquely observed in invertebrates, has been thought to be conserved from invertebrates to vertebrates. Hedgehod cascade components (SMO) and negative regulators (PTCH1) and transcription factor (GLI) are shown to be expressed in developing and adult fat tissues. Results from Ptch1 heterozygous mice (Ptch1+/–) suggested that elevated Hh signaling results in defective glucose tolerance [311]. Also, reduced levels of expression of SMO and GLI, and elevated levels of PTCH1 expression was observed during adipogenesis, suggesting that Hh signals might block mammalian adipogenesis and alter obesity [312], hence future studies will need to clarify whether a relation between disrupted Hh signaling and development of Type 2 diabetes exists.

5.2.4. Disease Interventions Derived from TDFN

The proteins involved in the modular form in TDFN has been implicated to be present in many other complex diseases, therefore the diseases can be linked to each other through shared partners, indicating the shared underlying mechanisms and proteins. To elucidate the shared partners, the proteins in condensed form of TDFN were used to calculate the disease overlapping score among diseases. The manual curation and elimination of the MeSH terms initially yielded 3630 disease terms, and subsequently incorporated with OMIM database records, forming a disease classification scheme, which was also used scoring the functional modules enumerated from TDFN. The 172 proteins present in condensed map were found to be related with 340 disease terms. According to the shared proteins, the disease overlapping score was calculated for each pair of diseases, and significantly associated 132 diseases sharing at least two proteins were selected (*p-val* < 2.00E-03). The resulting diseases and their linkages are presented in Figure 5.9. The selected disease pairs are tabulated in Table 5.8. The corresponding disease classes were shown in parenthesis, DO represents the disease overlapping score and *p-val* indicates the significance of the linkage.

As shown in Table 5.8 and Figure 5.9, glucose intolerance, insulin resistance and obesity, all of which are nutritional and metabolic disorders, are connected to each other through INS, IRS1 and IRS2, as expected. However, addition to these, CNTF, FOX2A, PPARGC1A, PTPN1 and TGFB1 establish the connection between insulin resistance and obesity. Other than CNTF, all proteins are present in the core set of proteins. Ciliary neurotrophic factor encoded by CNTF is a polypeptide hormone facilitating and promoting neurotransmitter synthesis. Despite its role in neurotransmitter activity, recent studies support the finding that CNTF is significantly associated with insulin resistance. CNTF reverses insulin resistance by increasing fatty acid oxidation, increases muscle glucose uptake through phosphoinositide-3-kinase (PI3K) signaling and AMP-activated protein kinase (AMPK) pathway [313, 314], imitates biological actions of leptin [315] and signals though JAK-STAT cascade, increases SOCS3 expression [316]. The obvious existence of CNTF in insulin resistance, its function in signaling pathways for glucose uptake rate should be investigated to clarify its therapeutic effect as a treatment for obesity-related Type 2 diabetes.

The co-existence of insulin resistance, obesity, glucose intolerance, Type 2 diabetes, hyperglycemia and hypertension through shared proteins is represented in a scheme, where shared proteins are located at the intersections, and the proteins that are not present in the core network are indicated with red (Figure 5.10). The links established through the proteins in TDFN between the diseases were also presented in Table 5.8. IRS1 is observed to be common in these six diseases, whereas INS is shared among nutritional diseases. In this scheme, the non-core proteins, namely, CNTF, IFNGR1, NRIP1, PTGIS, CREBBP, IGF1R and JAK3, which have not previously reported associations with Type 2 diabetes were indicated in red. As explained above, CNTF is involved and signals through JAK-STAT, AMPK and PI3K signaling pathways and regulates the glucose uptake rate leading to glucose intolerance and insulin resistance.

IFNGR1, which is not present in core proteins, was found to be associated with hypertension (Figure 5.10). IFNGR1 encodes a heterodimeric interferon gamma receptor, which activates interferon gamma (IFNG) in cellular responses through non-ligand-binding. IFNG is a cytokine that elicits the production of macrophage mediators, which

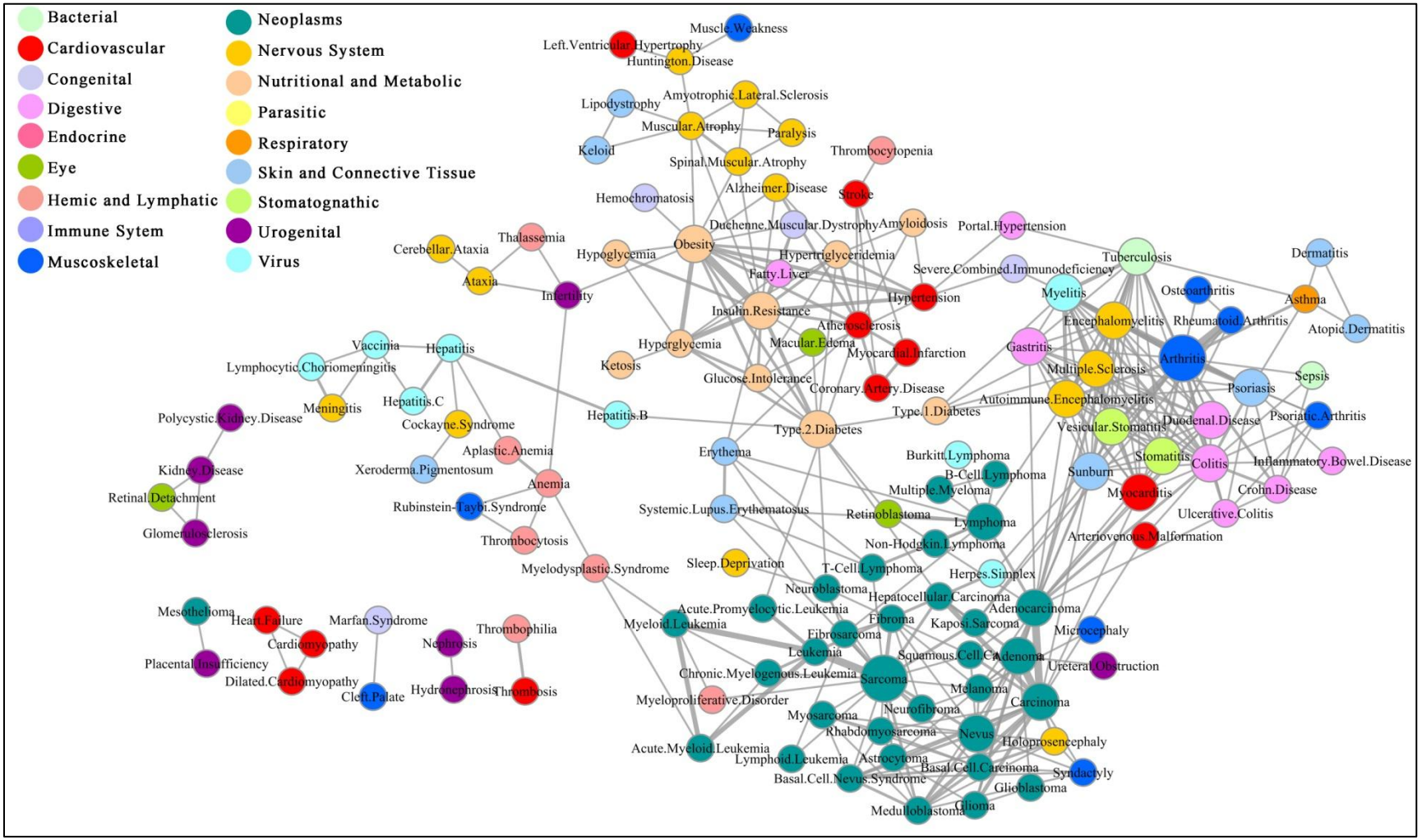

Figure 5.9. Disease associations derived through TDFN

Table 5.8. Selected pairs of disease terms derived from TDFN displaying significant associations in terms of shared genes

DiseaseTerm1	**DiseaseTerm2**	***DO***	***p-val***	**Shared genes**
Glucose Intolerance (Nutritional and Metabolic)	Obesity (Nutritional and Metabolic)	0.250	2.25E-05	INS, IRS1, IRS2
Glucose Intolerance (Nutritional and Metabolic)	Insulin Resistance (Nutritional and Metabolic)	0.231	9.80E-05	INS, IRS1, IRS2
Insulin Resistance (Nutritional and Metabolic)	Obesity (Nutritional and Metabolic)	0.615	0.00E+00	CNTF, FOXA2, INS, IRS1, IRS2, PARGC1A, PTPN1, TGFB1
Insulin Resistance (Nutritional and Metabolic)	Muscular Atrophy (Nervous System)	0.154	4.33E-03	CNTF, TGFB1
Muscular Atrophy (Nervous System)	Obesity (Nutritional and Metabolic)	0.167	5.31E-03	CNTF, TGFB1
Obesity (Nutritional and Metabolic)	Spinal Muscular Atrophy (Nervous System)	0.182	1.86E-04	CNTF, TGFB1
Insulin Resistance (Nutritional and Metabolic)	Spinal Muscular Atrophy (Nervous System)	0.167	6.64E-04	CNTF, TGFB1
Insulin Resistance (Nutritional and Metabolic)	Alzheimer Disease (Nervous System)	0.167	3.28E-04	INS, PTGDS, TGFB1
Alzheimer Disease (Nervous System)	Atherosclerosis (Cardiovascular)	0.375	8.18E-14	INS, PTGDS, TGFB1
Duchenne Muscular Dystrophy (Congenital)	Obesity (Nutritional and Metabolic)	0.200	2.45E-06	PPARGC1A, TGFB1
Insulin Resistance (Nutritional and Metabolic)	Duchenne Muscular Dystrophy (Congenital)	0.182	9.07E-06	PPARGC1A, TGFB1
Fatty Liver (Digestive System)	Duchenne Muscular Dystrophy (Congenital)	0.667	0.00E+00	PPARGC1A, TGFB1
Type 2 Diabetes (Nutritional and Metabolic)	Neuroblastoma (Neoplasms)	0.182	5.31E-03	CDK2NA, PTGES2

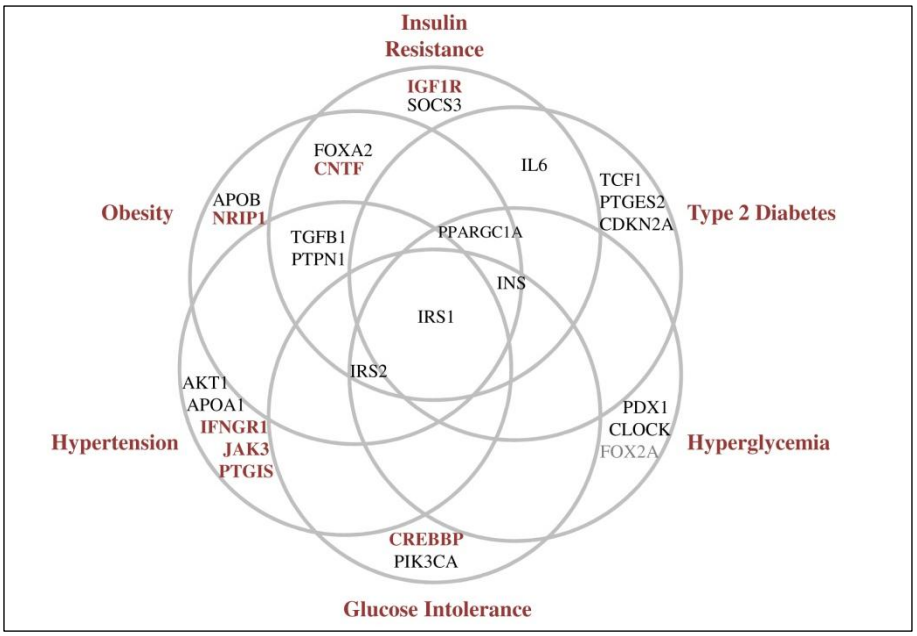

Figure 5.10. A representative scheme showing the links between the nutritional and metabolic disorders, with hypertension derived from the protein linkages in TDFN

induces leukocyte adhesion molecules, chemokines and increases antigen-presenting capacity by macrophages and endothelial cells. It also participates in the progression of atherosclerotic plaque, where it affects importantly all local cell types: endothelial cells, smooth muscle cells, and macrophages. The molecular pathways involved in the cellular response to interferon IFNG have been subject to intense research due to their importance in host defense against infection and disease. Recent studies report that IFNG interferes with JAK-STAT signaling by mediating nuclear transportation of STAT1 [317, 318]. The binding of IFNGR1 and IFNG initiates tyrosine phosphorylation events, catalyzed by JAK1 and JAK2 kinases that result in the phosphorylation and binding of STAT1 to the cytoplasmic domain of IFNGR1. Later, IFNG, IFNGR1 and STAT1 are transported to the nucleus of cells as a complex [319]. Additionally, IFNG is reported to promote inflammation in fat tissue [317], which provides compelling evidence that IFNGR1 may modulate the processes linking obesity and inflammation.

Nuclear receptor interacting protein 1 (NRIP1), also known as RIP140, is a nuclear protein that specifically interacts with the hormone-dependent activation domain of nuclear receptors, such as estrogen receptor. Nuclear receptors have a crucial role in lipid and glucose homeostasis by regulating the expression of gene networks in metabolic tissues.

RIP140 suppresses the expression of gene clusters that are involved in lipid and carbohydrate metabolism, inhibits glucose uptake and facilitates the expression of genes promoting energy expenditure. Therefore, the functional interplay between transcriptional activators and RIP140 is an essential process in metabolic regulation [320]. However, two recent studies reported contradictory results; the function and expression level of RIP140 was not correlated with obesity [321] but lower gene and protein expression levels of RIP140 was observed in obese subjects [322]. Although controversial studies were reported on the modulating effect on RIP140 in obesity, RIP140 has an apparent modulating role in lipid and carbohydrate metabolism, and yet no studies were conducted to investigate the subsequent effects of RIP140 on Type 2 diabetes.

Prostacyclin (PTGIS) a member of the cytochrome P450 superfamily of enzymes, which catalyzes the reactions involved in cholesterol synthesis. PTGIS also mediates a number of other effects including inhibition of platelet aggregation, smooth muscle relaxation and contraction, vasodilation and vasoconstriction [308], hence hypertension [323]. Yet, no studies were conducted on querying the effect of PTGIS on insulin or glucose metabolism, the protein was suggested to be one of the candidate genes linking obesity and diabetes with inflammation according to the analysis performed on overlapping chromosomal regions [324].

The associations of insulin resistance, obesity, muscular atrophy, amyotrophic lateral sclerosis and paralysis through shared proteins is presented in a scheme, where shared proteins are located at the intersections, and the proteins that are not present in the core network are indicated with red (Figure 5.11). The links established through the proteins in TDFN between the diseases were also presented in Table 5.8. Muscular atrophy is the loss of mass and strength that progresses with medical conditions such as cancer and aging. Spinal muscular atrophy, on the other hand is a neuromuscular disease that is characterized by degeneration of motor neurons, leading to progressive muscular atrophy. Amyotrophic lateral sclerosis (ALS) is a fatal progressive neurodegenerative disorder that is also characterized by the degeneration of motor neurons in the spinal cord, brainstem and motor cortex, progressive muscle weakness resulting in paralysis. Remarkably, ALS and SMA are characterized by marked weight loss. Besides its promising functions in insulin resistance, CNTF was one of the first neurotrophic factors administered in clinical trials for

ALS patients, however due to the undesired adverse effects CNTF is no longer considered as a potential target in the treatment of ALS [325]. However, contradictory results were recorded on the participation of CNTF in the progression of ALS. One study indicating that CNTF genotype affects susceptibility to disease initiation and acts as a modifier but not disease progression [326], however another claimed no effect [327]. In both SMA and ALS, altered expression levels of CNTF were observed in patients as well as animal models [325]. As a consequence, a vast number of studies have further substantiated to determine the significance of CNTF linking obesity, insulin resistance, ALS and SMA.

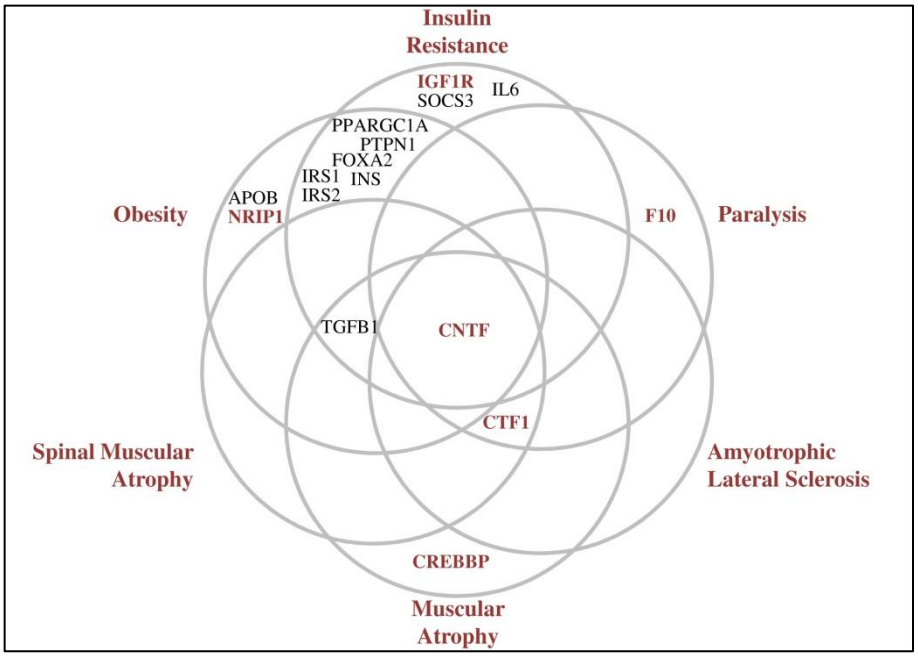

Figure 5.11. A representative scheme showing the links between insulin resistance, obesity, muscular atrophy, amyotrophic lateral sclerosis and paralysis derived from the protein linkages in TDFN

CTF1 has been associated with SMA, ALS and paralysis, but not with obesity and insulin resistance (Figure 5.12). Cardiotrophin-1 (CTF1) is another protein that is not previously associated with Type 2 diabetes and insulin resistance. CTF1 is a muscle-derived member of IL6 family cytokine, exerts its cellular effects by interacting with the glycoprotein 130 [328], and is highly expressed in embryonic skeletal muscle and secreted by myotubes [325]. It promotes the survival of cultured embryonic mouse and rat motor neurons. Circulating levels of CTF1 were associated with glucose levels, where glucose triggers CTF1 expression in adipocytes [328]. In SMA, CTF1 has a slowing down effect

on the progression of the disease [329] and decreased motor impairment was observed in ALS patients by CTF1 treatment [330].

CREBBP is found to be linked with muscular atrophy and not associated with Type 2 diabetes in previous studies. CREB binding protein is encoded by CREBBP gene, which is ubiquitously expressed and is involved in the transcriptional co-activation of many different transcription factors. Generally the deficiency of the protein resulted in Rubinstein-Taybi syndrome, which is characterized by facial abnormalities and mental retardation. CREBBP has also been implicated to play a central role in spinal and bulbar muscular atrophy, which is a neurodegenerative disorder caused by toxic effects of polyglutamine tracts [331]. In animal models, heterozygous CREBBP deficiency results in increased effects of hormones such as adiponectin and leptin, preventing obesity and insulin resistance. Hence, CREBBP functions as a “master-switch” between energy storage and expenditure through inhibition or activation of leptin and adiponectin pathways [332].

The links between insulin resistance, Type 2 diabetes, fatty liver and Duchenne Muscular Dystophy (DMD) was established through PPARGGC1A and TGFB1, both of which are known to be involved Type 2 diabetes. DMD is a fatal recessive X-linked form of muscular dystrophy characterized by rapid progression of muscle degeneration, leading to loss of movement (Table 5.8). Although these diseases seem divergent from each other, the cross link is established through transforming growth factor beta 1 (TGFB1) signaling coupled with connective tissue growth factor (CTGF). According to human biopsies taken from DMD patients, the co-localization of TGFB1 and CTGF in these samples suggests that CTGF expression is regulated with TGFB1, and these two proteins are implicated to have roles in DMD [333]. Remarkably, elevated levels of TGFB1 activity is observed in muscle wasting, which is a condition that is observed in DMD [334]. Another possible link is established through PPARGC1A, which is a transcriptional co-activator that coordinates the genes involved in energy metabolism. This protein is also involved in controlling blood pressure, regulating cellular cholesterol homoeostasis, and the development of obesity. However, the maintenance / elevation of PPARGC1A levels with statins have reported to be enhances the treatment for DMD. Therefore, PPARGC1A has been implicated to have a potential treatment of patients with muscle wasting, Type 2 diabetes and muscular dystrophies [335].

6. ALZHEIMER'S DISEASE RELATED FUNCTIONAL LINKAGE NETWORK

6.1. Construction of Alzheimer's Disease Related Functional Linkage Network

The construction of Alzheimer's disease functional linkage network was started with the proteins encoded by 244 genes (c-proteins) collected from five different sources; Human Experimental/Functional Mapper [336], AlzGene [337], GeneCards [338], NCBI databases and genome-wide associations reported in literature [305, 339-341]. The validation of one protein in at least two resources was accepted sufficient enough to consider these proteins as c-proteins. The linkages between the proteins were extracted from STRING database v8.1 to establish functional associations among proteins [158]. The core set of proteins were incorporated with the first neighbors to capture putative proteins that have potential links with the disease. To ensure that maximum number of core proteins is covered in a network housing optimum number of proteins, the confidence score for the functional linkages was selected by constructing various networks with confidence scores varying from 900 to 990. According to these selection criteria, as shown in Table 6.1., the confidence score threshold was set as 960; selection of such a stringent threshold confidence score for the linkages enabled us to construct a coherent functional linkage network. Although, all c-proteins are covered for a confidence score 900, such a lenient criteria might lead to irrelevant linkages, therefore a confidence score 960 was set as the threshold to eliminate the linkages while keeping the sufficient amount of core proteins in the network. The removal of orphan nodes that is not connected to the giant component resulted in a network of 1587 nodes (proteins) and 7785 edges (functional linkages). In this network, among the 1587 proteins, 226 of them are c-proteins, which have previously defined associations with the disease. These 226 proteins form 14.24 per cent of the functional linkage network. The functional linkage network is entitled as Alzheimer's disease related functional linkage network (ADFN).

Table 6.1. Selection of confidence score for the construction of ADFN

Confidence score	**900**	**920**	**940**	**960**	**980**	**990**
N	2848	2376	2026	1603	1044	764
l	28784	17522	11974	7816	3531	2049
c-proteins	244	243	237	231	203	178
<*k*>	10.11	7.37	5.91	4.88	3.38	2.68
% constitution	8.57	10.23	11.70	14.41	19.44	23.30
% coverage	100.00	99.59	97.13	94.67	83.20	72.95

6.1.1. Analysis of ADFN Topological Properties

To initiate the analyses, the topological features of ADFN were investigated to ensure that the network is scale-free and hence biologically relevant. The ADFN were analyzed in terms of the degree, *k*, clustering coefficient, *C*, and betweenness, *b*, for each node. The average degree <*k*>, average clustering coefficient <*C*> and average betweenness <*b*> for the network were calculated to be 9.81, 0.32 and 4433.09 respectively. The cumulative degree distribution $n(k)$ (Figure 6.1) showed that the network is scale-free with degree exponent, γ = 2.52 (R^2 = 0.8879). In addition, clustering coefficient distribution $C(k)$ (Figure 6.2) with respect to degree distribution, also showed the higher degree proteins have lower clustering coefficients, hence providing additional evidence that the network of interest exhibits a scale-free behavior. The betweenness distribution $b(k)$ was also well-characterized with Power law scaling (Figure 6.3), meaning that many proteins are located at the periphery while a few proteins are located at the center of the network, and have central roles in maintaining the communication within the network. The effect of removing the nodes with higher betweenness is similar to that of excluding a hub protein from the network and proteins displaying high betweenness are reported to be likely essential and evolutionary conserved.

