Network-based identification of genetic factors in Ageing, lifestyle and Type 2 Diabetes that Influence in the progression of Alzheimer's disease

1st Utpala Nanda Chowdhury
Department of Computer Science and Engineering
University of Rajshahi
Rajshahi, Bangladesh
unchowdhury@ru.ac.bd

2nd Md. Babul Islam
Department of Applied Physics and Electronic Engineering
University of Rajshahi
Rajshahi, Bangladesh
babul.apee@ru.ac.bd

3rd Shamim Ahmad
Department of Computer Science and Engineering
University of Rajshahi
Rajshahi, Bangladesh
shamim_cst@ru.ac.bd

5th Mohammad Ali Moni
Discipline of Biomedical Science, Faculty of Medicine and Health
The University of Sydney
NSW, Australia
mohammad.moni@sydney.edu.au

***Abstract*—Alzheimer's disease (AD) is an incurable disease, and the causative risk factors, especially the modifiable ones, are still poorly understood which impedes the effective prevention and remedy strategies. We proposed a network-based quantitative framework to reveal the complex relationship of various biasing genetic factors for the AD. We analyzed gene expression microarray data from the AD, ageing, severe alcohol consumption, type II diabetes, high body fat, high dietary fat, obesity, dietary red meat, sedentary lifestyle, smoking, and control datasets. We have developed genetic associations and diseasome network of various factors with the AD based on the neighborhood-based benchmarking and multilayer network topology approaches. The study identified 484 differentially expressed genes of the AD. Among them, 27 genes were highly expressed in both for the AD and smoking whereas the number of genes is 21 for the AD and type II diabetes, and 12 for the AD and sedentary lifestyle. However, AD shared less than ten significant biomarkers with other factors. Notably, 3 significant genes, HLA-DRB4, IGH and IGHA2 are commonly up-regulated among the AD, type II diabetes and alcohol consumption; 2 significant genes IGHD and IGHG1 are commonly up-regulated among the AD, type II diabetes, alcohol consumption and sedentary lifestyle. Protein-protein interaction network identified 10 hub genes: CREBBP, PRKCB, ITGB1, GAD1, GNB5, PPP3CA, CABP1, SMARCA4, SNAP25 and GRIA1. Ontological and pathway analyses have identified significant gene ontology and molecular pathways that enhance our understanding of the fundamental molecular procedure of AD progression. Therapeutic targets of the AD could be developed using these identified target genes, ontologies and pathways. Online Mendelian Inheritance in Man (OMIM) and dbGaP databases were used for gold benchmark gene-disease associations to validate the significance of these identified target genes in AD progression. Our formulated methodologies demonstrate a network-based approach to understand the disease mechanism and the causative reason for the AD, and the identification of therapeutic targets for the AD.**

***Index Terms*—Alzheimers disease, Ageing, Type 2 diabetes, Alcohol consumption, Smoking**

I. Introduction

Alzheimer's disease (AD) is the most common form of dementia characterized by gradual degeneration in memory, thinking, language, and learning capacity [1]. Initial signs begin with the inability of forming recent memories, but inevitably affect all cognitive functions resulting in complete dependency for basic daily activities, and fostering premature death [2]. It is an irremediable disease-causing 60 to 80% of all dementia cases and affecting over 24 million people worldwide. In the United States, 93,541 deaths from AD were officially recorded in 2014. The AD is ranked sixth among all causes of death in the United States and fifth among all causes of death after 65 years of age. The death rate of the AD has increased by 89% within five years till 2010, whereas death rates of other major diseases like cardiac disease, stroke, breast and prostate cancer, and AIDS have declined in the same time frame. Currently, one new case of the AD is developed in every 66 seconds, which is estimated to drop down to 33 seconds by 2050, resulting in nearly 1 million new cases per year [3].

The etiology of the AD is not clearly comprehended so far, but it is hypothesized that both genetic and environmental factors are the primary causes. Genes encoding amyloid precursor protein (APP), presenilin 1 (PSEN1) and presenilin 2 (PSEN2) have been identified as associated with AD [4]. Age is found to be the most influential risk factor of the AD along with a sedentary lifestyle. Typically AD is developed after the age of 65 years and almost half individuals over 85 years old have AD [5]. Obesity also increases the risk of dementia and eventually AD [6]. Type II diabetes, hypertension, smoking and dietary fats can increase the risk of developing AD [7] [8] [9]. Meta-analysis of prospective studies suggests that alcohol

consumption in late life yields reduced the risk of dementia and hence reduced the risk of AD [10]. The AD is a complex polygenic disorder, and many of the associated factors are yet to be unfolded. The pathogenesis of AD continues to be poorly comprehended despite constant discoveries in this field. Therefore, problems with accurate diagnosis, characterizing heterogeneous groups of patients who may respond differently to treatment have even convoluted the development of effective treatment. Only the discovery of further genetic factors could lead not only to the development of a better diagnostic profile but to a clearer understanding of the disease process [11]. The common genetic factors associated with susceptibility to complex diseases can be effectively unfolded by genome-wide association studies, and its usefulness has been proven empirically. This method aims to identify genetic factors influencing the common complex condition of such diseases against the background of random variation seen in a population as a whole [12].

Molecular associations, such as differential gene expressions, protein-protein interactions (PPIs), gene ontologies and metabolic pathways can ascribe genetic relation of various risk factors as the important causes of the disease [13] [14]. Any influential factor can be attributed to the disease if they both share the same set of differentially expressed genes [15] [16]. But from a proteomics point of view, they are associated through biological modules such as PPIs, gene ontologies or molecular pathways [17] [18].

Recently much attention is given to network-based integrative analyses by researchers in recent years to elucidate the possible roles of biomolecules in different complex diseases [19] [20]. A number of genetic studies have been conducted on the AD [21] [22] [23]. However, their findings were limited at the transcript level, since the functional interactions among the genes products were not considered. Since the biological molecules interact with each other to carry out functions in biological processes in cells and tissues, integrative analysis within network medicine context is essential to understand the molecular mechanisms behind diseases and to identify critical biomolecules. Therefore, in this article, a network-based analysis to determine