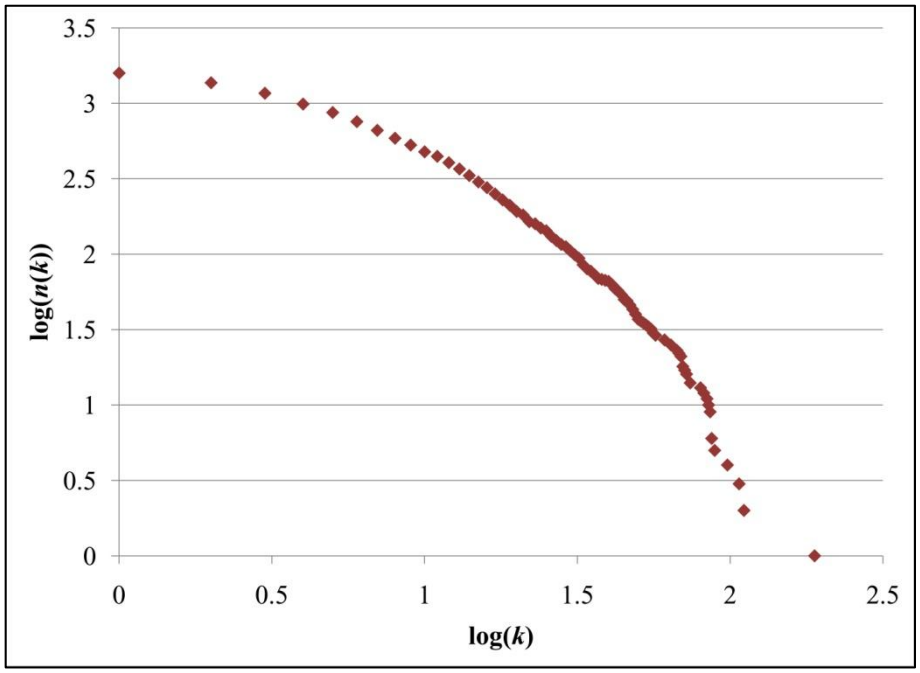

Figure 6.1. Cumulative degree distribution $n(k)$ of the network with respect to degree distribution in ADFN

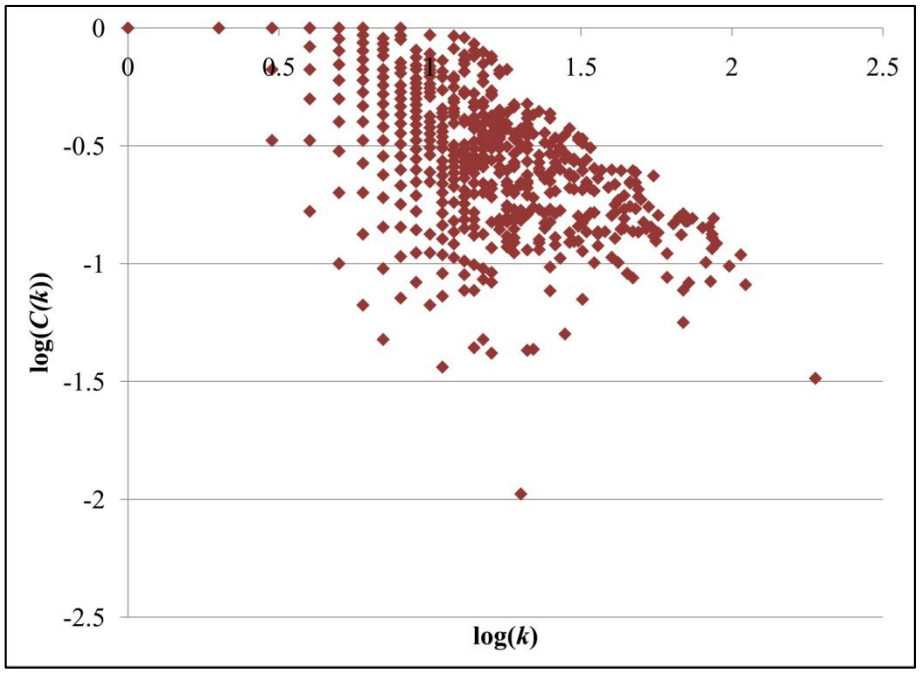

Figure 6.2. Clustering coefficient distribution $C(k)$ with respect to degree distribution in ADFN

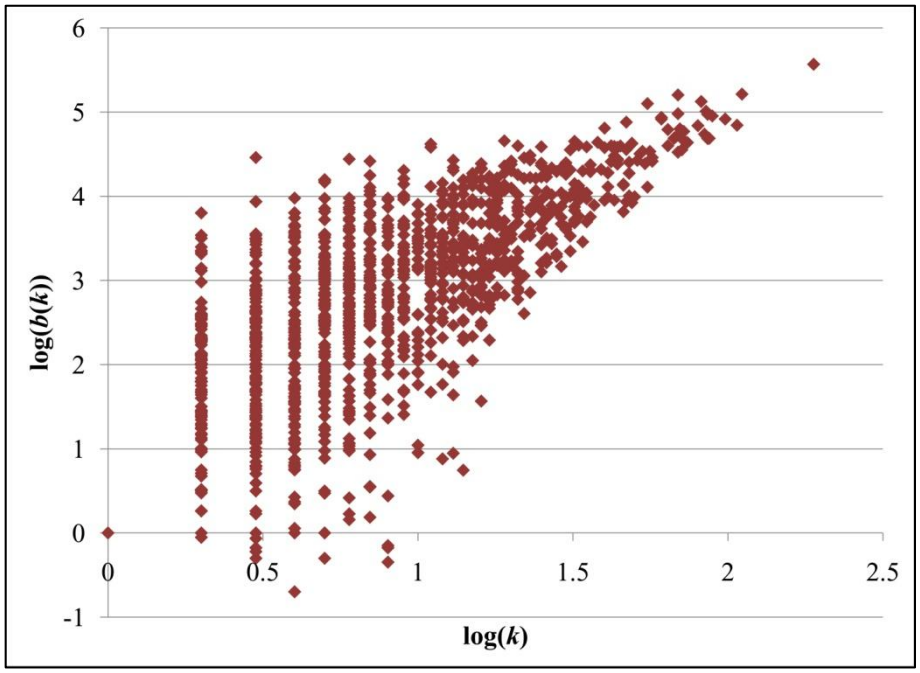

Figure 6.3. Betweenness distribution $b(k)$ with respect to degree distribution in ADFN

6.1.2. Selection of Hub Proteins in ADFN

Hub protein in network play crucial roles in the network s while maintaining stability and providing intercommunication. Therefore, these key players in the ADFN were selected to understand the governing major biological systems present in the network. The hub protein selection criterion is based on the accumulative degree and betweenness distributions with respect to degree. These two measures were incorporated to identify hub proteins. To select the most informative entities in the network, the proteins that are listed at the top two per cent of the accumulative degree and accumulative betweenness distributions were entitled as hub proteins. The accumulative degree distribution graph shows that 13.7 per cent of the proteins in ADFN have single linkage (Figure 6.4), whereas 22.4 per cent of the proteins were located at the periphery (Figure 6.5), hence no contribution to the communication in the network.

The intersection of the highest top two per cent accumulative degree and betweenness distribution yielded 24 proteins, which have anticipated playing vital roles in the network. The degree of hub proteins ranges between 55 to 189 with an average of 83 interactions, whereas the betweenness (log(b)) ranges between 4.638 to 5.568. 11 out of 24

hub proteins are previously associated with the disease, listed in c-proteins. The hub proteins identified in the network are presented in Table 6.2.

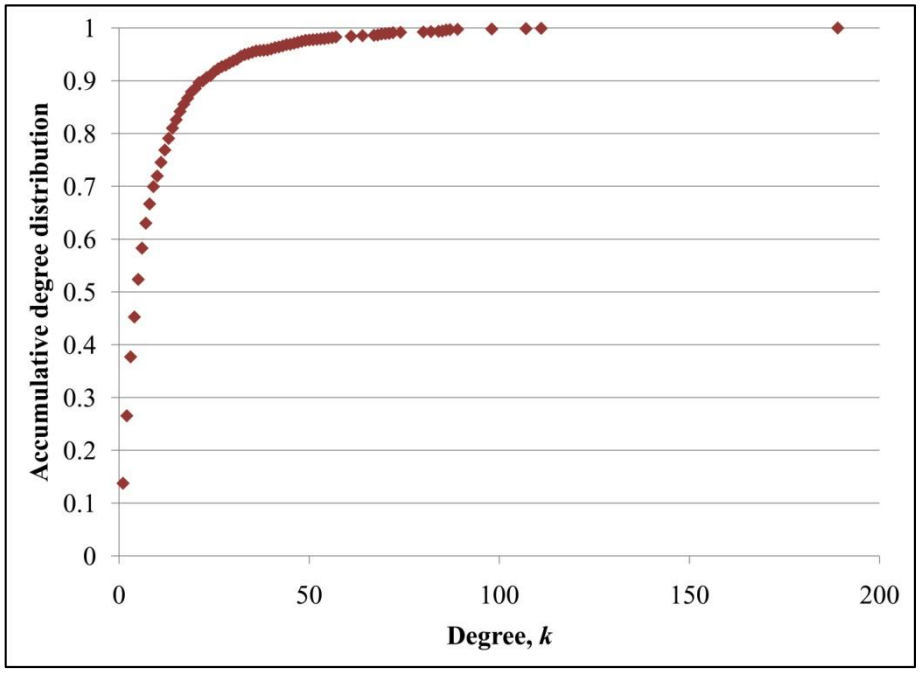

Figure 6.4. Accumulative degree distribution of ADFN with respect to degree distribution

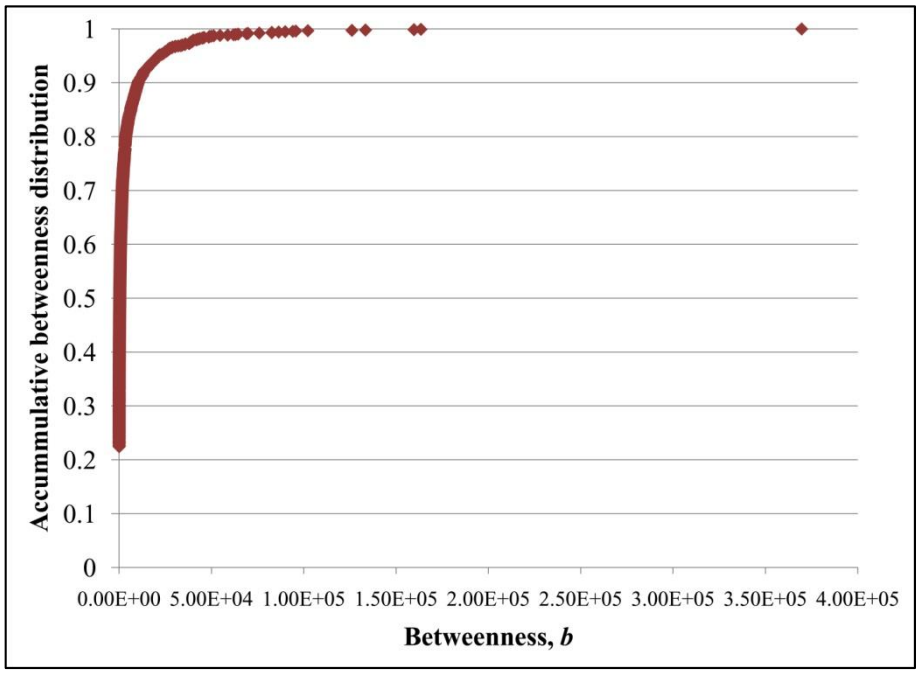

Figure 6.5. Accumulative betweenness distribution of ADFN with respect to degree distribution

Table 6.2. Topological properties of hub proteins in ADFN

Symbol	**Name**	**Degree, *k***	**Betweenness, log(*b*)**	**Clustering coefficient, *C***
TP53	tumor protein p53	189	5.568	0.03259
AKT1	RAC-alpha serine/threonine-protein kinase	111	5.213	0.081409
IFNG	interferon, gamma	107	4.844	0.108975
SRC	proto-oncogene tyrosine-protein kinase Src	98	4.917	0.097623
JUN	transcription factor AP-1	89	4.954	0.121808
IL2	interleukin-2 precursor	87	4.688	0.155573
EGFR	epidermal growth factor receptor	86	4.973	0.115732
GRB2	growth factor receptor-bound protein 2	86	4.698	0.128865
IL4	interleukin 4	86	4.686	0.133242
BCL2	B-cell CLL/lymphoma 2	85	5.009	0.084034
STAT3	signal transducer and activator of transcription 3	84	4.737	0.143144
MAPK8	mitogen-activated protein kinase 8 isoform JNK1 alpha1	82	5.125	0.101174
IL6	interleukin 6	80	4.839	0.142089
SHC1	SHC transforming protein 1	74	4.638	0.155498
CCND1	G1/S-specific cyclin-D1	72	4.663	0.152973
FYN	FYN oncogene related to SRC, FGR, YES	72	4.769	0.082942
INS	Insulin	69	5.203	0.077153
CASP3	caspase 3	69	4.981	0.056266
TGFB1	transforming growth factor, beta 1	68	4.709	0.132572
IL1B	interleukin 1, beta	67	4.689	0.157847
FOS	v-fos FBJ murine osteosarcoma viral oncogene homolog	64	4.792	0.147321
CTNNB1	Catenin	61	4.917	0.110383
ESR1	estrogen receptor 1	61	4.936	0.087432
PTGS2	prostaglandin-endoperoxide synthase 2	55	5.101	0.13064

The 24 hub proteins in ADFN have degrees, *k*, varying in the range of 55 (PTGS2 – prostaglandin endoperoxide synthase 2) to 189 (TP53 –tumor protein 53). TP53 is a tumor repressor and responds to a variety of cellular stresses that regulate cell cycle arrest, apoptosis and DNA repair. Gene Ontology (GO) terms enriched in the 24 hub proteins indicate the major cellular mechanisms that govern throughout the network. The biological processes GO categories significantly enriched in hub proteins (*p-val* < 1.41E-14) indicate that the selected hub proteins are involved in regulation to response mechanisms to external stimulus and signal transduction. The selected GO terms enriched in 24 hub proteins are listed in Table 6.3.

Table 6.3. Gene Ontology terms enriched in the hub proteins in ADFN

GOTerm	***p-val***	**Proteins**
GO:0010033 response to organic substance	5.46E-23	BCL2 IL6 ESR1 CCND1 INS GRB2 CASP3 FOS STAT3 JUN AKT1 IFNG PTGS2 TGFB1 SRC SHC1 CTNNB1 IL1B FYN EGFR
GO:0007243 intracellular protein kinase cascade	1.67E-17	IL6 INS IL2 MAPK8 GRB2 STAT3 IL4 AKT1 IFNG TGFB1 SRC SHC1 CTNNB1 IL1B FYN EGFR
GO:0023014 signal transmission via phosphorylation event	1.67E-17	IL6 INS IL2 MAPK8 GRB2 STAT3 IL4 AKT1 IFNG TGFB1 SRC SHC1 CTNNB1 IL1B FYN EGFR
GO:0009725 response to hormone stimulus	1.23E-16	BCL2 IL6 ESR1 CCND1 INS GRB2 FOS STAT3 AKT1 PTGS2 TGFB1 SHC1 CTNNB1 IL1B
GO:0009719 response to endogenous stimulus	3.99E-16	BCL2 IL6 ESR1 CCND1 INS GRB2 FOS STAT3 AKT1 PTGS2 TGFB1 SHC1 CTNNB1 IL1B
GO:0042325 regulation of phosphorylation	1.41E-14	BCL2 IL6 CCND1 INS IL2 CASP3 JUN IL4 AKT1 IFNG TGFB1 SHC1 IL1B EGFR

6.1.3. Enumeration of Functional Modules in ADFN

Genes exerting similar functions in biological processes have a tendency for sharing GO terms and localizing around dense groups in interaction networks [5]. These functionally linked proteins collaborate with each other and these entities are considered as functional modules. The proteins which have the maximum allowable number of interactions with each other, e.g. $Q = 1$, were derived, as explained in Methodology. The algorithm used in this study allocates the proteins into multiple clusters, rather than assigning proteins into distinct clusters. The algorithm produced 5046 functional modules, the size of the modules ranges from two to 11, with an average module size of 3.85, hence with the supporting information that modules consisting four or more members are biologically meaningful, the 2674 modules of size four and above were considered for further analyses. In the selected functional modules, out of 1587 proteins in ADFN, 783 (49.3 per cent) are represented, indicating that approximately half of the proteins participate the modular structure.

The hub modules play important roles in maintaining the communication in the network, therefore to understand the extent of the effect due to any loss of activity in these proteins, the distribution of hub proteins in the functional modules was calculated. As expected, because of their high numbers of interaction and power in maintaining the

communication, all of the hub proteins display the highest frequency in functional modules. Figure 6.6 shows the frequency of occurrence of the hub proteins in the functional modules with respect to the other proteins. The red diamonds in the graph indicates the hub proteins. For instance, IFNG is present in 287 modules with a frequency of occurrence 0.1073. The graph is presented in descending appearance of proteins from highest to lowest. The self distribution of each hub protein in functional modules according to the size of the modules is presented in Figure 6.7. IFNG is distributed among functional modules with varying sizes from four to nine. Notably, the largest modules in ADFN belong to ribosomal protein complexes.

The two largest modules consist of 11 proteins, where all members are ribosomal proteins (RPL27A, RPL8, RPL4, RPLP0, RPS29, RPS13, RPS3, RPL11, RPL30, RPS5, RPL5, RPS23), with RPL5 and RPS23 are interchangeable in two modules. These proteins function in the addition of amino acids to a nascent polypeptide, translational elongation (*p-val* = 4.39E-27). The impairment in ribosome function is associated with a decreased rate and capacity for protein synthesis, decreased ribosomal RNA and tRNA levels, and increased RNA oxidation. The complications in the protein synthesis were implicated to be one of the earliest neurochemical alterations in AD [342].

Two modules that have 10 proteins composed of IL2, IL2RA, IL2RB, IL2RG, JAK1, JAK3, LCK, STAT1, STAT3, STAT5A, STAT5B are involved in the JAK-STAT signaling pathway (*p-val* = 6.72E-08) and regulate T-cell differentiation (*p-val* = 5.42E-08). T-cells are a group of lymphocytes that maintains and increases the competence of immune system. During aging, telomeres of T-cells were shortened and observed to be correlated with the AD disease progression, indicating the prevalence of the disease among aged subjects [343]. The JAK-STAT pathway is critically involved in transmission of cytokine receptor signals in T-cells, thus influencing T-cell differentiation.

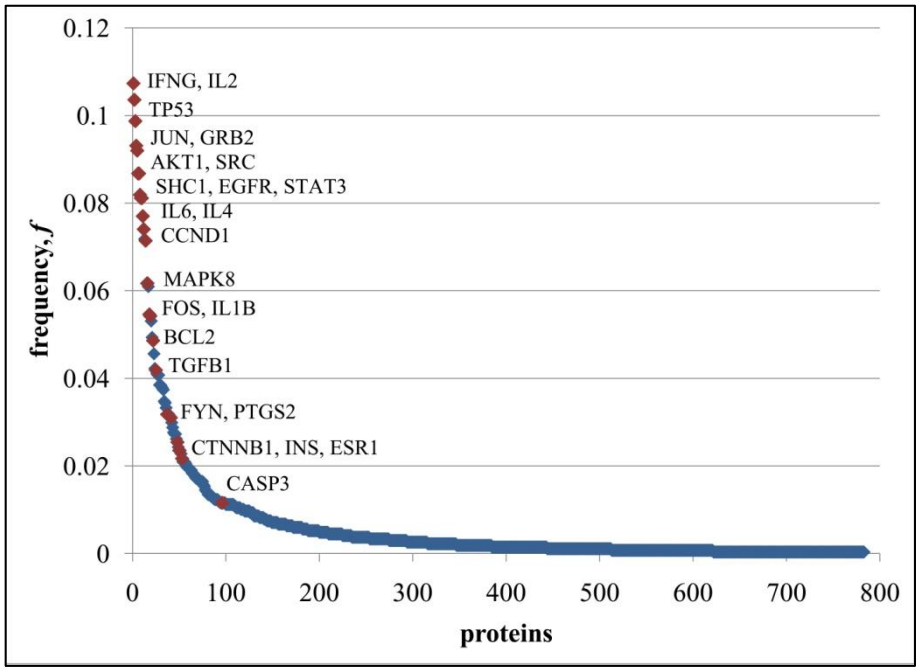

Figure 6.6. Frequency of presence of the proteins in ADFN in functional modules

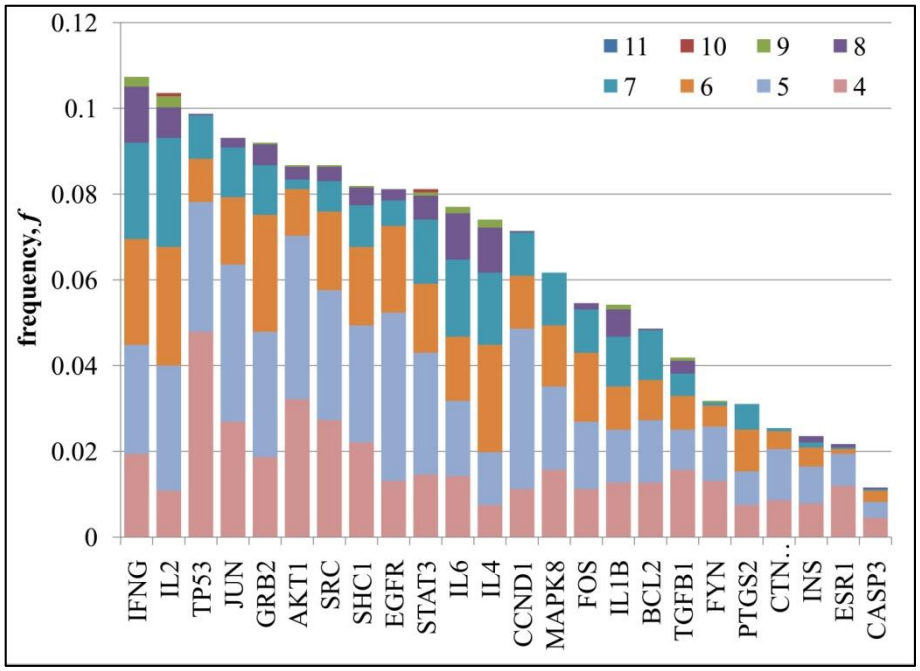

Figure 6.7. Self distribution of hub proteins in ADFN according to the size of the functional modules

6.2. Evaluation of ADFN Functional Modules

The functional modules were evaluated and subsequently scored in terms of the member's co-occurrence in similar pathways, co-localized, shared similar disease terms and exhibit similar expression patterns. The steps followed in this section are the same with the other diseases, computational layout is presented in Figure 3.2.

6.2.1. Evaluation and Scoring of ADFN Modules

The module enumeration algorithm yielded in 2674 functional modules, which makes the selection of most biologically informative ones impossible. Due to the mounting evidence that proteins function together to exhibit a single action often tend to participate in similar pathways, co-localized and share GO Terms and genes contributing to a disorder have increased tendency to be functionally related, the functional modules produced with the algorithm were evaluated in terms of sharing similar expression patterns, participation in pathways, co-localization and association with similar diseases. Incorporation of various independent data was anticipated to generate a deliberate and consistent evaluation method.

The scoring of the functional modules was initiated with the assembly of the data that will be incorporated. KEGG Pathway database was used to associate biological pathways, where the classification scheme involves 338 pathways, including major metabolic processes, as well as disease pathways [163]. LOCATE database was utilized to determine the co-localization information, where the proteins are assigned to 30 different subcellular compartments [164]. Medical Subject Headings (MeSH) [161] was incorporated with OMIM database to achieve the disease associations. The generation of the classification schemes was explained in Methodology. To validate the scoring scheme used in this study, and to show that the scores assigned to the functional modules could not be obtained by chance, the associations in the classification schemes were randomly shuffled 10^3 times and calculated score was compared with the distribution of the random scores.

The expression profiles of 22 postmortem AD samples at various stages of severity were analyzed for co-expression patterns using GSE1297 [166, 344]. The co-expression pattern in the functional modules was evaluated by calculating Pearson correlation coefficients (PCC) for each interaction in the network. The functional module average PCCs (PCC_{AVG}) were calculated from the interactions and this score is considered as the co-expression score of the modules.

Upon the scoring of the functional modules, four different scores were obtained: R_{KEGG}, R_{LOC}, R_{OMIM} and PCC_{AVG}. These scores were varied from zero to one, where one indicates the consistency in the module. To select the informative top scoring functional modules a non-linear model was proposed, where the parameters were estimated using Genetic Algorithm. The 10 artificially generated functional modules, five of which have the one for each scoring scheme, were planted in the population, and the population of the modules was evolved for 100 generations. The estimated model parameters were then used to evaluate the modules enumerated from ADFN and summarized in Table 6.4.

Table 6.4. Estimated nonlinear model parameters for ADFN

Classification	α	β
N	1.8077	1.7211
R_{KEGG}	1.4510	0.0310
R_{LOC}	1.5465	1.6720
R_{OMIM}	0.9455	1.7880
PCC_{AVG}	0.9712	1.5524

6.2.2. Assembly of Top Scoring Functional Modules in ADFN

One of the top scoring functional modules consists of 11 proteins, where all members are ribosomal proteins (RPL27A, RPL8, RPL4, RPLP0, RPS29, RPS13, RPS3, RPL11, RPL30, RPS5, RPS23). The members in this module involved in translation, and localized in nucleolus. The proteins in this module were associated with carcinoma and anemia. To maintain cellular homeostasis, all cells must continually synthesize new proteins. Ribosomes are specialized complexes composed of nucleic acids and proteins that are responsible for mediating all protein synthesis. The impairment in ribosome function is associated with a decreased rate and capacity for protein synthesis, decreased ribosomal

RNA and tRNA levels, and increased RNA oxidation. The complications in the protein synthesis were implicated to be one of the earliest neurochemical alterations in AD [342].

Table 6.5. Top scoring functional module in ADFN

Module1	R_{KEGG}	R_{LOC}	R_{OMIM}	PCC_{AVG}
Score	1.000	0.7074	0.7868	0.4823
p-val	0.00E+00	1.78E-14	0.00E+00	-
RPL27A, RPL8, RPL4, RPLP0, RPS29, RPS13, RPS3, RPL11, RPL30, RPS5, RPS23	Ribosome	Nucleolus	Carcinoma Adenocarcinoma Anemia	

The selection criteria for the number of top scoring functional modules that will be assembled were based on the number of proteins, clusters and hub proteins. Since, the motivation of such as assemble is to derive the underlying shared mechanism; the number of clusters obtained through incorporation of the overlapping members of the selected functional modules is an important parameter. Addition to this, the number of core proteins in the network is another parameter that should be considered to keep the association with the disease. Hence, the members of top scoring 150 functional modules were selected to be incorporated to construct a condensed functional linkage network. The basis of the criteria were presented in the Table 6.6.

Table 6.6. Selection criteria of the top scoring functional modules in ADFN according to the results obtained from Genetic Algorithm

Modules	***Top-25***	***Top-50***	***Top-75***	***Top-100***	***Top-150***	***Top-250***
N	58	101	142	181	269	332
l	199	357	503	628	1007	1377
c-proteins	10	16	20	28	44	64
hub proteins	0	0	0	0	3	7
clusters	6	6	7	9	5	5

Among 2674 modules, the overlapping members of 150 top scoring functional modules were assembled reflecting the biological processes involved in the disease. This condensed network, presented in Figure 6.8, contained 269 proteins, 225 of which were not previously reported with Alzheimer's disease, and 1007 interactions.

6.2.3. Indicators of Fundamental Cellular Processes in ADFN

To elucidate the fundamental biological processes involved in the disease and to determine the shared partners of the processes, the proteins in the condensed network was analyzed in terms of distinct Gene Ontology terms enriched in the map. 14 GO Terms with non-overlapping 142 members indicate the major cellular processes involved in the progress of the disease. The GO Terms enriched in this condensed map is represented in the Table 6.7. The proteins in the modular structure of ADFN are shown in Figure 6.8. The 142 proteins associated with a GO term in this map are colored according to biological process.

The 17 ribosomal proteins and elongation factor 2 (EEF2) are assembled in a modular structure, as shown in Figure 6.8. The ribosome plays an important role in translation elongation mechanism (*p-val* = 3.89E-13), and conserved throughout evolution. Alterations in ribosome resulted in elevated levels of protein oxidation [345]. Oxidative modification of proteins has a major effect on cellular homeostasis, upon the inhibition of protein function, deleterious fragmentation and formation of protein aggregates. In addition, the formation of highly oxidized proteins has negative effects on the proteolytic pathways, leading to inhibition of proteases mediating protein degradation. Besides, the oxidation of proteins may result in alterations in cells to synthesize new proteins. To decrease the protein toxicity due to protein oxidation, the oxidized proteins should be replaced with new ones. Both, removal of oxidized proteins and production of new proteins are affected, which result in a progressive accumulation of aberrantly functional proteins. This potentially lethal combination of increased levels of oxidized proteins, and decreased levels of protein synthesis, almost certainly contributes to the oxidative stress believed to occur in AD [342, 346]. The initial findings reported a direct influence of ribosomal proteins on the disease [347], and recent studies reported that ribosomal dysfunction is an early indication of neurodegeneration [342] and also observed in late stages of AD [348]. Elongation factor 2 (EEF2), which is a core protein present in the cluster, serves as a mediator of protein synthesis where it catalyses the translocation of the two tRNAs and the mRNA after peptidyl transfer on the 80S ribosome. EEF2 has been suggested to up-regulate the translation of protein tau in AD patients, which is responsible for the formation of neurofibrilliary tangles in the brain [349].

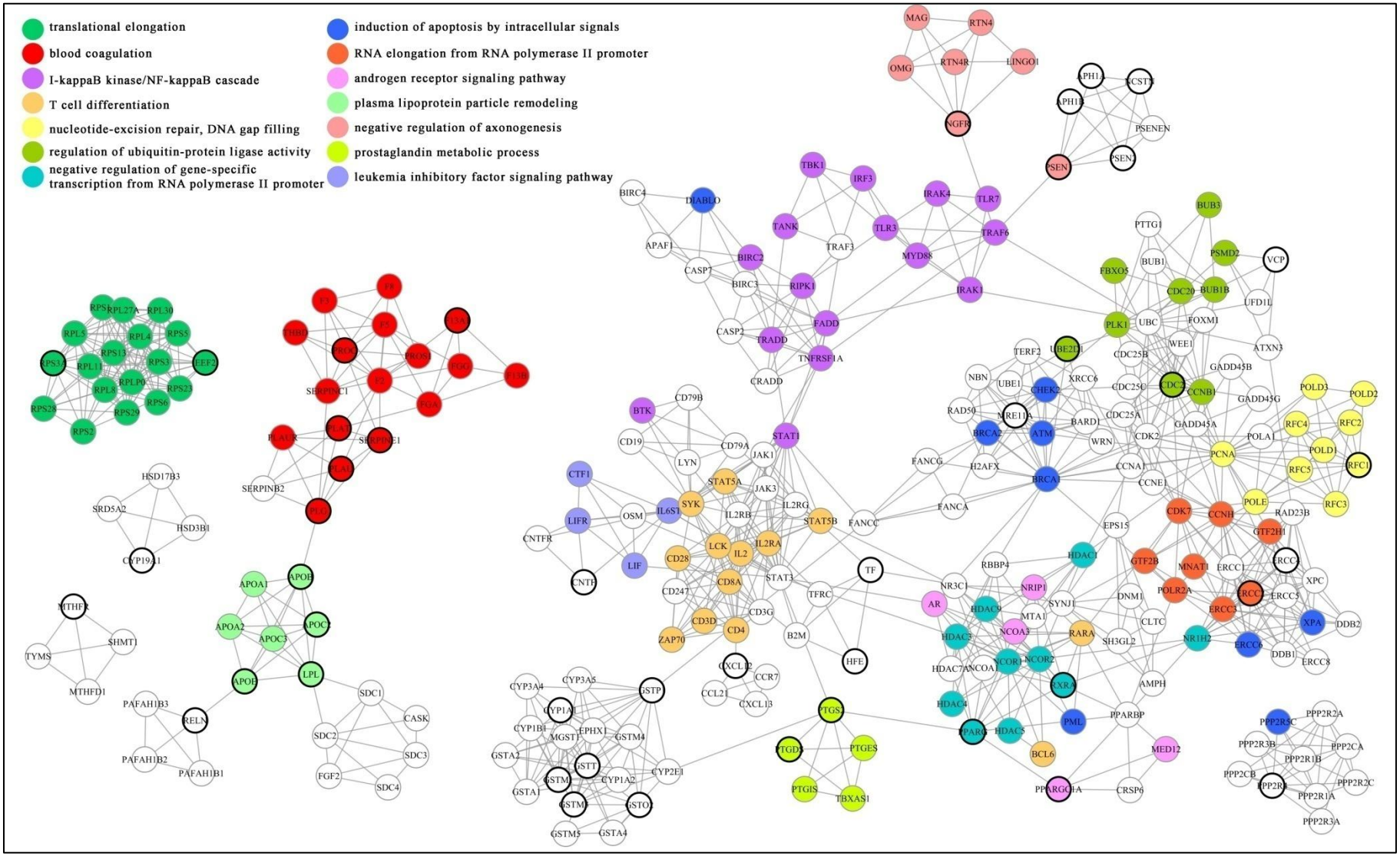

Figure 6.8. Condensed functional linkage network constructed from the top scoring functional modules in ADFN

Table 6.7. 14 distinct GO terms corresponding to separate cellular processes enriched in the ADFN condensed network

GO Term	***p-val***	**Genes**
GO:0006414 translational elongation	3.89E-13	RPS19 RPS13 RPL4 RPL27A RPL8 RPL5 RPS3 RPL11 EEF2 RPS23 RPS28 RPS6 RPL30 RPLP0 RPS5 RPS3A RPS2 RPS29
GO:0007596 blood coagulation	6.37E-10	F8 SERPINE1 F13B PLAU PLG PROC PLAT PROS1 FGG F13A1 F2 PLAUR SERPINC1 F5 THBD FGA F3
GO:0007249 I-kappaB kinase/NF-kappaB cascade	1.47E-07	RIPK1 IRAK4 IRAK1 TLR3 MYD88 TBK1 BIRC2 IRF3 FADD TLR7 TRADD TRAF6 TANK TNFRSF1A BTK STAT1
GO:0030217 T-cell differentiation	3.07E-07	STAT5B STAT5A IL2 IL2RA CD3D SYK CD4 CD8A RPS6 CD28 ZAP70 LCK RARA BCL6
GO:0006297 nucleotide-excision repair, DNA gap filling	8.55E-14	RFC4 RFC5 POLD1 RFC1 POLE RFC3 PCNA POLD2 RFC2 POLD3
GO:0051438 regulation of ubiquitin-protein ligase activity	1.96E-05	PLK1 FBXO5 PSMD2 CDC2 CCNB1 BUB1B BUB3 UBE2D1 CDC20 RPS27A
GO:0010553 negative regulation of gene-specific transcription from RNA polymerase II promoter	1.27E-06	HDAC4 NR1H2 HDAC5 NCOR1 HDAC1 PPARG NCOR2 HDAC3 RXRA HDAC9
GO:0008629 induction of apoptosis by intracellular signals	7.12E-06	PML BRCA2 DIABLO PPP2R5C CHEK2 BRCA1 ERCC6 ATM XPA
GO:0006368 RNA elongation from RNA polymerase II promoter	3.41E-05	GTF2H1 MNAT1 ERCC2 CDK7 CCNH ERCC3 GTF2B POLR2A
GO:0030521 androgen receptor signaling pathway	1.11E-07	AR MED17 MED1 MED12 NRIP1 PPARGC1A NCOA3
GO:0034369 plasma lipoprotein particle remodeling	6.97E-06	APOC3 LPL APOB APOA2 APOA1 APOE APOC2
GO:0050771 negative regulation of axonogenesis	1.86E-05	PSEN1 OMG NGFR RTN4R RTN4 LINGO1 MAG
GO:0006693 prostaglandin metabolic process	1.65E-11	PTGIS PTGS2 TBXAS1 PTGES PTGDS
GO:0048861 leukemia inhibitory factor signaling pathway	6.17E-05	IL6ST LIF CTF1 LIFR

The second largest cluster identified in this map belongs to blood coagulation (*p-val* = 6.37E-10), where 17 proteins are present. F13A1, PLAT, PLAU, PLG PROC, SERPINE1 have previously determined associations with the disease and F8, SERPINC1, F13B, PROS1, FGG, F2, PLAUR, F5, THBD, FGA, F3 are included in the network through extension. These proteins are well known members of blood coagulation pathway and may responsible for the formation of blood clots in the brain. As early studies

suggested a possible thrombin formation from prothrombin in AD [350], recent studies recorded that elevated levels of fibrinogen increases the risk of AD and vascular dementia [351, 352]. Fibrinogen, the precursor of fibrin, is normally removed from the brain by blood-brain barrier (BBB) to prevent any formation of blood clots by fibrin. The proper function of BBB is to tightly control the communication between the circulation and the central nervous system and to protect the brain from macromolecules in circulation. However, impaired BBB may result in the formation of agglomerates, as reported in animal models. This study also reported elevated levels of Plasminogen activator 1 (SERPINE1) contributes to the accumulation of Aβ [353]. Notably; the association of fibrinogen with AD is less strong than with vascular dementia [351, 352]. In fact, these two diseases have the similar form, where vascular dementia is the formation of vascular lesions in the brain, leading to cognitive impairment. Vascular dementia is generally associated with blood homeostasis, including hypertension. The hemostatic factors involved in vascular dementia may also be associated with AD, however the effect of hemostatic balance in AD might be less than the formation of Aβ-plaques and neurofibrillary tangles.

The third largest cluster identified in ADFN condensed map are the members of the NF-κB pathway (*p-val* = 1.14E-07). The direct link between presenilin 1 (PSEN) and nerve growth factor (NGFR) indicates the pronounced links between inflammation and AD. NF-κB pathway is one of the major signaling pathways that is activated in response to proinflammatory cytokines and Toll-like receptors (TLRs). Although the proteins, BIRC2, BTK, FADD, IRAK4, IRAK1, IRF3, MYD88, RIPK1, STAT1, TANK, TBK1, TLR3, TLR7, TRADD, TRAF6, TNFRSF1A have not previously associated with AD, the presence of NF-κB cluster in ADFN map suggests the involvement of inflammatory events in neurodegeneration in AD. Recent studies showed that Toll-like receptors (TLRs), having central roles in innate immune system, have also been linked to the pathogenesis of AD [114, 115]. The apparent functional association with presenilin 1 (PSEN) with tumor necrosis factor receptor associated factor 6 (TRAF6) and nerve growth factor (NGFR) with Toll-like receptor 7 (TLR7) observed in ADFN condensed network may indicate the coupling of presenilin with the NF-κB pathway. One recent study, revealing this interconnection, suggests that the TRAF6 binding site of PSEN1 mediates nerve growth factor (NGF)-induced association between PSEN1 and TRAF6. Disruption of the

interaction between PSEN1 and TRAF6 inhibits TRAF6 autoubiquitination and PSEN1-deficiency antagonizes NGF-induced I-κB degradation [354]. Therefore, the functional association between PSEN1, NGFR, TRAF6 and TLR7 revealed in this study is suggested to initiate the modulation of NF-κB signaling pathway. Considering the facts that Aβ is able to stimulate a NF-κB signaling pathway [355] and PSEN1 is critical for Aβ production [87], a realistic approach can be established to infer the biological processes linking inflammation to AD should start from the modulating effects on PSEN1 on NF-κB signaling pathway.

NF-κB signaling pathway cluster is linked to T-cell differentiation (*p-val* = 3.07E-07) and leukemia inhibitory factor signaling pathway (*p-val* = 6.17E-05). T-cell differentiation is implicated to the early indication of active immune response. In case of an immune response, T-cells proliferate and differentiate into Th1, Th2, Th17 and regulatory Th cells depending on the type of the effect. Interleukin 2 (IL2), produced in T-cells, mediates the cell cycle. After entering the cell cycle, the production of cytokines such as IFNG or IL4, IL5 and IL13 has been initiated. The most potent determinant of T-cell differentiation is the influence of IL12 and IL4 signaling, acting via STAT4 and STAT6. In Th1 cells, CD40 ligand (CD40L) is a potent inducer of IL12 gene expression that involves the activation of NF-B signaling pathway [356], whereas IL12 signaling through JAK-STAT pathway is crucial in the induction of Th1 differentiation [357]. JAK-STAT signaling pathway mediates the IL3 activation of microglia, which produces immune and inflammatory responses in the central nervous system [358].

APOC3, LPL, APOB, APOA2, APOA1, APOE, APOC2, which clustered together in this map, are involved in plasma lipid particle remodeling (*p-val* = 6.97E-06). Apolipoproteins (APOB, APOE, APOC2, APOC3) facilitate the removal of triglycerides in the form of chylomicrons from the circulation. A number of epidemiological studies reported that elevated levels of cholesterol contribute to the incidence of AD. Individuals with high levels of plasma cholesterol have an increased susceptibility to the disease, particularly with APOE4 genotype. In addition, high levels of low-density lipoprotein (LPL) were also observed [359]. This evidence includes *in vitro* studies indicating that cellular cholesterol levels modulate Aβ production [360] and animal studies demonstrating that cholesterol levels modulate Aβ accumulation in the brain [361]. Aβ, APOE and

cholesterol have been shown to co-localize in the core of fibrillar plaques in transgenic mice models of AD, supporting the suggestion that cholesterol and APOE are involved in fibrillar plaque formation [362]. Statins, which are drugs often used as a medication for atherosclerosis, inhibit cholesterol biosynthesis and result in accelerated clearance of plasma LDLs, hence tested for their validity in AD; results showed that statins exert a protective against AD [363]. Hence, these studies support the fact that there is a direct link between the altered cholesterol homeostasis and AD.

6.2.4. Disease Interventions Derived from ADFN

The intracellular biological processes involved in the modular form in ADFN has been implicated to be an indication of associations with many other complex diseases, therefore the diseases were linked to each other through shared partners. To delineate the complex disease associations through shared partners, the proteins in condensed form of ADFN were used to calculate the disease overlapping score among diseases. As explained in Methodology, the disease classification scheme was used to uncover the relations among the diseases. The 269 proteins participate the modular structure of ADFN were found to be linked with 362 disease terms. Subsequent calculation of the disease overlapping scores and validation of the linkages, the diseases that share at least two proteins (*p-val* < 2.00E-03) have been assembled, as shown in Figure 6.9 and selected disease pairs are listed in Table 6.8. The corresponding disease classes were shown in parenthesis, *DO* represents the disease overlapping score and *p-val* indicates the significance of the linkage.

Presenilins (PSEN and PSEN2) are shared between Alzheimer's disease, seborrheic keratosis, glomerulonephritis, erythema and cardiomyopathy. Seborrheic keratosis is a skin disorder consist of benign overgrowths of epithelium that have a pigmented velvety. The prevalence of the disease increases with aging. On the other hand, glomerulonephritis is a renal disorder characterized by inflammation of the glomeruli, or small blood vessels in the kidneys. Erythema is a general condition of the redness of skin; generally it occurs with environmental stimuli directed to inflammation. Cardiomyopathy is an idiopathic cardiovascular disease affecting the myocardium primarily. Although the diseases seem quite different and affect various parts of the system, these diseases intersect with each other through presenilins, which are the members of the γ-secretase protease complex. It is

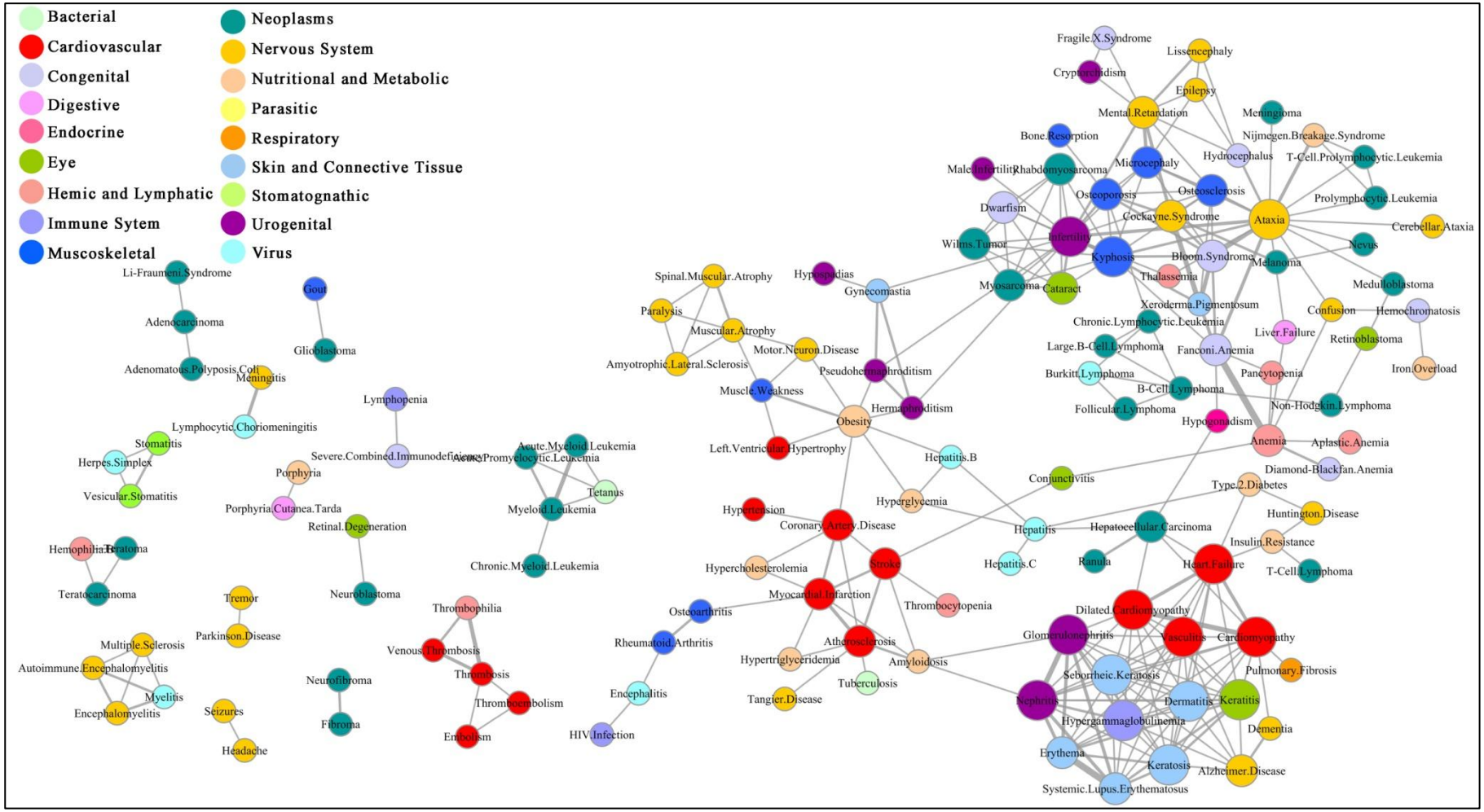

Figure 6.9. Disease associations derived through ADFN

Table 6.8. Selected pairs of disease terms derived from ADFN displaying significant associations in terms of shared genes

DiseaseTerm1	**DiseaseTerm2**	***DO***	***p-val***	**Shared genes**
Alzheimer Disease (Nervous System)	Dementia (Nervous System)	0.231	1.85E-04	APOE, PSEN1, PSEN2
Alzheimer Disease (Nervous System)	Seborrheic Keratosis (Skin and Connective Tissue)	0.182	2.51E-05	PSEN1, PSEN2
Alzheimer Disease (Nervous System)	Vasculitis (Cardiovascular)	0.182	3.71E-05	PSEN1, PSEN2
Keratosis (Skin and Connective Tissue)	Alzheimer Disease (Nervous System)	0.167	1.94E-03	PSEN1, PSEN2
Glomerulonephritis (Urogenital)	Seborrheic Keratosis (Skin and Connective Tissue)	0.333	7.40E-10	PSEN1, PSEN2
Cardiomyopathy (Cardiovascular)	Dilated Cardiomyopathy (Cardiovascular)	0.857	0.00E+00	CYP1A1, HDAC1, HFE, IL6ST, PSEN1, PSEN2
Seborrheic Keratosis (Skin and Connective Tissue)	Dilated Cardiomyopathy (Cardiovascular)	0.333	3.03E-09	PSEN1, PSEN2
Nephritis (Urogenital)	Dilated Cardiomyopathy (Cardiovascular)	0.200	1.54E-03	PSEN1, PSEN2
Cardiomyopathy (Cardiovascular)	Dementia (Nervous System)	0.200	2.20E-03	CXCL12, PSEN1
Glomerulonephritis (Urogenital)	Systemic Lupus Erythematosus (Skin and Connective Tissue)	0.400	7.90E-10	GADD45A,LYN, PSEN1, PSEN2
Erythema (Skin and Connective Tissue)	Systemic Lupus Erythematosus (Skin and Connective Tissue)	1.000	0.00E+00	CD4, GADD45A, LYN, PSEN1, PSEN2, PTGS2, TLR7, XPC
Thalassemia (Hemic and Lymphatic)	Ataxia (Nervous System)	0.118	1.62E-03	APOB, ERCC2
Ataxia (Nervous System)	Osteosclerosis (Musculoskeletal)	0.118	2.32E-03	ERCC2, XPA

well known that presenilin-dependent cleavage of amyloid-β (Aβ) precursor protein (APP) contributes to the generation of Aβ peptides and the formation of Aβ plaques. However, presenilin partial deficient mice developed skin lesions similar to seborrheic keratosis and glomerulonephritis, keratitis and vasculitis, as seen in human systemic lupus erythematosus [364]. This diverse impact of presenilins can be attributed to the interaction with a variety of proteins ranging from the modulation of cell viability to cell adhesion, protein trafficking, calcium homeostasis and regulation of gene transcription [90, 365, 366].

The links between Alzheimer's disease and cardiomyopathy were established through dementia, where along with PSEN1 and CXCL12 are shared. CXCL12 is a shared protein that is not previously associated with AD. Stromal cell-derived factor 1 (CXCL12) is a chemokine that controls many bone-marrow derived cell functions, has been involved in recruitment of brain cells to the lesions in the central nervous system and involved in neuroinflammatory responses. Due to its modulating effect, CXCL12 has been investigated for a potential role in AD, results indicated that the protein is associated with the disease and suggested as a potential therapeutic target [367]. In animal models, CXCL12 has a modulating effect on cardiomyocyte calcium homeostasis regulation [368] and elevated levels of CXCL12 were observed after myocardial infarction in humans [369]. Hence, the regulatory effect of CXCL12 may suggest a link between inflammation and calcium homeostasis with AD.

7. COMPARATIVE ANALYSIS AND DISCUSSION

A general perspective to the interference of complex disorders was developed using the modular architectures of the functional linkage networks representing three complex diseases: cardiovascular disease, Type 2 diabetes and Alzheimer's disease. These three disorders have been the subject of intense research due to their prevalence. Although the contribution of environmental factors, such as nutrition and sedentary life, to the diseases has been investigated, there is a considerable effort to understand the contribution of genetic factors to the development and progression of the diseases. The discovery of genetic factors has been anticipated to understand the fundamental underlying mechanisms, hence to improve the pathogenesis, diagnosis and treatment. Recent years the systems biology approaches, rather than focusing on single gene or a single system, enabled the researchers to develop generalized perspectives to achieve a comprehensive understanding and identification of biomarkers. In fact, the results showed that the genetic inheritance also contributes to the associations among complex disorders, where an interference of the systems was observed yielding shared mechanisms.

7.1. Comparison of Network Topology

To elucidate these mechanisms that are shared and to uncover other complex disease associations, the functional linkage networks for CVD, T2D and AD were constructed using the proteins encoded by the genes reported to be associated, which are called as core proteins, and the first neighbors were also incorporated to capture putative proteins. The functional linkages extracted from STRING database, where a confidence score was assigned based on the variety of information that supports the linkage. To achieve a comprehensive representation of the disease, a threshold for confidence score of functional linkage was set while keeping a substantial amount of core proteins in the network. The selection of the confidence score was based on two measures: coverage and constitution. Table 7.1 summarizes the networks as well as basic topological properties of the networks.

The networks were entitled as cardiovascular disease related functional linkage network (CFN), Type 2 diabetes related functional linkage network (TDFN) and

Alzheimer's disease related functional linkage network (ADFN). Network topological analyses showed that the degree distribution of the networks follow Power law with degree exponents, 2.28, 2.54 and 2.52, respectively, indicating the scale-freeness of the networks, hence biological relevance.

Table 7.1. Summary of the networks

	CFN	**TDFN**	**ADFN**
c-proteins	234	494	244
N	1536	2734	1587
l	3345	14823	7785
confidence score	900	940	960
% coverage	88.03	95.95	92.62
% constitution	13.41	17.34	14.24
$<k>$	4.36	10.84	9.81
γ	2.28	2.54	2.52

11 proteins are present in these disease related linkage networks: apolipoproteins APOA1, APOA2, and APOB, coagulation factors F3 and F5, thrombomodulin (THBD), glutathiones GSTM1 and GSTT1, interlukins IL2 and IL2RA, oncostatin M (OSM). The presence of apolipoproteins represents the existence of lipid metabolism, the coagulation factors indicate the roles of blood colagulation, glutathiones, interleukins and oncostatin M reflect the involvement of inflammation in the pathogenesis of the diseases.

The hub proteins, which are the determinants of the major biological processes involved in the disease, were identified through accumulative degree and betweenness distributions. It is a well-known phenomenon that hub proteins have higher degrees in the network, but rather than identifying hub proteins solely on degree, betweenness, which emphasizes the importance of a node to transmit the information throughout the network, was also considered as a significant determinant. In these three networks, insulin (INS), interleukin-6 (IL6) and signal transducer and activator of transcription 3 (STAT3), act as hub proteins, supporting the importance of signaling pathways in these networks were identified.

To build a modular perspective, the functional modules, which are highly condensed subgraphs, were derived in these networks. The functional modules were selected on the modularity measure, where the maximum number of possible interactions was attained among the proteins included in the module. The functional modules derived in the networks were restricted to size four and above to maintain biological relevance. 566, 5355 and 2674 functional modules were identified for CFN, TDFN and ADFN, respectively, and considered for further analyses. Functional analysis of proteins are members of a specific module indicated that they are all functionally linked. As expected, hub proteins in the selected networks present in the modular structures, due to their high interactions.

7.2. A Systems Biology Approach to Link CVD, T2D and AD

There is mounting evidence that proteins function in same pathway are often co-localized and co-expressed. It is also a well-known fact that genes contributing to a disease have an increased tendency for their products to interact and function together. Based on this premise, the evaluation of the functional modules was performed by calculating the consistency in the module in terms of sharing pathways, co-localization, co-expression and shared diseases. Incorporation of various independent data was anticipated to generate a deliberate and consistent evaluation method. The information deposited in KEGG pathway, mammalian localization database, a gene-disease association classification, which is developed in this study by incorporating Medical Subject Headings with the OMIM database, and Gene Expression Omnibus were used as resources. For each of the networks, the top scoring functional modules were assembled and analyzed in terms of Gene Ontology terms, representing the fundamental processes in the corresponding disease.

Blood coagulation cascade, vascular homeostasis and lipoprotein metabolism is present in CFN, as expected, and these processes underlie the delicate balance of blood homeostasis and lipid homeostasis which are maintained through inflammatory mechanisms JAK-STAT signaling and NF-κB cascades. The connection of blood coagulation and fibrinolysis to the proinflammatory cytokines was established through vitronectin (VTN) and oncostatin M (OSM), which are the proteins that are not previously associated with CVD. The modulating roles of these proteins in maintaining blood homeostasis and inflammation require further research. The presence of proinflammatory

cytokines and its connection with vascular homesostasis, involving the members of renin-angiotensin system and blood coagulation suggests there is the possible link between inflammation, hypertension and atherosclerosis. Accumulating evidence also suggests that inflammation participates in hypertension [370].

The presence of blood coagulation cascade is also observed in TDFN and ADFN, implying the importance of blood homeostasis in the pathogenesis of the diseases. The activation of blood coagulation system is suggested as an early indication of vascular damage [371]. Endothelial dysfunction leading to vascular damage is a common pattern shared by CVD and T2D [372]. In brain, the impaired blood coagulation results in the formation of fibrinolysis and blood clots, which is associated with impaired hemostatic imbalance observed in AD patients [352]. Hence, any disturbance in the blood coagulation cascade may result in endothelial vascular damage and uncontrolled fibrinolysis leading to formation of blood clots is a common underlying mechanism that is shared by CVD, T2D and AD.

The contribution of lipid metabolism, impaired cholesterol regulation to CVD and T2D has been the subject of intense research for years. Cholesterol uptake is known to increase the risk of obesity and Type 2 diabetes. The regulation of lipid metabolism involves the transport of lipids, particularly cholesterol and triglycerides, in the blood. The low density lipoprotein (LPL) and apolipoproteins (APOB, APOE, APOA1) are involved in lipid metabolism are present in CFN, TDFN and ADFN. The accumulation and subsequent oxidation of LPL is an early indication of atherosclerosis. The initiation of the process involves the accumulation of LDLs, which are modified by oxidation or enzymatic activity, and form agglomerates, leading to an increase phagocytosis by macrophages. In plasma and brain, LDL and chylomicrons, which are circulating triglycerides entrapped by APOB, APOE and APOC1, are removed from circulation by liver. APOE is also responsible for the formation of Aβ plaques in AD etiology. The role of APOE in maintaining cholesterol homeostasis in the brain may contribute to the increased risk for AD. Hence, it is necessary to gain insights for brain cholesterol mechanism to clarify the contribution of cholesterol to the disease and potential use of drugs targeting cholesterol. Consequently, the comprehensive understanding of the delicate balance for cholesterol is

needed to decipher its role as a shared biological process between CVD, T2D and AD etiology.

Another cluster of proteins, that is not associated with a biological process, observed in all of the disease networks include GSTM1, GSTM4, GSTT1, EPHX1, CYP1A1, CYP1A2, GPX1 proteins with varying compositions. GSTM1, GSTM4 and GSTT1 are members of Glutathione-S-transferase family of proteins, which play a major role in cellular antioxidant defense mechanisms by catalyzing the reduction of potentially harmful peroxides. Glutathione peroxidase (GPX1), which functions as an antioxidant enzyme, is responsible for the detoxification of hydrogen peroxide. Glutathiones are known as the detoxification enzymes responsible for the metabolism of a broad range of xenobiotics and carcinogens [373]. These proteins signify the importance of oxidative stress in the pathobiological processes in CVD, T2D and AD. Oxidative stress refers to a persistent imbalance between excessive production of ROS (reactive oxygen species) and limited antioxidant defenses [76]. The accumulation of the products of the oxidative stress, especially reactive oxygen species (ROS), which are basically free radicals, can cause the damage to biological macromolecules: proteins, lipids and DNA. Because of their unpaired electrons, free radicals are very unstable and highly reactive; hence bind to molecules with free electrons yielding unstable molecules. The free radicals produced in the body are toxic, and if not removed or neutralized, they react with lipids, proteins, and nucleic acids and damage cellular functions. In case of an increased oxidative damage, the free radical form of O_2, superoxide, can be converted in the hydrogen peroxide, H_2O_2, which is highly toxic. However, the H_2O_2 can be neutralized by GPX1, therefore any impairment in the GPX1 may result in the increase in the H_2O_2, leading the apoptosis [84]. The effect of oxidative stress in Type 2 diabetes results from the glucose toxicity mediated by ROS, which has been shown by indirect studies that the high glucose levels produce an oxidative stress inducing a subsequent increase in antioxidant enzyme levels and the harmful effects of high glucose concentration can be reversed by antioxidants [374], as well as the impaired insulin gene expression and β-cell function due to high glucose levels [375]. Initial studies investigating the impact of glutathione in AD produced controversial results [376-378]. These results may possibly be attributed to the fact that the alterations in the glutathione system are secondary to other events leading to neurodegeneration. Hence the prominent role of oxidative stress revealed in this study indicates that oxidative stress can

be considered as one of the major links in the pathobiological processes in CVD, T2D and AD.

NF-κB pathway, which is one of the major signaling pathways that is activated in response to proinflammatory cytokines and Toll-like receptors (TLRs), is shared between CVD and AD networks. The control of the induction of proinflammatory cytokines and chemokines is carried out through the Toll–like receptor pathway. Seven Toll-like receptors (TLRs) and large number of cytokines including TNF, IRAK2, IRAK4, TRAF6 and MYD88 are present in the modular structure of CFN. The presence of NF-κB cluster in ADFN map suggests the involvement of inflammatory events in neurodegeneration in AD. Recent studies showed that Toll-like receptors (TLRs), having central roles in innate immune system, have also been linked to the pathogenesis of AD [114, 115]. The relation of NF-κB signaling is observed through TRAF6 interaction with PSEN and NGFR linkage with TLR7. Experimental studies investigating inflammation as a marker of these two diseases revealed the role of LPL and Aβ peptide plaque formation mediating inflammation in CVD and AD, respectively. In fact, the accumulation of macrophages in the brain due to impaired blood brain barrier [113] is similar to those found in intima [41], as well as, both processes share the same cytokines. The initial inflammatory responses, such as removal of LDL and Aβ-peptide, can be protective against the diseases. However, the initiation stimulus of inflammation is yet to be discovered.

The presence of prostaglandins in T2D and AD networks suggest the possible role of inflammation in these diseases. PTGES, PTGDS, PTGIS, PTGS1 are members of prostaglandins, which are generally considered to be potent pro-inflammatory mediators. PTGD2 and prostacyclin (PTGIS) mediate a number of other effects including inhibition of platelet aggregation, smooth muscle relaxation and contraction, vasodilation and vasoconstriction. Prostaglandin E2 synthases (PGES and PTGES2) are a group enzymes that are engaged in paracrine signaling system. Recent studies reported PTGES1 involvement in inflammatory pain and ischemic brain injury, and suggested as potential pharmacological targets for stress-related disorders [309, 310]. The modulating role of prostaglandins in AD can be expected due to their effects in platelet aggregation, vasodilation and vasoconstriction; however to confirm the direct relation between

prostaglandins and Type 2 diabetes, experimental studies should be substantiated regarding the supplementary role of prostaglandins in muscle and adipose glucose transport.

The presence of inflammatory cytokines and prostaglandins in these disease networks highlights the importance of inflammation in the pathophysiology of these diseases. Considering the fact that, inflammatory response generally is not tissue specific, inflammation can be regarded as the major association between these diseases. This result indicating the obvious contribution of inflammation to CVD, T2D and AD is in concordance with a recent report investigating the accelerative effect of T2D in AD [379].

Another biological process that is shared between T2D and AD is the ubiquitin-proteasome system (UPS), which is an essential process responsible for the degradation of the majority of intracellular proteins and plays a pivotal role in the regulation of many cellular processes, particularly in the activation of nuclear factor B (NF-κB) and the transcription factors, which induces the production of proinflammatory cytokines. In AD, the accumulation of aggregates and inclusions of aberrant proteins imply an impaired UPS. The finding that UPS is up-regulated in T2D in animal models supported the finding that the defective UPS leading to inflammation might provide a link between AD and T2D that has not been revealed yet.

7.3. Complex Disease Interventions

Within the framework of this study, the complex disorders are linked to each other through shared proteins present in the disease networks, CVD, T2D and AD. Since the connection between the diseases is established through the proteins involved in the individual functional linkage networks, the relations gathered herein are more likely to be local. However, the connections between some disorders are “conserved” in each network, as well as some clusters formed in the disease networks. For instance, the strong link between myocardial infarction and atherosclerosis is conserved in all three disease networks, while shared proteins are changed. The disease associations conserved in CVD, T2D and AD networks are listed in Table 7.2. As expected, majority of the linkages contain the generalized form and specific forms of the diseases.

The links between muscular atrophy, spinal muscular atrophy (SMA) and amyotrophic lateral sclerosis (ALS) and obesity observed in disease associations derived from TDFN were also detected in the disease linkages based on ADFN, suggesting the consistent links between these disorders. ALS is a neurodegenerative and SMA is a muscular disease that both are characterized by marked weight loss. The proteins ciliary neurotrophic factor (CNTF) and cardiotrophin 1 (CTF1) are common in both of these disorders, suggesting a potential implication in obesity.

Table 7.2. Disease pairs conserved in three disease networks

		CVD			T2D			AD		
		DO_{CVD}	i_{CVD}*	*p-val*	DO_{T2D}	i_{CVD}	*p-val*	DO_{AD}	i_{CVD}	*p-val*
Hyperglycemia	Obesity	0.316	6	4.28E-07	0.417	5	5.12E-13	0.154	2	3.29E-03
Myocardial Infarction	Atherosclerosis	0.353	12	7.90E-06	0.300	3	4.73E-07	0.250	4	2.07E-04
Myocardial Infarction	Stroke	0.304	7	2.00E-05	0.333	2	8.58E-09	0.300	3	3.94E-06
Thrombocytopenia	Stroke	0.235	4	3.41E-04	0.667	2	0.00E+00	0.333	2	7.33E-08
Coronary Artery Disease	Stroke	0.286	6	6.95E-05	0.500	2	2.69E-14	0.333	2	2.13E-08
Multiple Sclerosis	Encephalomyelitis	0.222	2	3.35E-04	0.429	3	4.01E-13	0.500	2	4.28E-13
Thrombophilia	Thrombosis	0.353	6	4.48E-10	1.000	3	0.00E+00	0.500	5	2.78E-15
Autoimmune Encephalomyelitis	Encephalomyelitis	0.833	5	0.00E+00	0.667	4	0.00E+00	1.000	3	0.00E+00
Myelitis	Encephalomyelitis	0.857	6	0.00E+00	0.857	6	0.00E+00	1.000	3	0.00E+00
Systemic Lupus Erythematosus	Erythema	0.857	6	0.00E+00	0.750	3	0.00E+00	1.000	8	0.00E+00
Lymphocytic Choriomeningitis	Meningitis	0.750	3	0.00E+00	1.000	3	0.00E+00	1.000	4	0.00E+00
Anemia	Aplastic Anemia	0.231	3	2.54E-06	0.300	3	1.85E-10	0.091	2	8.73E-03
Coronary Artery Disease	Myocardial Infarction	0.348	8	1.27E-06	0.600	3	0.00E+00	0.375	3	6.22E-11
Ataxia	Cerebellar Ataxia	0.400	2	8.86E-14	0.667	2	0.00E+00	0.118	2	2.29E-03
Paralysis	Amyotrophic Lateral Sclerosis	0.500	2	0.00E+00	0.667	2	0.00E+00	0.500	2	1.56E-12
Hypertriglyceridemia	Atherosclerosis	0.207	6	2.18E-05	0.333	3	1.24E-09	0.167	2	1.41E-04
Multiple Sclerosis	Autoimmune Encephalomyelitis	0.250	2	4.37E-05	0.333	2	1.99E-08	0.500	2	5.22E-13
Myelitis	Autoimmune Encephalomyelitis	0.714	5	0.00E+00	0.571	4	0.00E+00	1.000	3	0.00E+00
Coronary Artery Disease	Atherosclerosis	0.303	10	2.77E-05	0.375	3	9.20E-14	0.154	2	2.89E-03
Cardiomyopathy	Dilated Cardiomyopathy	0.400	2	1.53E-12	1.000	2	0.00E+00	0.857	6	0.00E+00

* *i* indicates the number of proteins that are shared between the disease pairs

8. CONCLUSION AND FUTURE PROSPECTS

8.1. Conclusion

In this study, a generalized perspective to complex disease interference was developed by combining various levels of biological information for three prevalent complex diseases: cardiovascular disease, Type 2 diabetes and Alzheimer's disease. The underlying shared biological processes in these diseases were investigated using a systems based modular approach.

One aspect of systems biology involves the integration of diverse experimental and biological databases using computational tools to infer new knowledge about the systems. Network based approaches incorporating these resources have been widely used to predict protein function / interactions and selection of putative target proteins within a particular pathway. In the last decades, systems biology approaches received considerable attention to understand the contribution of environmental stimuli as well as genetic factors to the development of complex disorders with the hope that discovering these genetic factors will provide fundamental insights for pathogenesis, diagnosis and treatment.

The study was initiated with the construction of the cardiovascular disease functional linkage network (CFN), with 1536 nodes and 3345 interactions using proteins encoded by 234 genes associated with cardiovascular disease. Following the derivation of functional modules in CFN, integration of CFN with bibliomics revealed that out of 566, 227 functional modules are significantly associated with one or more diseases. Functional modules involved in blood coagulation, lipid metabolism and renin-angiotensin systems were found to be linked to cardiovascular diseases (CVD) as expected. Analysis of functional modules revealed possible regulatory roles of HIF1A, SP1 and CXCL12 in the pathogenesis of CVD and modulation of their activities may be considered as potential therapeutic tools.

A more comprehensive perspective to CVD was accomplished through incorporation of diverse levels of omics and semantics information with the CFN. Following the

subsequent scoring of the functional modules in terms of shared pathways, co-localization, co-expression and associations with diseases, the members of the top scoring functional modules were assembled in a condensed network indicating that, the biological processes involved in the pathogenesis of the disease are not restricted to blood coagulation, lipid metabolism and vascular homeostasis, but the inflammatory mechanisms also play pivotal role in these processes. The proteins, VTN, OSM, EPHX1 and FURIN, although not reported to have pronounced roles in cardiovascular diseases, were found to play important functions in this network. The proteins assembled in this condensed network were also associated with other complex disorders, albeit the interference of the complex diseases can be inferred through shared partners. Along with the well documented associations of obesity, significant linkages between CVD with proteinuria and rheumatoid arthritis were deciphered through shared proteins among these diseases.

The approach was later applied to two other prevalent complex disorders: Type 2 diabetes and Alzheimer's disease. The functional linkage networks include 2734 proteins / 14823 linkages (TDFN) and 1587 proteins / 7785 linkages (ADFN) for Type 2 diabetes and Alzheimer's disease, respectively. The subsequent scoring and evaluation of the functional modules enumerated from TDFN revealed not only the well documented fundamental biological processes, but also featured processes including ubiquitin proteasome system, sonic hedhehog signalling and prostaglandin mechanism. In TDFN, the proteins CNTF, IFNGR1, NRIP1, and CT1 were proposed to play regulatory roles in the pathogenesis of the disease. The linkages between the complex diseases derived from TDFN suggested the conjunction of the disease with various types of muscoskeletal and neurological diseases, including spinal muscular atrophy, Duchenne Muscular Dystrophy and amyotrophic lateral sclerosis.

The functional modules enumerated from ADFN were also evaluated by using the same scoring system. Along with the prominent biological processes, the involvement of ubiquitin-proteasome system, apoptosis and inflammatory mechanisms in the condensed map of ADFN suggested that the well-documented biological events might be insufficient to explain the progression of the disease. In addition, the links established between the presenilins, which are regarded as the mediators of Aβ-plaque formation in brain, and TRAF6 may suggest that presenilins not only involved in plaque formation but also act as

mediators of the inflammatory responses. CXCL12 was proposed to play a modulating role in initiating the inflammatory response and linking AD with CVD. The complex disorders associated with the proteins included in the condensed map of ADFN supported the possible links between the disease with several skin, renal and cardiovascular diseases.

The individual fundamental processes and signalling cascades observed in each of these diseases were converged at blood coagulation, glucose homeostasis, lipid metabolism, oxidative stress and inflammation. These biological processes were suggested as the underlying links between CVD, T2D and AD.

Blood coagulation and impaired homeostatic balance resulting in endothelial vascular damage and uncontrolled fibrinolysis was a common shared mechanism between these diseases. An impaired insulin action and imbalance in cholesterol homeostasis were appeared to affect the pathogenesis of all of the diseases. The presence of proteins related with oxidative stress in the disease networks indicated that oxidative stress is one of the underlying links connecting these diseases. One of the significant findings of this study was that inflammation is the missing link between CVD, T2D and AD. The proteins generating inflammatory response, including various proinflammatory cytokines and chemokines were present and play central roles in all disease networks. Considering the fact that inflammatory responses were not restricted to tissues, inflammation can be regarded as the major association between these diseases.

8.2. Future Prospects

This study focused on gaining new information on underlying shared biological pathways between complex disorders through systems based modular approach. The results provide further understanding for the biological pathways that contribute to the diseases investigated and point out putative proteins for their roles in the pathobiological processes. The underlying biological conditions of CVD, T2D and AD were observed to converge at blood coagulation, oxidative stress, glucose homeostasis, lipid metabolism and inflammation.

The disease specific networks constructed in this study are function based, albeit the proteins are linked to each other through functional associations. Another valuable resource of linkage is the direct protein interactions; indeed these disease-related linkage networks can be constructed via protein-protein interactions from high throughput experiments. A comparative analysis between these networks may shed light to novel protein function annotations.

These results in this study were elucidated by incorporating data from various resources that provides a deliberate and consistent evaluation method. However, the results obtained herein reflect a general understanding to the complicated events underneath complex diseases. Thereby, to ameliorate these findings and to get a full perspective, the differential gene expression analysis may provide more insight to understand the contribution of suggested biological processes to the diseases.

The linkages between the complex diseases derived from this study suggested the conjunction of Type 2 diabetes with various muscoskeletal and neurological diseases, including spinal muscular atrophy, Duchenne Muscular Dystrophy and amyotrophic lateral sclerosis, as well as the possible links between Alzheimer’s disease and several skin, renal and cardiovascular diseases. Thereby, the application of this approach to these diseases could be beneficial to reveal unspecified biological pathways and novel proteins.

REFERENCES

1. Goh, K., M. Cusick, D. Valle, B. Childs, M. Vidal and A.-L. Barabási, "The Human Disease Network", *Proceedings in National Academy of Science*, Vol. 104, pp. 8685-8690, 2008.

2. Libby, P., "Inflammation and Cardiovascular Disease Mechanisms", *American Journal of Clinical Nutrition*, Vol. 83, pp. 456S-460S, 2006.

3. Mayer, B., J. Erdmann and H. Schunkert, "Genetics and Heritability of Coronary Artery Disease and Myocardial Infarction", *Clinical Research in Cardiology*, Vol. 96, pp. 1-7, 2007.

4. Butte, A. and I. Kohane, "Creation and Implications of a Phenome-Genome Network", *Computational Biology*, Vol. 24, pp. 55-62, 2006.

5. Hwang, S., S.-W. Son, S. Kim, Y. Kim, H. Jeong and D. Lee, "A Protein Interaction Network Associated with Asthma", *Journal of Theoretical Biology*, Vol. 252, pp. 722-731, 2008.

6. Xue, H., B. Xian, D. Dong, K. Xia, S. Zhu, Z. Zhang, L. Hou, Q. Zhang, Y. Zhang and J.-D. Han, "A Modular Network Model of Aging", *Molecular Systems Biology*, Vol. 3, pp. 147, 2007.

7. Wang, J., S. Zhang, Y. Wang, L. Chen and X. Zhang, "Disease-aging Network Reveals Significant Roles of Aging Genes in Connecting Genetic Diseases", *PLoS Computational Biology*, Vol. 5, pp. e1000521, 2009.

8. Lim, J., T. Hao, C. Shaw, A. Patel, G. Szabó, J. Rual, C. Fisk, N. Li, A. Smolyar, D. Hill, A.-L. Barabási, M. Vidal and H. Zoghbi, "A Protein-protein Interaction Network for Human Inherited Ataxias and Disorders of Purkinje Cell Degeneration", *Cell*, Vol. 125, pp. 801-814, 2006.

9. Taylor, I., R. Linding, D. Warde-Farley, Y. Liu, C. Pesquita, D. Faria, S. Bull, T. Pawson, Q. Morris and J. Wrana, "Dynamic Modularity in Protein Interaction Networks Predicts Breast Cancer Outcome", *Nature Biotechnology*, Vol. 27, pp. 199-204, 2009.

10. Baranzini, S., N. Galwey, J. Wang, P. Khankhanian, R. Lindberg, D. Pelletier, W. Wu, B. Uitdehaag, L. Kappos, C. Polman, P. Matthews, S. Hauser, R. Gibson, J. Oksenberg and M. Barnes, "Pathway and Network-based Analysis of Genome-wide Association Studies in Multiple Sclerosis", *Human Molecular Genetics*, Vol. 18, pp. 2078-2090, 2009.

11. Girvan, M. and M. Newman, "Community Structure in Social and Biological Networks", *Proceedings in National Academy of Science*, Vol. 99, pp. 7821-7826, 2002.

12. Onnela, J., J. Saramäki, J. Hyvönen, G. Szabó, M. de Menezes, K. Kaski, A.-L. Barabási and J. Kertèsz, "Analysis of a Large-Scale Weighted Network of One-to-one Human Communication", *New Journal of Physics*, Vol. 9, pp. 1-27, 2007.

13. Barabási, A.-L., R. Albert and H. Jeong, "Scale-free Characteristics of Random Networks: The Topology of the World Wide Web", *Physica A: Statistical Mechanics and its Applications*, Vol. 281, pp. 69-77, 2000.

14. Camacho, J., R. Guimerà and L. Amaral, "Analytical Solution of a Model for Complex Food Webs", *Physical Review Lettters E*, Vol. 65, pp. 030901, 2002.

15. Chen, J. and B. Yuan, "Detecting Functional Modules in the Yeast Protein-protein Interaction Network", *BioInformatics*, Vol. 22, pp. 2283-2290, 2006.

16. Karimpour-Fard, A., S. Leach, L. Hunter and R. Gill, "The Topology of the Bacterial Co-conserved Protein Network and Its Implications for Predicting Protein Function", *BMC Genomics*, Vol. 9, pp. 313, 2008.

17. Tsankov, A., C. Brown, M. Yu, M. Win, P. Silver and J. Casolari, "Communication Between Levels of Transcriptional Control Improves Robustness and Adaptivity", *Molecular Systems Biology*, Vol. 2, pp. 65, 2006.

18. Ulitsky, I. and R. Shamir, "Identification of Functional Modules Using Network Topology and High-throughput Data", *BMC Systems Biology*, Vol. 1, pp. 8-24, 2007.

19. Zhang, C., S. Liu and Y. Zhou, "Fast and Accurate Method for Identifying High-quality Protein-interaction Modules by Clique Merging and its Application to Yeast", *Journal of Proteome Research*, Vol. 5, pp. 801-807, 2006.

20. Wang, X., E. Dalkic, M. Wu and C. Chan, "Gene Module Level Analysis: Identification to Networks and Dynamics", *Current Opinion in Biotechology*, Vol. 19, pp. 482-491, 2008.

21. Zhang, S., X. Ning and X. Zhang, "Identification of Functional Modules in a PPI Network by Clique Percolation Clustering", *Computational Biology and Chemistry*, Vol. 30, pp. 445-451, 2006.

22. Bron, C. and J. Kerbosch, "Algorithm 457: Finding All Cliques in an Undirected Graph", *Communications of the ACM*, Vol. 16, pp. 575-577, 1973.

23. Stix, V., "Finding All Maximal Cliques in Dynamic Graphs", *Computational Optimization and Applications*, Vol. 27, pp. 173-186, 2004.

24. Lozano, M. and F. Escolano, "Protein Classification by Matching and Clustering Surface Graphs", *Pattern Recognition*, Vol. 39, pp. 539-551, 2006.

25. Baum, D., *Finding All Maximal Cliques of A Family of Induced Subgraphs*, Zuse Institute Berlin, Berlin, 2003

26. Gaspin, C., "RNA Secondary Structure Determination and Representation Based on Constraints Satisfaction", *Constraints*, Vol. 6, pp. 201-221, 2001.

27. Oliviera, S. and S.-C. Seok, "Multilevel Approaches for Large-Scale Proteomic Networks", *International Journal of Computer Mathematics*, Vol. 84, pp. 683-695, 2007.

28. Ikonomidis, I., J. Lekakis, G. Vamvakou, F. Andreotti and P. Nihoyannopoulos, "Cigarette Smoking is Associated with Increased Circulating Proinflammatory and Procoagulant Markers in Patients with Chronic Coronary Artery Disease: Effects of Aspirin Treatment", *American Heart Journal*, Vol. 149, pp. 832-839, 2005.

29. Goldhammer, E., A. Tanchilevitch, I. Maor, Y. Beniamini, U. Rosenschein and M. Sagiv, "Exercise Training Modulates Cytokines Activity in Coronary Heart Disease Patients", *International Journal of Cardiology*, Vol. 100, pp. 93-99, 2005.

30. Fontana, L., D. Villareal, E. Weiss, S. Racette, K. Steger-May, S. Klein and J. Holloszy, "Calorie Restriction or Exercise: Effects on Coronary Heart Disease Risk Factors. A Randomized, Controlled Trial", *American Journal of Physiology, Endocrinology and Metabolism*, Vol. 293, pp. E197-E202, 2007.

31. Holmes, J., I. Arispe and E. Moy, "Heart Disease and Prevention: Race and Age Differences in Heart Disease Prevention, Treatment and Mortality", *Medical Care*, Vol. 43, pp. I33-I41, 2005.

32. Pilote, L., K. Dasgupta, V. Guru, K. Humphries, J. McGrath, C. Norris, D. Rabi, J. Tremblay, A. Alamian, T. Barnett, J. Cox, W. Ghali, S. Grace, P. Hamet, T. Ho, S. Kirkland, M. Lambert, D. Libersan, J. O'Loughlin, G. Paradis, M. Petrovich and V. Tagalakis, "A Comprehensive View of Sex-specific Issues Related to Cardiovascular Disease", *Canadian Medical Association Journal*, Vol. 176, pp. S1-S44, 2007.

33. Buyken, A., A. von Eckardstein, H. Schulte, P. Cullen and G. Assmann, "Type 2 Diabetes Mellitus and Risk of Coronary Heart Disease: Results of the 10-year

Follow-up of the PROCAM Study", *European Journal of Cardiovascular Prevention & Rehabilitation*, Vol. 14, pp. 230-236, 2007.

34. van der Net, J., A. Isaacs, G. Dallinga-Thie, J. Kastelein, J. Defesche, E. Steyerberg and E. Sijbrands, "Haplotype of the Angiotensinogen Gene is Associated with Coronary Heart Disease in Familial Hypercholesterolemia", *Journal of Hypertension*, Vol. 26, pp. 462-467, 2008.

35. Zhang, L., K. Pua, Z. Chaib and J. Dixon, "Family History of Coronary Heart Disease is Associated with a Higher Incidence of Nonalcoholic Fatty Liver Disease: Central Obesity the Likely Link", *Obesity Research and Clinical Practice*, Vol. 1, pp. 125-131, 2007.

36. Canoy, D., S. Boekholdt, N. Wareham, R. Luben, A. Welch, S. Bingham, I. Buchan, N. Day and K.-T. Khaw, "Body Fat Distribution and Risk of Coronary Heart Disease in Men and Women in the European Prospective Investigation Into Cancer and Nutrition in Norfolk Cohort", *Circulation*, Vol. 116, pp. 2933-2943, 2007.

37. Ray, M., J. Ruan and W. Zhang, "Variations in the Transcriptome of Alzheimer's Disease Reveal Molecular Networks Involved in Cardiovascular Disease", *Genome Biology*, Vol. 9, pp. R148, 2008.

38. Bergholdt, R., Z. Størling, K. Lage, E. Karlberg, P. Olason, M. Aalund, J. Nerup, S. Brunak, C. Workman and F. Pociot, "Integrative Analysis for Finding Genes and Networks Involved in Diabetes and Other Complex Diseases", *Genome Biology*, Vol. 8, pp. R253, 2007.

39. Garcia-Lafuente, A., E. Guillamon, A. Villares, M. Rostagno and J. Martinez, "Flavonoids as Anti-inflammatory Agents: Implications in Cancer and Cardiovascular Disease", *Inflammation Research*, Vol. 58, pp. 537-552, 2009.

40. Azuaje, F., Y. Devauex and D. Wagner, "Computational biology for cardiovascular biomarker discovery", *Briefings in Bioinformatics*, Vol. 10, pp. 367–377, 2009.

41. Rader, D. and A. Daugherty, "Translating Molecular Discoveries into New Therapies for Atherosclerosis", *Nature*, Vol. 451, pp. 904-913, 2008.

42. Sherer, Y. and Y. Shoenfeld, "Mechanisms of Disease: Atherosclerosis in Autoimmune Diseases", *Nature Clinical Practice*, Vol. 2, pp. 99-106, 2006.

43. Frick, M., O. Elo, K. Haapa, O. Heinonen, P. Heinsalmi, P. Helo, J. Huttunen, P. Kaitaniemi, P. Koskinen and V. Manninen, "Helsinki Heart Study: primary-prevention trial with gemfibrozil in middle-aged men with dyslipidemia. Safety of treatment changes in risk factors, and incidence of coronary heart disease", *New England Journal of Medicine*, Vol. 317, pp. 1237–1245, 1987.

44. Oliver, W. J., J. Shenk, M. Snaith, C. Russell, K. Plunket, N. Bodkin, M. Lewis, D. Winegar, M. Sznaidman, M. Lambert, H. Xu, D. Sternbach, S. Kliewer, B. Hansen and T. Willson, "A selective peroxisome proliferator activated receptor delta agonist promotes reverse cholesterol transport", *Proceedings of the National Academy of Science*, Vol. 98, pp. 5306–5311, 2001.

45. Brown, M., A. Inazu, C. Hesler, L. Agellon, C. Mann, M. Whitlock, Y. Marcel, R. Milne, J. Koizumi and H. Mabuchi, "Molecular basis of lipid transfer protein deficiency in a family with increased high-density lipoproteins", *Nature*, Vol. 342, pp. 448–451, 1989.

46. Inazu, A., X. Jiang, T. Haraki, K. Yagi, N. Kamon, J. Koizumi, H. Mabuchi, R. Takeda, K. Takata and Y. Moriyama, "Increased high-density lipoprotein levels caused by a common cholesteryl-ester transfer protein gene mutation", *New England Journal of Medicine*, Vol. 323, pp. 1234–1238, 1990.

47. McCoy, M., G. Sun, D. Marchadier, C. Maugeais, J. Glick and D. Rader, "Characterization of the lipolytic activity of endothelial lipase", *Journal of Lipid Research*, Vol. 43, pp. 921–929, 2002.

48. Dahlbäck, B., "Advances in Understanding Pathogenic Mechanisms of Thrombophilic Disorders", *Blood*, Vol. 112, pp. 19-27, 2008.

49. Dahlbäck, B., "Blood Coagulation", *Lancet*, Vol. 355, pp. 1627-1632, 2000.

50. Lust, M., M. Vulcano and S. Danese, "The Protein C Pathway in Inflammatory Bowel Disease: the Missing Link between Inflammation and Coagulation ", *Trends in Molecular Medicine*, Vol. 14, pp. 237-244, 2008.

51. Gribouval, O., M. Gonzales, T. Neuhaus, J. Aziza, E. Bieth, N. Laurent, J. Bouton, F. Feuillet, S. Makni, H. Amar, G. Laube, A. Delezoide, R. Bouvier, F. Dijoud, F. Ollagnon-Roman, J. Roume, M. Joubert, C. Antignac and M. Gubler, "Mutations in Genes in the Renin-angiotensin System are Associated with Autosomal Recessive Renal Tubular Dysgenesis", *Nature Genetics*, Vol. 37, pp. 964-968, 2005.

52. Lonn, E., S. Yusuf, P. Jha, T. Montague, K. Teo, C. Benedict and B. Pitt, "Emerging role of angiotensin-converting enzyme inhibitors in cardiac and vascular protection", *Circulation*, Vol. 90, pp. 2056–2069, 1994.

53. Weber, M., "The angiotensin II receptor blockers: opportunities across the spectrum of cardiovascular disease", *Reviews in Cardiovascular Medicine*, Vol. 3, pp. 183–191, 2002.

54. Ellis, E., "Reactive carbonyls and oxidative stress: Potential for therapeutic intervention", *Pharmacology and Therapeutics*, Vol. 115, pp. 13-24, 2007.

55. Girn, H., N. Orsi and S. Homer-Vanniasinkam, "An Overview of Cytokine Interactions in Atherosclerosis and Implications for Peripheral Arterial Disease", *Vascular Medicine*, Vol. 12, pp. 299-309, 2007.

56. Mehra, V., V. Ramgolam and J. Bender, "Cytokines and Cardiovascular Disease", *Journal of Leukocyte Biology*, Vol. 78, pp. 805-818, 2005.

57. Lee, Y. and Å. Gustafsson, "Role of apoptosis in cardiovascular disease", *Apoptosis*, Vol. 14, pp. 536-548, 2009.

58. Niu, J., A. Azfer, M. Deucher, P. Goldschmidt-Clermont and P. Kolattukudy, "Targeted cardiac expression of soluble FAS prevents the development of heart failure in mice with cardiac specific expression of MCP-1", *Journal of Molecular and Cellular Cardiology*, Vol. 40, pp. 810–820, 2006.

59. Kockx, M., G. De Meyer, J. Muhring, W. Jacob, H. Bult and A. Herman, "Apoptosis and related proteins in different stages of human atherosclerotic plaques", *Circulation*, Vol. 97, pp. 2307–2315, 1998.

60. Olivetti, G., F. Quaini, R. Sala, C. Lagrasta, D. Corradi, E. Bonacina, S. Gambert, E. Cigola and P. Anversa, "Acute myocardial infarction in humans is associated with activation of programmed myocyte cell death in the surviving portion of the heart", *Journal of Molecular and Cellular Cardiology*, Vol. 28, pp. 2005–2016, 1996.

61. Wild, S., G. Roglic, A. Green, R. Sicree and H. King, "Global Prevalence of Diabetes", *Diabetes Care*, Vol. 27, pp. 1047-1053, 2004.

62. Hussain, A., B. Claussen, A. Ramachandran and R. Williams, "Prevention of Type 2 Diabetes:A Review", *Diabetes Research and Clinical Practice*, Vol. 76, pp. 317–326, 2007.

63. Barroso, I., J. Luan, R. Middelberg, A. Anne-Helen Harding, P. Franks, R. Jakes, D. Clayton, A. Schafer, S. O'Rahilly and N. Wareham, "Candidate Gene Association Study in Type 2 Diabetes Indicates a Role for Genes Involved in b-Cell Function as Well as Insulin Action", *PLoS Biology*, Vol. 1, pp. 41–55, 2003.

64. Saltiel, A. and C. Kahn, "Insulin Signalling and the Regulation of Glucose and Lipid Metabolism", *Nature*, Vol. 414, pp. 799-806, 2001.

65. Howard, J. and J. Flier, "Attenuation of Leptin and Insulin Signaling by SOCS Proteins", *Trends in Endocrinology and Metabolism*, Vol. 17, pp. 365-371, 2006.

66. Mooney, R., J. Senn, S. Cameron, N. Inamdar, L. Boivin, Y. Shang and R. Furlanetto, "Suppressors of cytokine signaling-1 and -6 associate with and inhibit the insulin receptor", *Journal of Biological Chemistry*, Vol. 276, pp. 25889–25893, 2001.

67. Tartaglia, L., "The leptin receptor", *Journal of Biological Chemistry*, Vol. 272, pp. 6093–6096, 1997.

68. Bjorbaek, C., K. El-Haschimi, J. Frantz and J. Flier, "The role of SOCS-3 in leptin signaling and leptin resistance", *Journal of Biological Chemistry*, Vol. 274, pp. 30059–30065, 1999.

69. Trujillo, M., S. Sullivan, I. Harten, S. Schneider, A. Greenberg and S. Fried, "Interleukin-6 Regulates Human Adipose Tissue Lipid Metabolism and Leptin Production in Vitro", *Journal of Clinical Endocrinology and Metabolism*, Vol. 89, pp. 5577-5582, 2004.

70. Rossetti, L. and I. Goldberg, "A New Piece in the Diabetes Puzzle", *Nature Medicine*, Vol. 8, pp. 112-114, 2002.

71. Koishi, R., Y. Ando, M. Ono, M. Shimamura, H. Yasumo, T. Fujiwara, H. Horikoshi and H. Furukawa, "Role of Angptl3 gene in the regulation of lipid metabolism", *Nature Genetics*, Vol. 30, pp. 151–157, 2002.

72. Burke, J., K. Williams, S. Gaskill, H. Hazuda, S. Haffner and M. Stern, "Rapid rise in the incidence of type 2 diabetes from 1987 to 1996: results from the San Antonio Heart Study", *Archives of Internal Medicine*, Vol. 159, pp. 1450–1456, 1999.

73. Ludvik, B., J. Nolan, J. Baloga, D. Sacks and J. Olefsky, "Effect of obesity on insulin resistance in normal subjects and patients with NIDDM", *Diabetes*, Vol. 44, pp. 1121–1125, 1995.

74. Butler, A., J. Janson, S. Bonner-Weir, R. Ritzel, R. Rizza and P. Butler, "b-Cell Deficit and Increased b-Cell Apoptosis in Humans With Type 2 Diabetes", *Diabetes*, Vol. 52, pp. 102-110, 2003.

75. Ceriello, A. and E. Motz, "Is Oxidative Stress the Pathogenic Mechanism Underlying Insulin Resistance, Diabetes, and Cardiovascular Disease? The Common Soil Hypothesis Revisited", *Arteriosclerosis Thrombosis and Vascular Biology*, Vol. 24, pp. 816-823, 2004.

76. Li, N., F. Frigerio and P. Maechler, "The sensitivity of pancreatic ?-cells to mitochondrial injuries triggered by lipotoxicity and oxidative stress", *Biochemical Society Transactions*, Vol. 36, pp. 930-934, 2008.

77. Blasiak, J., M. Arabski, R. Krupa, K. Wozniak, M. Zadrozny, J. Kasznicki, M. Zurawskac and J. Drzewoski, "DNA Damage and Repair in Type 2 Diabetes Mellitus", *Mutation Research*, Vol. 554, pp. 297-304, 2004.

78. Merkel, P., N. Khoury, C. Bertolotto and R. Perfetti, "Insulin and glucose regulate the expression of the DNA repair enzyme XPD", *Molecular and Cellular Endocrinology*, Vol. 201, pp. 75-85, 2003.

79. Behl, C., "Alzheimer's Disease and Oxidative Stress: Implications for Novel Therapeutic Approches", *Progress in Neurobiology*, Vol. 57, pp. 301-323, 1999.

80. Salloway, S., J. Mintzer, M. Weiner and J. Cummings, "Disease-modifying therapies in Alzheimer's disease", *Alzheimer's and Dementia*, Vol. 4, pp. 65-79, 2008.

81. Thorsett, E. and L. Latimer, "Therapeutic approaches to Alzheimer's disease", *Current Opinion in Chemical Biology*, Vol. 4, pp. 377–382, 2000.

82. McGeer, P. and E. McGeer, "The possible role of complement activation in Alzheimer disease", *Trends in Molecular Medicine*, Vol. 8, pp. 519-523, 2002.

83. Adalbert, R., J. Gilley and M. Coleman, "Aß, tau and ApoE4 in Alzheimer's disease: the axonal connection", *Trends in Molecular Medicine*, Vol. 13, pp. 135-142, 2007.

84. Chauhan, V. and A. Chauhan, "Oxidative stress in Alzheimer's disease", *Pathophysiology*, Vol. 13, pp. 195–208, 2006.

85. Gleichmann, M. and M. Mattson, "Alzheimer's Disease and Neuronal Network Activity", *Neuromolecular Medicine*, Vol. 12, pp. 44–47, 2010.

86. Sherrington, R., E. I. Rogaev, Y. Liang, E. A. Rogaeva, G. Levesque, M. Ikeda, H. Chi, C. Lin, G. Li, K. Holman, T. Tsuda, L. Mar, J. F. Foncin, A. C. Bruni, M. P. Montesi, S. Sorbi, I. Rainero, L. Pinessi, L. Nee, I. Chumakov, D. Pollen, A. Brookes, P. Sanseau, R. J. Polinsky, W. Wasco, H. A. R. Da Silva, J. L. Haines, M. A. Pericak-Vance, R. E. Tanzi, A. D. Roses, P. E. Fraser, J. M. Rommens and P. H. St George-Hyslop, "Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's disease", *Nature*, Vol. 375, pp. 754-760, 1995.

87. Kimberly, W., W. Xia, T. Rahmati, M. Wolfe and D. Selkoe, "The transmembrane aspartases in presenilin 1 and 2 are obligatory for gamma-secretase activity and amyloid-beta protein generation", *Journal of Biological Chemistry*, Vol. 275, pp. 3173–3178, 2000.

88. Wolfe, M., W. Xia, B. Ostaszewski, T. Diehl, W. Kimberly and D. Selkoe, "Two transmembrane aspartates in presenilin-1 required for presenilin endoproteolysis and gamma-secretase activity", *Nature*, Vol. 398, pp. 513–517, 1999.

89. Pastorino, L. and K. Lu, "Pathogenic mechanisms in Alzheimer's disease", *European Journal of Pharmacology*, Vol. 545, pp. 29-38, 2006.

90. McCarthy, J., "Involvement of presenilins in cell-survival signalling pathways", *Biochemical Society Transactions*, Vol. 33, pp. 568–572, 2005.

91. Parihar, M. and T. Hemnai, "Alzheimer's disease pathogenesis and therapeutic", *Interventions Journal of Clinical Neuroscience*, Vol. 11, pp. 456–467, 2004.

92. Rosenberg, K., J. Ross, H. Feinstein, S. Feinstein and J. Israelachvili, "Complementary dimerization of microtubule-associated tau protein: Implications for microtubule bundling and tau-mediated pathogenesis", *Proceedings of the National Academy of Sciences*, Vol. 105, pp. 7445-7450, 2008.

93. Wang, J. and F. Liu, "Microtubule-associated protein tau in development, degeneration and protection of neurons", *Progress in Neurobiology*, Vol. 85, pp. 148-175, 2008.

94. Iqbal, K., A. del C. Alonso, S. Chen, M. Chohan, E. El-Akkad, C. Gong, S. Khatoon, B. Li, F. Liu, A. Rahman, H. Tanimukai and I. Grundke-Iqbal, "Tau pathology in Alzheimer disease and other tauopathies", *Biochimica et Biophysica Acta*, Vol. 1739, pp. 198-210, 2005.

95. Rapoport, M., H. Dawson, L. Binder, M. Vitek and A. Ferreira, "Tau is essential to β-amyloid-induced neurotoxicity", *Proceedings of the National Academy of Sciences*, Vol. 99, pp. 6364-6369, 2002.

96. SantaCruz, K., J. Lewis, T. Spires, J. Paulson, L. Kotilinek, M. Ingelsson, A. Guimaraes, M. DeTure, M. Ramsden, E. McGowan, C. Forster, M. Yue, J. Orne, C. Janus, A. Mariash, M. Kuskowski, B. Hyman, M. Hutton and K. Ashe, "Tau Suppression in a Neurodegenerative Mouse Model Improves Memory Function", *Science*, Vol. 309, pp. 476-481, 2005.

97. Sotiropoulos, I., C. Catania, T. Riedemann, J. Fry, K. Breen, T. Michaelidis and O. Almeida, "Glucocorticoids trigger Alzheimer disease-like pathobiochemistry in rat

neuronal cells expressing human tau", *Journal of Neurochemistry*, Vol. 107, pp. 385-397, 2008.

98. Hanger, D., B. Anderton and W. Noble, "Tau phosphorylation: the therapeutic challenge for neurodegenerative disease", *Trends in Molecular Medicine*, Vol. 15, pp. 112-119, 2009.

99. Bartus, R. and D. Emerich, "Cholinergic markers in Alzheimer's disease", *Journal of American Medicine Association*, Vol. 282, pp. 2208–2209, 1999.

100. Auld, D., T. Kornecook, S. Bastianetto and R. Quirion, "Alzheimer's disease and the basal fore brain cholinergic system: relations to b-amyloid peptides, cognition and treatment strategies", *Progress in Neurobiology*, Vol. 68, pp. 209-245, 2002.

101. Hardy, J., "Amyloid, the presenillins and Alzheimer's disease", *Trends in Neuroscience*, Vol. 20, pp. 15-159, 1997.

102. Sommer, B., "Alzheimer's disease and the amyloid cascade hypothesis: ten years on", *Current Opinion in Pharmacology*, Vol. 2, pp. 87-92, 2002.

103. Pimplikar, S., "Reassessing the amyloid cascade hypothesis of Alzheimer's disease", *International Journal of Biochemistry and Cell Biology*, Vol. 41, pp. 1261-1268, 2009.

104. Park, L., J. Anrather, C. Forster, K. Kazama, G. Carlson and C. Iadecola, "Aβ-Induced Vascular Oxidative Stress and Attenuation of Functional Hyperemia in Mouse Somatosensory Cortex", *Journal of Cerebral Blood Flow and Metabolism*, Vol. 24, pp. 334-342, 2004.

105. Jhoo, J., H. Kim, T. Nabeshima, K. Yamada, E. Shin, W. Jhoo, W. Kim, K. Kang, S. Jo and J. Woo, "β-Amyloid (1-42)-induced learning and memory deficits in mice: involvement of oxidative burdens in the hippocampus and cerebral cortex", *Behavioural Brain Research*, Vol. 155, pp. 185-196, 2004.

106. Sultana, R., A. Ravagna, H. Mohmmad-Abdul, V. Calabrese and D. Butterfield, "Ferulic acid ethyl ester protects neurons against amyloid β-peptide(1-42)-induced oxidative stress and neurotoxicity: relationship to antioxidant activity", *Journal of Neurochemistry*, Vol. 92, pp. 749-758, 2005.

107. Dikalov, S., M. Vitek and R. Mason, "Cupric-amyloidβ peptide complex stimulates oxidation of ascorbate and generation of hydroxyl radical", *Free Radical Biology and Medicine*, Vol. 36, pp. 340-347, 2004.

108. Yanamadala, V. and R. Friedlander, "Complement in neuroprotection and neurodegeneration", *Trends in Molecular Medicine*, Vol. 6, pp. 69-76, 2010.

109. Rogers, J., N. R. Cooper, S. Webster, J. Schultz, P. L. McGeer, S. D. Styren, W. H. Civin, L. Brachova, B. Bradt and P. Ward, "Complement activation by beta-amyloid in Alzheimer disease", *Proceedings of the National Academy of Sciences*, Vol. 89, pp. 10016-10020, 1992.

110. Sastre, M., T. Klockgether and M. Heneka, "Contribution of inflammatory processes to Alzheimer's disease: molecular mechanisms", *International Journal of Developmental Neuroscience*, Vol. 24, pp. 167-176, 2006.

111. Alvarez, X., A. Franco, L. Fernández-Novoa and R. Cacabelos, "Blood levels of histamine, IL-1β, and TNF-α in patients with mild to moderate Alzheimer disease", *Molecular and Chemical Neuropathology*, Vol. 29, pp. 237-252, 1996.

112. Singh, V. and P. Guthikonda, "Circulating cytokines in Alzheimer's disease", *Journal of Psychiatric Research*, Vol. 31, pp. 657-660, 1997.

113. Cohen, R., "The Role of the Immune System in Alzheimer's Disease", *Focus*, Vol. 7, pp. 28-35, 2009.

114. Scholtzova, H., R. Kascsak, K. Bates, A. Boutajangout, D. Kerr, H. Meeker, P. Mehta, D. Spinner and T. Wisniewski, "Induction of toll-like receptor 9 signaling as a method for ameliorating Alzheimer's disease-related pathology", *Journal of Neuroscience*, Vol. 11, pp. 1846-1854, 2009.

115. Balistreri, C., M. Grimaldi, M. Chiappelli, F. Licastro, L. Castiglia, F. Listì, S. Vasto, D. Lio, C. Caruso and G. Candore, "Association between the polymorphisms of TLR4 and CD14 genes and Alzheimer's disease", *Current Pharmaceutical Design*, Vol. 14, pp. 2672-2677, 2008.

116. Albert, R., A.-L. Barabási, H. Jeong and G. Bianconi, "Power-law Distribution of the World Wide Web", *Science*, Vol. 287, pp. 2115a, 2000.

117. Barabási, A.-L., H. Jeong, R. Ravasz, Z. Néda, T. Vicsek and A. Schubert, "On the Topology of the Scientific Collaboration Networks", *Physica A*, Vol. 311, pp. 590-614, 2002.

118. Barabási, A.-L., H. Jeong, Z. Néda, R. Ravasz, A. Schubert and T. Vicsek, "Evolution of the Social Network of Scientific Collaborations", *Physica A: Statistical Mechanics and its Applications*, Vol. 311, pp. 590-614, 2002.

119. Jeong, H., B. Tombor, R. Albert, Z. Oltvai and A.-L. Barabási, "The Large-scale Organization of Metabolic Networks", *Nature*, Vol. 407, pp. 651-654, 2000.

120. Jeong, H., S. Mason, A.-L. Barabasi and Z. Oltvai, "Lethality and Centrality in Biological Networks", *Nature*, Vol. 411, pp. 41–42, 2001.

121. von Mering, C., R. Krause, B. Snel, M. Cornell, S. Oliver, S. Fields and P. Bork, "Comparative assessment of large-scale data sets of protein–protein interactions", *Nature*, Vol. 417, pp. 399-403, 2002.

122. Erdos, P. and A. Renyi, "On the Evolution of Random Graphs", *Publications in Mathematical Institute Hungarian Academy of Sciences*, Vol. 5, pp. 17, 1960.

123. Watts, D. and S. Strogatz, "Collective dynamics of ‘small-world’ networks", *Nature*, Vol. 393, pp. 440–442, 1998.

124. Rodriguez-Caso, C., M. Medina and R. Solé, "Topology, Tinkering and Evolution of The Human Transcription Factor Network", *FEBS Journal*, Vol. 272, pp. 6423-6434, 2005.

125. Joy, M., A. Brock, D. Ingber and S. Huang, "High Betweenness Proteins in the Yeast Protein Interaction Network", *Journal of Biomedicine and Biotechnology*, Vol. 2, pp. 96-103, 2005.

126. He, X. and J. Zhang, "Why Do Hubs Tend to be Essential in Protein Networks?", *PLoS Genetics*, Vol. 2, pp. E88, 2006.

127. Hsing, M., K. Byler and A. Cherkasov, "The use of Gene Ontology terms for predicting highly-connected 'hub' nodes in protein-protein interaction networks", *BMC Systems Biology*, Vol. 2, pp. 80, 2008.

128. Eisenberg, E. and E. Levanon, "Preferential attachment in the protein network evolution", *Physical Review Letters*, Vol. 91, pp. 138701, 2003.

129. Albert, R., H. Jeong and A.-L. Barabási, "Error and attack tolerance of complex networks", *Nature*, Vol. 406, pp. 378–382, 2000.

130. Spirin, V. and L. Mirny, "Protein Complexes and Functional Modules in Molecular Networks", *Proceedings in National Academy of Science*, Vol. 100, pp. 12123-12128, 2003.

131. Hartwell, L., J. Hopfield, S. Leibler and A. Murray, "From molecular to modular cell biology", *Nature*, Vol. 402, pp. C47–C52, 1999.

132. Hanisch, D., A. Zien, R. Zimmer and T. Lengauer, "Co-clustering of biological networks and gene expression data", *Bioinformatics*, Vol. 18, pp. S145–S154, 2002.

133. Rives, A. and T. Galitski, "Modular organization of cellular networks", *Proceeding of National Academy of Sciences*, Vol. 100, pp. 1128–1133, 2003.

134. Bader, G. and C. Hogue, "An automated method for finding molecular complexes in large protein interaction networks", *BMC Bioinformatics*, Vol. 4, pp. 2, 2003.

135. Petti, A. and G. Church, "A network of transcriptionally coordinated functional modules in *Saccharomyces cerevisiae*", *Genome Research*, Vol. 15, pp. 1298-1306, 2005.

136. Newman, M., "Detecting community structure in networks", *European Physical Journal B*, Vol. 38, pp. 321–330, 2004.

137. Adamcsek, B., G. Palla, I. Farkas, I. Imre Derenyi and T. Vicsek, "CFinder: locating cliques and overlapping modules in biological networks", *Bioinformatics*, Vol. 22, pp. 1021–1023, 2006.

138. Lubovac, Z., J. Gamalielsson and B. Olsson, "Combining Functional and Topological Properties to Identify Core Modules in Protein Interaction Networks", *Proteins: Structure, Function, and Bioinformatics*, Vol. 64, pp. 948–959, 2006.

139. Cho, Y., W. Hwang, M. Ramanathan and A. Zhang, "Semantic integration to identify overlapping functional modules in protein interaction networks", *BMC Bioinformatics*, Vol. 8, pp. 265, 2007.

140. Mete, M., F. Tang, X. Xu and N. Yuruk, "A structural approach for finding functional modules from large biological networks", *BMC Bioinformatics*, Vol. 9, pp. S19, 2008.

141. Maraziotis, I., K. Dimitrakopoulou and A. Bezerianos, "An in silico method for detecting overlapping functional modules from composite biological networks", *BMC Systems Biology*, Vol. 2, pp. 93, 2008.

142. Georgii, E., S. Dietmann, T. Uno, P. Pagel and K. Tsuda, "Enumeration of condition-dependent dense modules in protein interaction networks", *Bioinformatics*, Vol. 25, pp. 933–940, 2009.

143. Hood, L., J. Heath, M. Phelps and B. Lin, "Systems biology and new technologies enable predictive and preventative medicine", *Science*, Vol. 306, pp. 640–643, 2004.

144. Baudot, A., G. Gomez-Lopez and A. Valencia, "Translational disease interpretation with molecular networks", *Genome Biology*, Vol. 10, pp. 221, 2009.

145. Huang, Y., D. Hang, L. Lu, L. Tong, M. Gerstein and G. Montelione, "Targeting the Human Cancer Pathway Protein Interaction Network by Structural Genomics", *Molecular and Cellular Proteomics*, Vol. 7, pp. 2048–2060, 2008.

146. Nabieva, E., K. Jim, A. Agarwal, B. Chazelle and M. Singh, "Whole-proteome prediction of protein function via graph-theoretic analysis of interaction maps", *Bioinformatics*, Vol. 21, pp. 302–310, 2005.

147. Kar, G., A. Gursoy and O. Keskin, "Human cancer protein-protein interaction network: A structural perspective", *PloS Computational Biology*, Vol. 5, pp. e1000601, 2009.

148. Chen, Y., J. Zhu, P. Y. Lum, X. Yang, S. Pinto, D. J. MacNeil, C. Zhang, J. Lamb, S. Edwards, S. K. Sieberts, A. Leonardson, L. W. Castellini, S. Wang, M.-F. Champy, B. Zhang, V. Emilsson, S. Doss, A. Ghazalpour, S. Horvath, T. A. Drake, A. J. Lusis and E. E. Schadt, "Variations in DNA elucidate molecular networks that cause disease", *Nature*, Vol. 452, pp. 429-435, 2008.

149. Watkinson, J., X. Wang, T. Zheng and D. Anastassiou, "Identification of gene interactions associated with disease from gene expression data using synergy networks", *BMC Systems Biology*, Vol. 2, pp. 10, 2008.

150. Hwang, D., I. Lee, H. Yoo, N. Gehlenborg, J. Cho, B. Petritis, D. Baxter, R. Pitstick, R. Young, D. Spicer, N. Price, J. Hohmann, S. DeArmond, G. Carlson and L. Hood, "A systems approach to prion disease", *Molecular Systems Biology*, Vol. 5, pp. 252, 2009.

151. Dezsö, Z., Y. Nikolsky, T. Nikolskaya, J. Miller, D. Cherba, C. Webb and A. Bugrim, "Identifying disease-specific genes based on their topological significance in protein networks", *BMC Systems Biology*, Vol. 3, pp. 36, 2009.

152. Piruzian, E., S. Bruskin, A. Ishkin, R. Abdeev, S. Moshkovskii, S. Melnik, Y. Nikolsky and T. Nikolskaya, "Integrated network analysis of transcriptomic and proteomic data in psoriasis", *BMC Systems Biology*, Vol. 4, pp. 41, 2010.

153. Sam, L., Y. Liu, C. Friedman and Y. Lussier, "Discovery of Protein Interaction Networks Shared by Diseases", *Pacific Symposium on Biocomputing*, Vol. 12, pp. 76–87, 2007.

154. Lage, K., E. Karlberg, Z. Størling, P. Ólason, A. Pedersen, O. Rigina, A. Hinsby, Z. Tümer, F. Pociot, N. Tommerup, Y. Moreau and S. Brunak, "A Human Phenome-Interactome Network of Protein Complexes Implicated in Genetic Disorders", *Nature Biotechnology*, Vol. 25, pp. 309-316, 2007.

155. Lee, D., J. Park, K. Kay, N. Christakis, Z. Oltvai and A.-L. Barabási, "The implications of human metabolic network topology for disease comorbidity", *Proceedings of the National Academy of Sciences*, Vol. 105, pp. 9880–9885, 2008.

156. Park, J., D. Lee, N. Christakis and A.-L. Barabási, "The impact of cellular networks on disease comorbidity", *Molecular Systems Biology*, Vol. 5, pp. 262, 2009.

157. Li, Y. and P. Agarwal, "A pathway-based view of human diseases and disease relationships", *PLoS One*, Vol. 4, pp. e4346, 2009.

158. Jensen, L., M. Kuhn, M. Stark, S. Chaffron, C. Creevey, J. Muller, T. Doerks, P. Julien, A. Roth, M. Simonovic, P. Bork and C. von Mering, "STRING 8--a Global View on Proteins and their Functional Interactions in 630 Organisms", *Nucleic Acids Research*, Vol. 37, pp. D412-D416, 2009.

159. von Mering, C., L. Jensen, B. Snel, S. Hooper, M. Krupp, M. Foglierini, N. Jouffre, M. Huynen and P. Bork, "STRING: Known and Predicted Protein-protein Associations, Integrated and Transferred across Organisms", *Nucleic Acids Research*, Vol. 33, pp. D433-D437, 2005.

160. Carbon, S., A. Ireland, C. Mungall, S. Shu, B. Marshall and S. Lewis, "AmiGO: Online Access to Ontology and Annotation Data", *Bioinformatics*, Vol. 25, pp. 288-289, 2008.

161. Lowe, H. and G. Barnett, "Understanding and Using the Medical Subject Headings (MeSH) Vocabulary to Perform Literature Searches", *Journal of American Medical Association*, Vol. 271, pp. 1103-1108, 1994.

162. Shannon, P., A. Markiel, O. Ozier, N. Baliga, J. Wang, R. Ramage, N. Amin, B. Schwikowski and T. Ideker, "Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks", *Genome Research*, Vol. 13, pp. 2498-2504, 2003.

163. Kanehisa, M., S. Goto, M. Furumichi, M. Tanabe and M. Hirakawa, "KEGG for representation and analysis of molecular networks involving diseases and drugs", *Nucleic Acids Research*, Vol. 38, pp. D355-D360, 2010.

164. Sprenger, J., J. Fink, S. Karunaratne, K. Hanson, N. Hamilton and R. Teasdale, "LOCATE: a mammalian protein subcellular localization database", *Nucleic Acids Research*, Vol. 36, pp. D230-D233, 2008.

165. McKusick, V., *Mendelian Inheritance in Man: A Catalog of Human Genes and Genetic Disorders*, Johns Hopkins University and National Center for Biotechology Information, Baltimore, 1998

166. Edgar, R., M. Domrachev and A. Lash, "Gene Expression Omnibus: NCBI gene expression and hybridization array data repository", *Nucleic Acids Research*, Vol. 30, pp. 207-210, 2002.

167. Yu, H., P. Kim, E. Sprecher, V. Trifonov and M. Gerstein, "The Importance of Bottlenecks in Protein Networks: Correlation with Gene Essentiality and Expression Dynamics", *PLoS Computational Biology*, Vol. 3, pp. E59, 2007.

168. Semenza, G., "HIF-1 and Human Disease: One Highly Involved Factor", *Genes and Development*, Vol. 14, pp. 1983-1991, 2000.

169. Wilhide, M. and W. Jones, "Potential Therapeutic Gene for the Treatment of Ischemic Disease: Ad2/Hypoxia-Inducible Factor-1α (HIF-1)/VP16 Enhances B-Type Natriuretic Peptide Gene Expression via a HIF-1-Responsive Element ", *Molecular Pharmacology*, Vol. 69, pp. 1773-1778, 2006.

170. Deniaud, E., J. Baguet, R. Chalard, B. Blanquier, L. Brinza, J. Meunier, M. Michallet, A. Laugraud, C. Ah-Soon, A. Wierinckx, M. Castellazzi, J. Lachuer, C. Gautier, J. Marvel and Y. Leverrier, "Overexpression of Transcription Factor Sp1 Leads to Gene Expression Perturbations and Cell Cycle Inhibition", *PloS One*, Vol. 4, pp. e7035, 2009.

171. Szalad, A., M. Katakowski, X. Zheng, F. Jiang and M. Chopp, "Transcription factor Sp1 induces ADAM17 and contributes to tumor cell invasiveness under hypoxia", *Journal of Experimental & Clinical Cancer Research*, Vol. 28, pp. 129, 2009.

172. Schaëfer, D., B. Hamm-Kuënzelmann and K. Brand, "Glucose regulates the promoter activity of aldolase A and pyruvate kinase M2 via dephosphorylation of Sp1", *FEBS Letters*, Vol. 417, pp. 325-328, 1997.

173. Solomon, S., G. Majumdar, A. Martinez-Hernandez and R. Raghow, "A critical role of Sp1 transcription factor in regulating gene expression in response to insulin and other hormones", *Life Sciences*, Vol. 83, pp. 305-312, 2008.

174. Song, I., H. Chen, I. Aiba, A. Hossain, Z. Liang, L. Klomp and M. Kuo, "Transcription Factor Sp1 Plays an Important Role in the Regulation of Copper Homeostasis in Mammalian Cells", *Molecular Pharmacology*, Vol. 74, pp. 705-713, 2008.

175. Scotton, C., J. Wilson, K. Scott, G. Stamp, G. Wilbanks, S. Fricker, G. Bridger and F. Balkwill, "Multiple Actions of the Chemokine CXCL12 on Epithelial Tumor Cells in Human Ovarian Cancer", *Cancer Research*, Vol. 62, pp. 5930-5938, 2002.

176. Mantovani, A., P. Allavena, A. Sica and F. Balkwill, "Cancer-related Inflammation", *Nature*, Vol. 454, pp. 436-444, 2008.

177. Weber, C., "Platelets and Chemokines in Atherosclerosis Partners in Crime", *Circulation Research*, Vol. 96, pp. 612-616, 2005.

178. Slofstra, S., H. Cate and C. Spek, "Signal Transduction Induced by Activated Protein C: no role in Protection Against Sepsis? ", *Trends in Molecular Medicine*, Vol. 12, pp. 374-381, 2006.

179. Kallel, L., S. Matri, S. Karoui, M. Fekih, J. Boubaker and A. Filali, "Deep Venous Thrombosis Related to Protein S Deficiency Revealing Celiac Disease", *American Journal of Gastroenterology*, Vol. 104, pp. 256-257, 2009.

180. Musso, R., D. Cultrera, M. Stella, M. Musso, A. Cipolla and R. Giustollisi, "Recurrent Venous Thromboembolic Events During Treatment with Combined Oral

Contraceptives in Hereditaria Dystrophic Epidermolysis Bullosa", *Journal of Thrombosis and Haemostasis*, Vol. 3, pp. P0457, 2005.

181. Shook, B., M. Wells, R. Rapini and E. Cobos, "Epidermolysis Bullosa Acquisita Occurring in 2 Patients with Hepatitis C", *Journal of the American Academy of Dermatology*, Vol. 54, pp. 888-891, 2006.

182. Cugno, M., A. Zanichelli, F. Foieni, S. Caccia and M. Cicardi, "C1-inhibitor Deficiency and Angioedema: Molecular Mechanisms and Clinical Progress", *Trends in Molecular Medicine*, Vol. 15, pp. 69-78, 2009.

183. Pennings, M., I. Meurs, D. Ye, R. Out, M. Hoekstra, T. Van Berkel and M. Van Eck, "Regulation of Cholesterol Homeostasis in Macrophages and Consequences for Atherosclerotic Lesion Development", *FEBS Letters*, Vol. 580, pp. 5588-5596, 2006.

184. Menge, T., H. Hartung and O. Stuve, "Statins — a Cure-all for the Brain?", *Nature Reviews Neuroscience*, Vol. 6, pp. 325-331, 2005.

185. Linsel-Nitschke, P. and A. Tall, "HDL as a Target in the Treatment of Atherosclerotic Cardiovascular Disease", *Nature Reviews Drug Discovery*, Vol. 4, pp. 193-206, 2005.

186. Maeda, N., I. Shimomura, K. Kishida, H. Nishizawa, M. Matsuda, H. Nagaretani, N. Furuyama, H. Kondo, M. Takahashi, Y. Arita, R. Komuro, N. Ouchi, S. Kihara, Y. Tochino, K. Okutomi, M. Horie, S. Takeda, T. Aoyama, T. Funahashi and Y. Matsuzawa, "Diet-induced Insulin Resistance in Mice Lacking Adiponectin/ACRP30", *Nature Medicine*, Vol. 8, pp. 731-737, 2002.

187. Zou, C. and J. Shao, "Role of Adipocytokines in Obesity-associated Insulin Resistance", *Journal of Nutritional Biochemistry*, Vol. 19, pp. 277-286, 2008.

188. Rutherford, S., M. Johnson, R. Curtain and L. Griffiths, "Chromosome 17 and the Inducible Nitric Oxide Synthase gene in Human Essential Hypertension", *Human Genetics*, Vol. 109, pp. 408-415, 2001.

189. Pieper, G. and A. Roza, "The Complex Role of iNOS in Acutely Rejecting Cardiac Transplants", *Free Radical Biology and Medicine*, Vol. 44, pp. 1536-1552, 2008.

190. Cuzzocrea, S., E. Mazzon, L. Sautebin, I. Serraino, L. Dugo, G. Calabró, A. Caputi and A. Maggi, "The Protective Role of Endogenous Estrogens in Carrageenan-induced Lung Injury in the Rat", *Molecular Medicine*, Vol. 7, pp. 478-487, 2001.

191. Feng, Q., X. Lu and D. Jones, "Increased Inducible Nitric Oxide Synthase Expression Contributes to Myocardial Dysfunction and Higher Mortality after Myocardial Infarction in Mice", *Circulation*, Vol. 104, pp. 700-704, 2001.

192. Cybulsky, M., J. Fries, A. Williams, P. Sultan, R. Eddy, M. Byers, T. Shows, M. Gimbrone and T. Collins, "Gene Structure, Chromosomal Location, and Basis for Alternative mRNA Splicing of the Human VCAM1 gene", *Proceedings in National Academy of Science*, Vol. 88, pp. 7859-7863, 1991.

193. Campbell, D., M. Woodward, J. Chalmers, S. Colman, A. Jenkins, B. Kemp, B. Neal, A. Patel and S. MacMahon, "Soluble Vascular Cell Adhesion Molecule 1 and N-terminal pro-B-type Natriuretic Peptide in Predicting Ischemic Stroke in Patients with Cerebrovascular Disease", *Archives of Neurology*, Vol. 63, pp. 60-65, 2006.

194. Hohmann, H., R. Remy, M. Brockhaus and A. van Loon, "Two Different Cell Types have Different Major Receptors for Human Tumor Necrosis Factor (TNF alpha)", *Journal of Biological Chemistry*, Vol. 264, pp. 14927-14934, 1989.

195. Poirier, O., V. Nicaud, J. Gariepy, D. Courbon, A. Elbaz, C. Morrison, F. Kee, A. Evans, D. Arveiler, P. Ducimetiere, P. Amarenco and F. Cambien, "Polymorphism R92Q of the Tumour Necrosis Factor Receptor 1 Gene is Associated with Myocardial Infarction and Carotid Intima-media Thickness: the ECTIM, AXA, EVA

and GENIC Studies", *European Journal of Human Genetics*, Vol. 12, pp. 213-219, 2004.

196. Nathan, C., "Inducible Nitric Oxide Synthase: What Difference Does It Make?", *Journal of Clinical Investigations*, Vol. 100, pp. 2417-2423, 1997.

197. Griendling, K., T. Murphy and R. Alexander, "Molecular Biology of the Renin-angiotensin System", *Circulation*, Vol. 87, pp. 1816-1828, 1993.

198. Neutel, J., "Effect of the Renin-angiotensin System on the Vessel Wall: Using ACE inhibition to Improve Endothelial Function", *Journal of Hypertension*, Vol. 18, pp. 599-606, 2004.

199. Delaney, D., S. Kennedy, V. Tobias and R. Farnsworth, "Congenital Unilateral Renal Tubular Dysgenesis and Severe Neonatal Hypertension", *Pediatric Nephrology*, Vol. 24, pp. 863-867, 2009.

200. David, J., D. Mehic, L. Bakiri, A. Schilling, V. Mandic, M. Priemel, M. Idarraga, M. Reschke, O. Hoffmann, M. Amling and E. Wagner, "Essential Role of RSK2 in c-Fos-dependent Osteosarcoma Development", *Journal of Clinical Investigations*, Vol. 115, pp. 664-672, 2005.

201. Springer, M., A. Chen, P. Kraft, M. Bednarski and H. Blau, "VEGF Gene Delivery to Muscle: Potential Role for Vasculogenesis in Adults", *Molecular Cell*, Vol. 2, pp. 549-558, 1998.

202. Neptune, E., P. Frischmeyer, D. Arking, L. Myers, T. Bunton, B. Gayraud, F. Ramirez, L. Sakai and H. Dietz, "Dysregulation of TGF-beta Activation Contributes to Pathogenesis in Marfan Syndrome", *Nature Genetics*, Vol. 33, pp. 407-411, 2003.

203. Mucida, D., Y. Park, G. Kim, O. Turovskaya, I. Scott, M. Kronenberg and H. Cheroutre, "Reciprocal TH17 and Regulatory T Cell Differentiation Mediated by Retinoic Acid", *Science*, Vol. 317, pp. 256-260, 2007.

204. Sethi, J. and G. Hotamisligil, "The Role of TNFα in Adipocyte Metabolism", *Cell and Developmental Biology*, Vol. 10, pp. 19-29, 1999.

205. Bruce, A., W. Boling, M. Kindy, J. Peschon, P. Kraemer, M. Carpenter, F. Holtsberg and M. Mattson, "Altered Neuronal and Microglial Responses to Excitotoxic and Ischemic Brain Injury in Mice Lacking TNF Receptors", *Nature Medicine*, Vol. 2, pp. 788-794, 1996.

206. Pollock, A., J. Turck and D. Lovett, "The Prodomain of Interleukin 1 alpha Interacts with Elements of the RNA Processing Apparatus and Induces Apoptosis in Malignant Cells", *FASEB Journal*, Vol. 17, pp. 203-213, 2003.

207. Patti, M. and C. Kahn, "The Insulin Receptor - A Critical Link in Glucose Homeostasis and Insulin Action", *Journal of Basic and Clinical Physiology and Pharmacology*, Vol. 9, pp. 89-109, 1998.

208. Abel, E., O. Peroni, J. Kim, Y.-B. Kim, O. Boss, E. Hadro, T. Minnemann, G. Shulman and B. Kahn, "Adipose-selective Targeting of the GLUT4 Gene Impairs Insulin Action in Muscle and Liver", *Nature*, Vol. 409, pp. 729-733, 2001.

209. Maor, G. and E. Karnieli, "The Insulin-Sensitive Glucose Transporter (GLUT4) Is Involved in Early Bone Growth in Control and Diabetic Mice, But Is Regulated through the Insulin-Like Growth Factor I Receptor", *Endocrinology*, Vol. 140, pp. 1841-1851, 1999.

210. Faggioni, R., J. Jones-Carson, D. Reed, C. Dinarello, K. Feingold, C. Grunfeld and G. Fantuzzi, "Leptin-deficient (ob/ob) Mice are Protected from T Cell-mediated Hepatotoxicity: Role of Tumor Necrosis Factor Alpha and IL-18", *Proceedings in National Academy of Science*, Vol. 97, pp. 2367-2372, 2000.

211. Chen, K., F. Li, J. Li, H. Cai, S. Strom, A. Bisello, D. Kelley, M. Friedman-Einat, G. Skibinski, M. McCrory, A. Szalai and A. Zhao, "Induction of Leptin Resistance

through Direct Interaction of C-reactive Protein with Leptin", *Nature Medicine*, Vol. 12, pp. 425-432, 2006.

212. Straub, R., "The Complex Role of Estrogens in Inflammation", *Endocrine Reviews*, Vol. 25, pp. 521-574, 2007.

213. Talbert, M., C. Langefeld, J. Ziegler, J. Mychaleckyj, S. Haffner, J. Norris and D. Bowden, "Polymorphisms Near SOCS3 are Associated with Obesity and Glucose Homeostasis Traits in Hispanic Americans from the Insulin Resistance Atherosclerosis Family Study", *Human Genetics*, Vol. 125, pp. 153-162, 2009.

214. Barry, S., P. Townsend, D. Latchman and A. Stephanou, "Role of the JAKSTAT Pathway in Myocardial Injury", *Trends in Molecular Medicine*, Vol. 13, pp. 82-89, 2007.

215. Schalken, J., A. Roebroek, P. Oomen, S. Wagenaar, F. Debruyne, H. Bloemers and W. Van de Ven, "FUR gene expression as a discriminating marker in human lung cancer", *Journal of Clinical Investigation*, Vol. 80, pp. 1545–1549, 1987.

216. Laprise, M., F. Grondin, P. Cayer, P. McDonald and C. Dubois, "Furin gene (fur) regulation in differentiating human megakaryoblastic Dami cells: involvement of the proximal GATA recognition motif in the P1 promoter and impact on the maturation of furin substrates", *Blood*, Vol. 100, pp. 3578–3587, 2002.

217. Wasley, L., A. Rehemtulla, J. Bristol and R. Kaufman, "PACE/Furin Can Process the Vitamin K-dependent Pro-factor IX Precursor within the Secretory Pathway", *Journal of Biological Chemistry*, Vol. 268, pp. 8458–8465, 1993.

218. Hoffman, M., "Remodeling the Blood Coagulation Cascade", *Journal of Thrombosis and Thrombolysis*, Vol. 16, pp. 17–20, 2003.

219. Monroe, D. and N. Key, "The tissue factor-factor VIIa complex: procoagulant activity, regulation, and multitasking", *Journal of Thrombosis and Haemostasis*, Vol. 5, pp. 1097–1105, 2007.

220. Mann, K., K. Brummel-Ziedins, T. Orfeo and S. Butenas, "Models of blood coagulation", *Blood Cells, Molecules and Diseases*, Vol. 36, pp. 108–117, 2006.

221. Gailani, D. and T. Renne, "Intrinsic Pathway of Coagulation and Arterial Thrombosis", *Arteriosclerosis, Thrombosis and Vascular Biology*, Vol. 27, pp. 2507–2513, 2007.

222. Cesarman-Maus, G. and K. Hajjar, "Molecular mechanisms of fibrinolysis", *British Journal of Haematology*, Vol. 129, pp. 307–321, 2005.

223. Franchini, M. and G. Lippi, "Von Willebrand factor and thrombosis", *Annals of Hematology*, Vol. 85, pp. 415-423, 2006.

224. Wu, Y., J. Zuo, G. Ji, G. Saiyin, X. Liu, F. Yin, N. Cao, Y. Wen, J. Li and L. Yu, "Proapoptotic Function of Integrin β3 in Human Hepatocellular Carcinoma Cells", *Clinical Cancer Research*, Vol. 15, pp. 60–69, 2009.

225. Zhou, A., J. Huntington, N. Pannu, R. Carrell and R. Read, "How vitronectin binds PAI-1 to modulate fibrinolysis and cell migration", *Nature Structural Biology*, Vol. 10, pp. 541–544, 2003.

226. Bassi, D., J. Fu, R. de Cicco and A. Klein-Szanto, "Proprotein Convertases: "Master Switches" in the Regulation of Tumor Growth and Progression", *Molecular Carcinogenesis*, Vol. 44, pp. 151–161, 2005.

227. Berthou, L., N. Duverger, F. Emmanuel, S. Langouët, J. Auwerx, A. Guillouzo, J. Fruchart, E. Rubin, P. Denèfle, B. Staels and D. Branellec, "Opposite regulation of human versus mouse apolipoprotein A-I by fibrates in human apolipoprotein A-I transgenic mice", *Journal of Clinical Investigation*, Vol. 97, pp. 2408–2416, 1996.

228. Rader, D., "Molecular regulation of HDL metabolism and function: implications for novel therapies", *Journal of Clinical Investigations*, Vol. 116, pp. 3090-3100, 2006.

229. Tso, P. and M. Liu, "Apolipoprotein A-IV, food intake, and obesity", *Physiology and Behavior*, Vol. 83, pp. 631–643, 2004.

230. Stan, S., E. Delvin, M. Lambert, E. Seidman and E. Levy, "Apo A-IV: an update on regulation and physiologic functions", *Biochimica et Biophysica Acta*, Vol. 1631, pp. 177–187, 2003.

231. Peelman, F., W. Waelput, H. Iserentant, D. Lavens, S. Eyckerman, L. Zabeau and J. Tavernier, "Leptin: linking adipocyte metabolism with cardiovascular and autoimmune diseases", *Progress in Lipid Research*, Vol. 43, pp. 283-301, 2004.

232. Shen, L., P. Tso, S. Woods, R. Sakai, W. Davidson and M. Liu, "Hypothalamic Apolipoprotein A-IV Is Regulated by Leptin", *Endocrinology*, Vol. 148, pp. 2681–2689, 2007.

233. Paul, M., A. Mehr and R. Kreutz, "Physiology of Local Renin-Angiotensin Systems", *Physiological Reviews*, Vol. 86, pp. 747–803, 2006.

234. Schmieder, R., K. Hilgers, M. Schlaich and B. Schmidt, "Renin-angiotensin System and Cardiovascular Risk", *Lancet*, Vol. 369, pp. 1208-1219, 2007.

235. Tsutsui, M., H. Shimokawa, Y. Otsuji, Y. Ueta, Y. Sasaguri and N. Yanagihara, "Nitric oxide synthases and cardiovascular diseases: insights from genetically modified mice", *Circulation*, Vol. 73, pp. 986–993, 2009.

236. Kawashima, S., "Malfunction of Vascular Control in Lifestyle-Related Diseases: Endothelial Nitric Oxide (NO) Synthase/NO System in Atherosclerosis", *Journal of Pharmacological Sciences*, Vol. 96, pp. 411–419, 2004.

237. Huang, P., Z. Huang, H. Mashimo, K. Bloch, M. Moskowitz, J. Bevan and M. Fishman, "Hypertension in mice lacking the gene for endothelial nitric oxide synthase", *Nature*, Vol. 377, pp. 239–242, 1995.

238. Shesely, E., N. Maeda, H. Kim, K. Desai, J. Krege, V. Laubach, P. Sherman, W. Sessa and O. Smithies, "Elevated blood pressures in mice lacking endothelial nitric oxide synthase", *Proceedings of the National Academy of Sciences*, Vol. 93, pp. 13176–13181, 1996.

239. Knowles, J., R. Reddick, J. Jennette, E. Shesely, O. Smithies and N. Maeda, "Enhanced atherosclerosis and kidney dysfunction in eNOS(-/-)Apoe(-/-) mice are ameliorated by enalapril treatment", *Journal of Clinical Investigation*, Vol. 105, pp. 451–458, 2000.

240. Grote, K., M. Luchtefeld and B. Schieffer, "JANUS under stress--role of JAK/STAT signaling pathway in vascular diseases", *Vascular Pharmacology*, Vol. 43, pp. 357–363, 2005.

241. Mui, A., H. Wakao, N. Harada, A. O'Farrell and A. Miyajima, "Interleukin-3, granulocyte-macrophage colony-stimulating factor, and interleukin-5 transduce signals through two forms of STAT5", *Journal of Leukocyte Biology*, Vol. 57, pp. 799–803, 1995.

242. Callus, B. and B. Mathey-Prevot, "Interleukin-3-induced activation of the JAK/STAT pathway is prolonged by proteasome inhibitors", *Blood*, Vol. 91, pp. 3182–3192, 1996.

243. Kaplan, M., U. Schindler, S. Smiley and M. Grusby, "Stat6 is required for mediating responses to IL-4 and for development of Th2 cells", *Immunity*, Vol. 4, pp. 313–319, 1996.

244. Ogata, N., T. Kouro, A. Yamada, M. Koike, N. Hanai, T. Ishikawa and K. Takatsu, "JAK2 and JAK1 constitutively associate with an interleukin-5 (IL-5) receptor alpha

and betac subunit, respectively, and are activated upon IL-5 stimulation", *Blood*, Vol. 91, pp. 2264–2271, 1998.

245. Marrero, M., B. Schieffer, W. Paxton, L. Heerdt, B. Berk, P. Delafontaine and K. Bernstein, "Direct stimulation of Jak/STAT pathway by the angiotensin II AT1 receptor", *Nature*, Vol. 375, pp. 247– 250, 1995.

246. Tedgui, A. and Z. Mallat, "Cytokines in Atherosclerosis: Pathogenic and Regulatory Pathway", *Physiological Reviews*, Vol. 86, pp. 515-581, 2006.

247. Kawai, T. and S. Akira, "Signaling to NF-κB by Toll-like Receptors", *Trends in Molecular Medicine*, Vol. 13, pp. 460-469, 2007.

248. Hayek, T., J. Stephens, C. Hubbart, J. Acharya, M. Caslake, E. Hawe, G. Miller, S. Hurel and S. Humphries, "A common variant in the glutathione S transferase gene is associated with elevated markers of inflammation and lipid peroxidation in subjects with diabetes mellitus", *Atherosclerosis*, Vol. 184, pp. 404–412, 2006.

249. Manfredi, S., C. Federici, E. Picano, N. Botto, A. Rizza and M. Andreassi, "GSTM1, GSTT1 and CYP1A1 detoxification gene polymorphisms and susceptibility to smoking-related coronary artery disease: A case-only study", *Mutation Research*, Vol. 621, pp. 106–112, 2007.

250. Saadat, I. and M. Saadat, "Influence of genetic polymorphisms of glutathione S-transferase T1 (GSTT1) and M1 (GSTM1) on hematological parameters", *Molecular Biology Reports*, Vol. 37, pp. 249–253, 2010.

251. Tamer, L., B. Ercan, A. Camsari, H. Yildirim, C. D, N. Sucu, N. Aras Ates and U. Atik, "Glutathione S-transferase gene polymorphism as a susceptibility factor in smoking-related coronary artery disease", *Basic Research in Cardiology*, Vol. 99, pp. 223–229, 2004.

252. Doney, A., S. Lee, G. Leese, A. Morris and C. Palmer, "Increased cardiovascular morbidity and mortality in type 2 diabetes is associated with the glutathione S transferase theta-null genotype: a Go-DARTS Study", *Circulation*, Vol. 111, pp. 2927–2934, 2005.

253. Farin, F. and C. Omiecinski, "Regiospecific expression of cytochrome P-450s and microsomal epoxide hydrolase in human brain tissue", *Journal of Toxicology and Environmental Health*, Vol. 40, pp. 317–335, 1993.

254. Coller, J., P. Fritz, U. Zanger, I. Siegle, M. Eichelbaum, H. Kroemer and T. Murdter, "Distribution of microsomal epoxide hydrolase in humans: an immunohistochemical study in normal tissues, and benign and malignant tumours", *The Histochemical Journal*, Vol. 33, pp. 329–336, 2001.

255. Decker, M., M. Arand and A. Cronin, "Mammalian epoxide hydrolases in xenobiotic metabolism and signalling", *Archives in Toxicology*, Vol. 83, pp. 297–318, 2009.

256. Salam, M., P. Lin, E. Avol, W. Gauderman and F. Gilliland, "Microsomal Epoxide Hydrolase, Glutathione S-transferase P1, Traffic and Childhood Asthma", *Thorax*, Vol. 62, pp. 1050–1057, 2007.

257. Cornelis, M., A. El-Sohemy and H. Campos, "Microsomal Epoxide hydrolase genotype and risk of myocardial infarction", *Archives in Toxicology*, Vol. 81, pp. 641–645, 2007.

258. Marowsky, A., J. Burgener, J. Falck, J. Fritschy and M. Arand, "Distribution of soluble and microsomal epoxide hydrolase in the mouse brain and its contribution to cerebral epoxyeicosatrienoic acid metabolism", *Neuroscience*, Vol. 163, pp. 646–661, 2009.

259. Liu, M., A. Sun, E. Shin, X. Liu, S. Kim, C. Runyons, W. Markesbery, H. Kim and G. Bing, "Expression of microsomal epoxide hydrolase is elevated in Alzheimer’s

hippocampus and induced by exogenous beta-amyloid and trimethyl-tin", *European Journal of Neuroscience*, Vol. 23, pp. 2027–2034, 2006.

260. Ross, J., N. Stagliano, M. Donovan, R. Breitbart and G. Ginsburg, "Atherosclerosis: a cancer of the blood vessels?", *American Journal of Clinical Pathology*, Vol. 116, pp. S97–S107, 2001.

261. Zarling, J., M. Shoyab, H. Marquardt, M. Hanson, M. Lioubin and G. Todaro, "Oncostatin M: a growth regulator produced by differentiated histiocytic lymphoma cells", *Proceedings of the National Academy of Sciences*, Vol. 83, pp. 9739–9743, 1986.

262. Kishimoto, T., S. Akira, M. Narazaki and T. Taga, "Interleukin-6 Family of Cytokines and gp130", *Blood*, Vol. 86, pp. 1243-1254, 1995.

263. Schneiderman, J., M. Sawdey, M. Keeton, G. Bordin, E. Bernstein, R. Dilley and D. Loskutoff, "Increased type 1 plasminogen activator inhibitor gene expression in atherosclerotic human arteries", *Proceedings in National Academy of Science*, Vol. 89, pp. 6998-7002, 1992.

264. Rossignol, P., A. Luttun, J. Martin-Ventura, F. Lupu, P. Carmeliet, D. Collen, E. Angles-Cano and H. Lijnen, "Plasminogen activation: a mediator of vascular smooth muscle cell apoptosis in atherosclerotic plaques", *Journal of Thrombosis and Haemostasis*, Vol. 4, pp. 664–670, 2006.

265. Demyanets, S., C. Kaun, K. Rychli, G. Rega, S. Pfaffenberger, T. Afonyushkin, V. Bochkov, G. Maurer, K. Huber and J. Wojta, "The inflammatory cytokine oncostatin M induces PAI-1 in human vascular smooth muscle cells in vitro via PI 3-kinase and ERK1/2-dependent pathways", *American Journal of Physiology - Heart and Circulatory Physiology*, Vol. 293, pp. H1962–H1968, 2007.

266. Kastl, S., W. Speidl, K. Katsaros, C. Kaun, G. Rega, A. Assadian, G. Hagmueller, M. Hoeth, R. de Martin, Y. Ma, G. Maurer, K. Huber and J. Wojta, "Thrombin induces

the expression of oncostatin M via AP-1 activation in human macrophages: a link between coagulation and inflammation", *Blood*, Vol. 114, pp. 2812–2818, 2009.

267. Mercado, N., J. Brugts, J. Ix, M. Shlipak, S. Dixon, B. Gersh, P. Lemos, M. Guarneri, P. Teirstein, W. Wijns, P. Serruys, E. Boersma and W. O'Neill, "Usefulness of Proteinuria as a Prognostic Marker of Mortality and Cardiovascular Events Among Patients Undergoing Percutaneous Coronary Intervention (Data from the Evaluation of Oral Xemilofiban in Controlling Thrombotic Events [EXCITE] Trial)", *American Journal of Cardiology*, Vol. 102, pp. 1151–1155, 2008.

268. Simko, F., "Statins: a perspective for left ventricular hypertrophy treatment", *European Journal of Clinical Investigation*, Vol. 37, pp. 681-691, 2007.

269. Sandhu, S., N. Wiebe, L. Fried and M. Tonelli, "Statins for improving renal outcomes: a meta-analysis", *Journal of the American Society for Nephrology*, Vol. 17, pp. 2006–2016, 2006.

270. Levi, M., "Do statins have a beneficial effect on the kidney?", *Nature Clinical Practice Nephrology*, Vol. 2, pp. 666–667, 2006.

271. Zoja, C., D. Corna, D. Rottoli, D. Cattaneo, C. Zanchi, S. Tomasoni, M. Abbate and G. Remuzzi, "Effect of combining ACE inhibitor and statin in severe experimental nephropathy", *Kidney International*, Vol. 61, pp. 1635–1645, 2002.

272. Perkovic, V., C. Verdon, T. Ninomiya, F. Barzi, A. Cass, A. Patel, M. Jardine, M. Gallagher, F. Turnbull, J. Chalmers, J. Craig and R. Huxley, "The Relationship between Proteinuria and Coronary Risk: A Systematic Review and Meta-Analysis", *PLoS Medicine*, Vol. 21, pp. e207, 2008.

273. Knuefermann, P., J. Vallejo and D. Mann, "The role of innate immune responses in the heart in health and disease", *Trends in Cardiovascular Medicine*, Vol. 14, pp. 1–7, 2004.

274. Brady, S., B. Courten, C. Reid, F. Cicuttini, M. Courten and D. Liew, "The Role of Traditional Cardiovascular Risk Factors Among Patients with Rheumatoid Arthritis", *Journal of Rheumatology*, Vol. 36, pp. 34–40, 2009.

275. Hamdulay, S. and J. Mason, "Disease-modifying anti-rheumatic drugs: do they reduce cardiac complications of RA?", *Heart*, Vol. 95, pp. 1471–1472, 2009.

276. Galarraga, B., F. Khan, P. Kumar, T. Pullar and J. Belch, "C-reactive protein: the underlying cause of microvascular dysfunction in rheumatoid arthritis", *Rheumatology*, Vol. 47, pp. 1780–1784, 2008.

277. Brown, A., W. Olver, C. Donnelly, M. May, J. Naggert, D. Shaffer and D. Roopenian, "Searching QTL by gene expression: analysis of diabesity", *BMC Genetics*, Vol. 6, pp. 12, 2005.

278. Ridderstrale, M. and L. Groop, "Genetic dissection of type 2 diabetes", *Molecular and Cellular Endocrinology*, Vol. 297, pp. 10-17, 2009.

279. Gloyn, A. and M. McCarthy, "The genetics of type 2 diabetes", *Best Practice and Research Clinical Endocrinology and Metabolism*, Vol. 15, pp. 293-308, 2001.

280. Owen, K. and M. McCarthy, "Genetics of type 2 diabetes", *Current Opinion in Genetics and Development*, Vol. 17, pp. 239-244, 2007.

281. McCarthy, M. and P. Froguel, "Genetic approaches to the molecular understanding of type 2 diabetes", *American Journal of Physiology - Endocrinology and Metabolism*, Vol. 283, pp. E217-25, 2002.

282. Prokopenko, I., M. McCarthy and C. Lindgren, "Type 2 diabetes: new genes, new understanding", *Trends in Genetics*, Vol. 24, pp. 613-621, 2008.

283. Volkening, M. and I. Hoffmann, "Involvement of Human MCM8 in Prereplication Complex Assembly by Recruiting hcdc6 to Chromatin", *Molecular and Cellular Biology*, Vol. 25, pp. 1560-1568, 2005.

284. Pollok, S., C. Bauerschmidt, J. Sanger, H. Nasheuer and F. Grosse, "Human CDC45 is a Proliferation-associated Antigen", *FEBS Journal*, Vol. 274, pp. 3669-3684, 2007.

285. Parikh, H., E. Carlsson, W. Chutkow, L. Johansson, H. Storgaard, P. Poulsen, R. Saxena, C. Ladd, P. Schulze, M. Mazzini, C. Jensen, A. Krook, M. Björnholm, H. Tornqvist, J. Zierath, M. Ridderstråle, D. Altshuler, R. Lee, A. Vaag, L. Groop and V. Mootha, "TXNIP regulates peripheral glucose metabolism in humans", *PloS Medicine*, Vol. 4, pp. e158, 2007.

286. Pihlajamäki, J., T. Boes, E. Kim, F. Dearie, B. Kim, J. Schroeder, E. Mun, I. Nasser, P. Park, A. Bianco, A. Goldfine and M. Patti, "Thyroid hormone-related regulation of gene expression in human fatty liver", *Journal of Endocrinology and Metabolism*, Vol. 94, pp. 3521-3529, 2009.

287. Nair, S., Y. Lee, E. Rousseau, M. Cam, P. Tataranni, L. Baier, C. Bogardus and P. Permana, "Increased expression of inflammation-related genes in cultured preadipocytes/stromal vascular cells from obese compared with non-obese Pima Indians", *Diabetologia*, Vol. 48, pp. 1784-1788, 2005.

288. Lecker, S., V. Solomon, S. Price, Y. Kwon, W. Mitch and A. Goldberg, "Ubiquitin conjugation by the N-end rule pathway and mRNAs for its components increase in muscles of diabetic rats", *Journal of Clinical Investigations*, Vol. 104, pp. 1411–1420, 1999.

289. Marfella, R., M. D'Amico, K. Esposito, A. Baldi, C. Filippo, M. Siniscalchi, F. Sasso, M. Portoghese, F. Cirillo, F. Cacciapuoti, O. Carbonara, B. Crescenzi, F. Baldi, A. Ceriello, G. Nicoletti, F. D'Andrea, M. Verza, L. Coppola, F. Rossi and D. Giugliano, "The ubiquitin-proteasome system and inflammatory activity in diabetic

atherosclerotic plaques: Effects of rosiglitazone treatment", *Diabetes*, Vol. 55, pp. 622–632, 2006.

290. Barbieri, M., R. Marfella, M. Rizzo, V. Boccardi, M. Siniscalchi, C. Schiattarella, S. Siciliano, P. Lemme and G. Paolisso, "The -8 UTR C/G polymorphism of PSMA6 gene is associated with susceptibility to myocardial infarction in type 2 diabetic patients", *Atherosclerosis*, Vol. 201, pp. 117–123, 2008.

291. Volinic, J., J. Lee, K. K. Eto, V and M. Thomas, "Overexpression of the Coactivator Bridge-1 Results in Insulin Deficiency and Diabetes", *Molecular Endocrinology*, Vol. 20, pp. 167–182, 2006.

292. Banz , C., B. Münchow and K. Diedrich, "Bridge-1 is expressed in human granulosa cells and is involved in the activin A signaling cascade", *Fertility and Sterility*, Vol. 93, pp. 1349–1352, 2010.

293. Kavsak P, R. R., Causing CG, Bonni S, Zhu H, Thomsen GH, Wrana JL, "Smad7 binds to Smurf2 to form an E3 ubiquitin ligase that targets the TGF beta receptor for degradation", *Molecular Cell*, Vol. 6, pp. 1365–1375, 2000.

294. Hwang, H., B. Bowen, N. Lefort, C. Flynn, E. De Filippis, C. Roberts, C. Smoke, C. Meyer, K. Højlund, Z. Yi and L. Mandarino, "Proteomics Analysis of Human Skeletal Muscle Reveals Novel Abnormalities in Obesity and Type 2 Diabetes", *Diabetes*, Vol. 59, pp. 33–42, 2010.

295. Wang, X., Z. Hu, J. Hu, J. Du and W. Mitch, "Insulin resistance accelerates muscle protein degradation: activation of the ubiquitin-proteasome pathway by defects in muscle cell signaling", *Endocrinology*, Vol. 147, pp. 4160–4168, 2006.

296. van Vugt, M., A. Gardino, R. Linding, G. Ostheimer, H. Reinhardt, S. Ong, C. Tan, H. Miao, S. Keezer, J. Li, T. Pawson, T. Lewis, S. Carr, S. Smerdon, T. Brummelkamp and M. Yaffe, "A mitotic phosphorylation feedback network connects

Cdk1, Plk1, 53BP1, and Chk2 to inactivate the G(2)/M DNA damage checkpoint", *PLoS Biology*, Vol. 8, pp. e1000287, 2010.

297. Ling, C., S. Del Guerra, R. Lupi, T. Rönn, C. Granhall, H. Luthman, P. Masiello, P. Marchetti, L. Groop and S. Del Prato, "Epigenetic regulation of PPARGC1A in human type 2 diabetic islets and effect on insulin secretion", *Diabetologia*, Vol. 51, pp. 615-622, 2008.

298. Henneke, G., S. Koundrioukoff and U. Hubscher, "Multiple roles for kinases in DNA replication", *EMBO Reports*, Vol. 4, pp. 1-5, 2003.

299. Ando, H., T. Takamura, N. Matsuzawa-Nagata, K. Shima, T. Eto, H. Misu, M. Shiramoto, T. Tsuru, S. Irie, A. Fujimura and S. Kaneko, "Clock gene expression in peripheral leucocytes of patients with type 2 diabetes", *Diabetologia*, Vol. 52, pp. 329-335, 2009.

300. Busik, J. V., M. Tikhonenko, A. Bhatwadekar, M. Opreanu, N. Yakubova, S. Caballero, D. Player, T. Nakagawa, A. Afzal, J. Kielczewski, A. Sochacki, S. Hasty, S. L. Calzi, S. Kim, S. K. Duclas, M. S. Segal, D. L. Guberski, W. J. Esselman, M. E. Boulton and M. B. Grant, "Diabetic retinopathy is associated with bone marrow neuropathy and a depressed peripheral clock", *Journal of Experimental Medicine*, Vol. 206, pp. 2897-2906, 2009.

301. Shearman, L. P., S. Sriram, D. R. Weaver, E. S. Maywood, I. Chaves, B. Zheng, K. Kume, C. C. Lee, G. T. van der Horst, nbsp, J., M. H. Hastings and S. M. Reppert, "Interacting Molecular Loops in the Mammalian Circadian Clock", *Science*, Vol. 288, pp. 1013-1019, 2000.

302. Cermakian, N. and D. B. Boivin, "A molecular perspective of human circadian rhythm disorders", *Brain Research Reviews*, Vol. 42, pp. 204-220, 2003.

303. Su, W., Z. Guo, D. Randall, L. Cassis, D. Brown and M. Gong, "Hypertension and disrupted blood pressure circadian rhythm in type 2 diabetic db/db mice", *American Journal of Physiology*, Vol. 295, pp. H1634-1641, 2008.

304. Emanuelli, B., P. Peraldi, C. Filloux, D. Sawka-Verhelle, D. Hilton and E. Van Obberghen, "SOCS3 is an insulin-induced negative regulator of insulin signaling", *Journal of Biological Chemistry*, Vol. 275, pp. 15985–15991, 2000.

305. Li, L. and C. Hölscher, "Common pathological processes in Alzheimer disease and type 2 diabetes: A review", *Brain Research Reviews*, Vol. 56, pp. 384-402, 2007.

306. Chen, C.-L., S. S. Huang and J. S. Huang, "Cholesterol modulates cellular TGF-beta responsiveness by altering TGF-beta binding to TGF-beta receptors", *Journal of Cellular Physiology*, Vol. 215, pp. 223-233, 2008.

307. Boomgaarden, I., A. Bosy-Westphal, M. Müller and F. Döring, "Influence of a type 2 diabetes associated prostaglandin E synthase 2 polymorphism on blood prostaglandin E2 levels", *Prostaglandins, Leukotrienes and Essential Fatty Acids*, Vol. 80, pp. 185-188, 2009.

308. Hata, A. and R. Breyer, "Pharmacology and signaling of prostaglandin receptors: Multiple roles in inflammation and immune modulation", *Pharmacology & Therapeutics*, Vol. 103, pp. 147-166, 2004.

309. O'Banion, M., "Prostaglandin E2 synthases in neurologic homeostasis and disease", *Prostaglandins & Other Lipid Mediators*, Vol. 91, pp. 113-117, 2010.

310. Furuyashiki, T. and S. Narumiya, "Roles of prostaglandin E receptors in stress responses", *Current Opinion in Pharmacology*, Vol. 9, pp. 31-38, 2009.

311. Hebrok, M., S. Kim, B. St-Jacques, A. McMahon and D. Melton, "Regulation of pancreas development by hedgehog signaling", *Development*, Vol. 127, pp. 4905-4913, 2000.

312. Suh, J., X. Gao, J. McKay, R. McKay, Z. Salo and J. Graff, "Hedgehog signaling plays a conserved role in inhibiting fat formation", *Cell Metabolism*, Vol. 3, pp. 25-34, 2006.

313. Watt, M., N. Dzamko, W. Thomas, S. Rose-John, M. Ernst, D. Carling, B. Kemp, M. Febbraio and G. Steinberg, "CNTF reverses obesity-induced insulin resistance by activating skeletal muscle AMPK", *Nature Medicine*, Vol. 12, pp. 541-548, 2006.

314. Steinberg, G., M. Watt, M. Ernst, M. Birnbaum, B. Kemp and S. Jørgensen, "CNTF Stimulates Muscle Glucose Uptake by a PI3-Kinase Dependent Pathway that is Impaired with Obesity", *Diabetes*, Vol. 58, pp. 829-839, 2009.

315. Matthews, V. and M. Febbraio, "CNTF: a target therapeutic for obesity-related metabolic disease? ", *Journal of Molecular Medicine*, Vol. 86, pp. 1432-1440, 2008.

316. Rezende, L., A. Vieira, A. Negro, F. Langone and A. Boschero, "Ciliary neurotrophic factor (CNTF) signals through STAT3-SOCS3 pathway and protects rat pancreatic islets from cytokine-induced apoptosis", *Cytokines*, Vol. 46, pp. 65-71, 2009.

317. Rocha, V., E. Folco, G. Sukhova, K. Shimizu, I. Gotsman, A. Vernon and P. Libby, "Interferon-gamma, a Th1 Cytokine, Regulates Fat Inflammation A Role for Adaptive Immunity in Obesity", *Circulation Research*, Vol. 103, pp. 467-476, 2008.

318. Gough, D., D. Levy, R. Johnstone and C. Clarke, "IFNgamma signaling-does it mean JAK-STAT?", *Cytokine Growth Factor*, Vol. 19, pp. 383-394, 2008.

319. Johnson, H. and C. Ahmed, "Gamma interferon signaling: insights to development of interferon mimetics", *Cellular and Molecular Biology*, Vol. 52, pp. 71-76, 2006.

320. Christian, M., R. White and M. Parker, "Metabolic regulation by the nuclear receptor corepressor RIP140", *Trends in Endocrinology and Metabolism*, Vol. 17, pp. 243-250, 2006.

321. Mejhert, N., J. Laurencikiene, A. Pettersson, M. Kaaman, B. Stenson, M. Rydén and I. Dahlman, "Role of Receptor-Interacting Protein 140 in human fat cells", *BMC Endocrine Disorders*, Vol. 10, pp. 1-8, 2010.

322. Catalán, V., J. Gómez-Ambrosi, A. Lizanzu, A. Rodríguez, C. Silva, F. Rotellar, M. Gil, J. Cienfuegos, J. Salvador and G. Frühbeck, "RIP140 Gene and Protein Expression Levels are Downregulated in Visceral Adipose Tissue in Human Morbid Obesity", *Obesity Surgery*, Vol. 19, pp. 771-776, 2009.

323. Yamada, Y., H. Matsuo, T. Segawa, S. Watanabe, K. Kato, T. Hibino, K. Yokoi, S. Ichihara, N. Metoki, H. Yoshida, K. Satoh and Y. Nozawa, "Assessment of the Genetic Component of Hypertension", *American Journal of Hypertension*, Vol. 19, pp. 1158-1165, 2006.

324. Elbers, C., N. Onland-Moret, L. Franke, A. Niehoff, Y. van der Schouw and C. Wijmenga, "A strategy to search for common obesity and type 2 diabetes genes", *Trends in Endocrinology and Metabolism*, Vol. 18, pp. 19-26, 2006.

325. Ekestern, E., "Neurotrophic factors and amyotrophic lateral sclerosis", *Neurodegenerative Diseases*, Vol. 1, pp. 88-100, 2004.

326. Giess, R., B. Holtmann, M. Braga, T. Grimm, B. Müller-Myhsok, K. Toyka and M. Sendtner, "Early onset of severe familial amyotrophic lateral sclerosis with a SOD-1 mutation: potential impact of CNTF as a candidate modifier gene", *American Journal of Human Genetics*, Vol. 70, pp. 1277-1286, 2002.

327. Bongioanni, P., C. Reali and V. Sogos, "Ciliary neurotrophic factor (CNTF) for amyotrophic lateral sclerosis/motor neuron disease", *Cochrane Database of Systemic Reviews*, Vol. 3, pp. CD004302, 2004.

328. Natal, C., M. Fortuno, P. Restituto, A. Bazan, I. Colina, J. Diez and N. Varo, "Cardiotrophin-1 is expressed in adipose tissue and upregulated in the metabolic syndrome", *American Journal of Endocrinology and Metabolism*, Vol. 294, pp. E52-E60, 2008.

329. Lesbordes, J., C. Cifuentes-Diaz, A. Miroglio, V. Joshi, T. Bordet, A. Kahn and J. Melki, "Therapeutic benefits of cardiotrophin-1 gene transfer in a mouse model of spinal muscular atrophy", *Human Molecular Genetics*, Vol. 12, pp. 1233-1239, 2003.

330. Bordet, T., J. Lesbordes, S. Rouhani, L. Castelnau-Ptakhine, H. Schmalbruch, G. Haase and A. Kahn, "Protective effects of cardiotrophin-1 adenoviral gene transfer on neuromuscular degeneration in transgenic ALS mice", *Human Molecular Genetics*, Vol. 10, pp. 1925-1933, 2001.

331. McCampbell, A., J. Taylor, A. Taye, J. Robitschek, M. Li, J. Walcott, D. Merry, Y. Chai, H. Paulson, G. Sobue and K. Fischbeck, "CREB-binding protein sequestration by expanded polyglutamine", *Human Molecular Genetics*, Vol. 9, pp. 2197-2202, 2000.

332. Yamauchi, T., Y. Oike, J. Kamon, H. Waki, K. Komeda, A. Tsuchida, Y. Date, M.-X. Li, H. Miki, Y. Akanuma, R. Nagai, S. Kimura, T. Saheki, M. Nakazato, T. Naitoh, K. Yamamura and T. Kadowaki, "Increased insulin sensitivity despite lipodystrophy in Crebbp heterozygous mice", *Nature Genetics*, Vol. 30, pp. 221-226, 2002.

333. Sun, G., K. Haginoya, Y. Wu, Y. Chiba, T. Nakanishi, A. Onuma, Y. Sato, M. Takigawa, K. Iinuma and S. Tsuchiya, "Connective tissue growth factor is overexpressed in muscles of human muscular dystrophy", *Journal of the Neurological Sciences*, Vol. 267, pp. 48-56, 2008.

334. Kollias, H. and J. McDermott, "Transforming growth factor-β and myostatin signaling in skeletal muscle", *Journal of Applied Physiology*, Vol. 104, pp. 579-587, 2008.

335. Handschin, C. and B. Spiegelman, "The role of exercise and PGC1[alpha] in inflammation and chronic disease", *Nature*, Vol. 454, pp. 463-469, 2008.

336. Huttenhower, C., E. Haley, M. Hibbs, V. Dumeaux, D. Barrett, H. Coller and O. Troyanskaya, "Exploring the human genome with functional maps", *Genome Research*, Vol. 19, pp. 1093-1106, 2009.

337. Bertram, L., M. McQueen, K. Mullin, D. Blacker and R. Tanzi, "Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database", *Nature Genetics*, Vol. 39, pp. 17-23, 2007.

338. Rebhan, M., V. Chalifa-Caspi, J. Prilusky and D. Lancet, "GeneCards: integrating information about genes, proteins and diseases", *Trends in Genetics*, Vol. 13, pp. 163-, 1997.

339. Bertram, L., "Alzheimer's Disease Genetics: Current Status and Future Perspectives", *International Review of Neurobiology*, Vol. 84, pp. 167-184, 2009.

340. Grupe, A., R. Abraham, Y. Li, C. Rowland, P. Hollingworth, A. Morgan, L. Jehu, R. Segurado, D. Stone, E. Schadt, M. Karnoub, P. Nowotny, K. Tacey, J. Catanese, J. Sninsky, C. Brayne, D. Rubinsztein, M. Gill, B. Lawlor, S. Lovestone, P. Holmans, M. O'Donovan, J. Morris, L. Thal, A. Goate, M. Owen and J. Williams, "Evidence for novel susceptibility genes for late-onset Alzheimer's disease from a genome-wide association study of putative functional variants", *Human Molecular Genetics*, Vol. 16, pp. 865-873, 2007.

341. Waring, S. and R. Rosenberg, "Genome-Wide Association Studies in Alzheimer Disease", *Archieves of Neurology*, Vol. 65, pp. 329-334, 2008.

342. Ding, Q., W. Markesbery, Q. Chen, F. Li and J. Keller, "Ribosome Dysfunction Is an Early Event in Alzheimer's Disease", *Journal of Neuroscience*, Vol. 25, pp. 9171-9175, 2005.

343. Monsonego, A., V. Zota, A. Karni, J. Krieger, A. Bar-Or, G. Bitan, A. Budson, R. Sperling, D. Selkoe and H. Weiner, "Increased T cell reactivity to amyloid Î² protein in older humans and patients with Alzheimer disease", *Journal of Clinical Investigation*, Vol. 112, pp. 415-422, 2003.

344. Blalock, E., J. Geddes, K. Chen, N. Porter, W. Markesbery and P. Landfield, "Incipient Alzheimer's disease: Microarray correlation analyses reveal major transcriptional and tumor suppressor responses", *Proceedings of the National Academy of Sciences*, Vol. 17, pp. 2173-2178, 2004.

345. Ballesteros, M., A. Fredriksson, J. Henriksson and T. Nystrom, "Bacterial senescence: protein oxidation in non-proliferating cells is dictated by the accuracy of the ribosomes", *EMBO Journal*, Vol. 20, pp. 5280-5289, 2001.

346. Keller, J., F. Schmitt, S. Scheff, Q. Ding, Q. Chen, D. Butterfield and W. Markesbery, "Evidence of increased oxidative damage in subjects with mild cognitive impairment", *Neurology*, Vol. 64, pp. 1152-1156, 2005.

347. Langstrom, N., J. Anderson, H. Lindroos, B. Winbland and W. Wallace, "Alzheimer's disease-associated reduction of polysomal mRNA translation", *Molecular Brain Research*, Vol. 5, pp. 259-269, 1989.

348. Honda, K., M. Smith, X. Zhu, D. Baus, W. Merrick, A. Tartakoff, T. Hattier, P. Harris, S. Siedlak, H. Fujioka, Q. Liu, P. Moreira, F. Miller, A. Nunomura, S. Shimohama and G. Perry, "Ribosomal RNA in Alzheimer Disease Is Oxidized by Bound Redox-active Iron", *Journal of Biological Chemistry*, Vol. 280, pp. 20978-20986, 2005.

349. Li, X., I. Alafuzoff, H. Soininen, B. Winblad and J. Pei, "Levels of mTOR and its downstream targets 4E-BP1, eEF2, and eEF2 kinase in relationships with tau in Alzheimer's disease brain", *FEBS Journal*, Vol. 272, pp. 4211-4220, 2005.

350. Akiyama, H., K. Ikeda, H. Kondo and P. McGeer, "Thrombin accumulation in brains of patients with Alzheimer's disease", *Neuroscience Letters*, Vol. 146, pp. 152-154, 1992.

351. Xu, G., H. Zhang, S. Zhang, X. Fan and X. Liu, "Plasma fibrinogen is associated with cognitive decline and risk for dementia in patients with mild cognitive impairment", *International Journal of Clinical Practice*, Vol. 62, pp. 1070-1075, 2008.

352. van Oijen, M., J. Witteman, A. Hofman, P. Koudstaal and M. Breteler, "Fibrinogen Is Associated With an Increased Risk of Alzheimer Disease and Vascular Dementia", *Stroke*, Vol. 36, pp. 2637-2641, 2005.

353. Melchor, J., R. Pawlak and S. Strickland, "The Tissue Plasminogen Activator-Plasminogen Proteolytic Cascade Accelerates Amyloid-$^{\beta}$ (A^{β}) Degradation and Inhibits A^{β}-Induced Neurodegeneration", *Journal of Neuroscience*, Vol. 23, pp. 8867-8871, 2003.

354. Powell, J., C. Twomey, R. Jain and J. McCarthy, "Association between Presenilin-1 and TRAF6 modulates regulated intramembrane proteolysis of the $p75^{NTR}$ neurotrophin receptor", *Journal of Neurochemistry*, Vol. 108, pp. 216-230, 2009.

355. Combs, C., J. Karlo, S. Kao and G. Landreth, "β-Amyloid Stimulation of Microglia and Monocytes Results in TNFα-Dependent Expression of Inducible Nitric Oxide Synthase and Neuronal Apoptosis", *Journal of Neuroscience*, Vol. 21, pp. 1179-1188, 2001.

356. Zhang, G. and G. Sankar, "Molecular mechanisms of NF-κB activation induced by bacterial lipopolysaccharide through Toll-like receptors", *Journal of Endotoxin Research*, Vol. 6, pp. 453-457, 2000.

357. Jacobson, N., S. Szabo, R. Weber-Nordt, Z. Zhong, R. Schreiber, J. Darnell and K. Murphy, "Interleukin 12 signaling in T helper type 1 (Th1) cells involves tyrosine phosphorylation of signal transducer and activator of transcription (Stat)3 and Stat4", *The Journal of Experimental Medicine*, Vol. 181, pp. 1755-1762, 1995.

358. Bright, J., C. Natarajan, S. Sriram and G. Muthian, "Signaling through JAK2-STAT5 pathway is essential for IL-3-induced activation of microglia", *Glia*, Vol. 45, pp. 188-196, 2004.

359. Jarvik, G., E. Wijsman, W. Kukull, G. Schellenberg, C. Yu and E. Larson, "Interactions of apolipoprotein E genotype, total cholesterol level, age, and sex in prediction of Alzheimer's disease: a case-control study", *Neurology*, Vol. 45, pp. 1092-1096, 1995.

360. Sparks, D., T. Martin, D. Gross and J. Hunsaker, "Link between heart disease, cholesterol, and Alzheimer's disease: a review", *Microscopy Reseach and Techique*, Vol. 50, pp. 287-290, 2000.

361. Refolo, L., M. Pappolla, B. Malester, J. LaFrancois, T. Bryant-Thomas, R. Wang, G. Tint, K. Sambamurti and K. Duff, "Hypercholesterolemia Accelerates the Alzheimer's Amyloid Pathology in a Transgenic Mouse Model", *Neurobiology of Disease*, Vol. 7, pp. 321-331, 2000.

362. Burns, M., W. Noble, V. Olm, K. Gaynor, E. Casey, J. LaFrancois, L. Wang and K. Duff, "Co-localization of cholesterol, apolipoprotein E and fibrillar Aβ in amyloid plaques", *Molecular Brain Research*, Vol. 110, pp. 119-125, 2003.

363. Jick, H., G. Zornberg, S. Jick, S. Seshadri and D. Drachman, "Statins and the risk of dementia", *The Lancet*, Vol. 356, pp. 1627-1631, 2000.

364. Tournoy, J., X. Bossuyt, A. Snellinx, M. Regent, M. Garmyn, L. Serneels, P. Saftig, K. Craessaerts, B. De Strooper and D. Hartmann, "Partial loss of presenilins causes seborrheic keratosis and autoimmune disease in mice", *Human Molecular Genetics*, Vol. 13, pp. 1321-1331, 2004.

365. Parks, A. and D. Curtis, "Presenilin diversifies its portfolio", *Trends in Genetics*, Vol. 23, pp. 140-150, 2007.

366. Verdile, G., S. Gandy and R. Martins, "The Role of Presenilin and its Interacting Proteins in the Biogenesis of Alzheimer's Beta Amyloid", *Neurochemical Research*, Vol. 32, pp. 609-623, 2007.

367. Laske, C., K. Stellos, G. Eschweiler, T. Leyhe and M. Gawaz, "Decreased CXCL12 (SDF-1) plasma levels in early Alzheimer's disease: a contribution to a deficient hematopoietic brain support?", *Journal of Alzheimer's Disease*, Vol. 15, pp. 83-95, 2008.

368. Segret, A., C. Rucker-Martin, C. Pavoine, J. Flavigny, E. Deroubaix, M. Chatel, A. Lombet and J. Renaud, "Structural Localization and Expression of CXCL12 and CXCR4 in Rat Heart and Isolated Cardiac Myocytes", *Journal of Histochemistry and Cytochemistry*, Vol. 55, pp. 141-150, 2007.

369. Zhuang, Y., X. Chen, M. Xu, L. Zhang and F. Xiang, "Chemokine stromal cell-derived factor 1/CXCL12 increases homing of mesenchymal stem cells to injured myocardium and neovascularization following myocardial infarction", *Chinese Medical Journal*, Vol. 122, pp. 183-187, 2009.

370. Liao, T., X. Yang, Y. Liu, G. Shesely, M. Cavasin, W. Kuziel, P. Pagano and O. Carretero, "Role of Inflammation in the Development of Renal Damage and Dysfunction in Angiotensin II–Induced Hypertension", *Hypertension*, Vol. 52, pp. 256-263, 2008.

371. Ford, I., T. Singh, S. Kitchen, M. Makris, J. Ward and F. Preston, "Activation of Coagulation in Diabetes Mellitus in Relation to the Presence of Vascular Complications", *Diabetic Medicine*, Vol. 8, pp. 322-329, 1991.

372. Calles-Escandon, J. and M. Cipolla, "Diabetes and Endothelial Dysfunction: A Clinical Perspective", *Endocrinology Reviews*, Vol. 22, pp. 36-52, 2001.

373. Hayek, T., J. Stephens, C. Hubbart, J. Acharya, M. Caslake, E. Hawe, G. Miller, S. Hurel and S. Humphries, "A common variant in the glutathione S transferase gene is associated with elevated markers of inflammation and lipid peroxidation in subjects with diabetes mellitus", *Atherosclerosis*, Vol. 184, pp. 404-412, 2006.

374. Ceriello, A., R. P. dello, P. Amstad and P. Cerutti, "High glucose induces antioxidant enzymes in human endothelial cells in culture. Evidence linking hyperglycemia and oxidative stress", *Diabetes*, Vol. 45, pp. 471-477, 1996.

375. Ihara, Y., Y. Yamada, S. Toyokuni, K. Miyawaki, N. Ban, T. Adachi, A. Kuroe, T. Iwakura, A. Kubota, H. Hiai and Y. Seino, "Antioxidant alpha-tocopherol ameliorates glycemic control of GK rats, a model of type 2 diabetes", *FEBS Letters*, Vol. 473, pp. 24-26, 2000.

376. Perry, T., V. Yong, C. Bergeron, S. Hansen and K. Jones, "Amino acids, glutathione, and glutathione transferase activity in the brains of patients with Alzheimer's disease", *Annalts of Neurology*, Vol. 21, pp. 331-336, 1987.

377. Lovell, M., W. Ehmann, S. Butler and W. Markesbery, "Elevated thiobarbituric acid-reactive substances and antioxidant enzyme activity in the brain in Alzheimer's disease", *Neurology*, Vol. 45, pp. 1594-1601, 1995.

378. Marcus, D., C. Thomas, C. Rodriguez, K. Simberkoff, J. Tsai, J. Strafaci and M. Freedman, "Increased peroxidation and reduced antioxidant enzyme activity in Alzheimer's disease", *Experimental Neurology*, Vol. 150, pp. 40-44, 1998.

379. Takeda, S., N. Sato, K. Uchio-Yamada, K. Sawada, T. Kunieda, D. Takeuchi, H. Kurinami, M. Shinohara, H. Rakugi and R. Morishita, "Diabetes-accelerated memory dysfunction via cerebrovascular inflammation and Aβ deposition in an Alzheimer mouse model with diabetes", *Proceedings of the National Academy of Sciences*, Vol. 107, pp. 7036-7041, 2010.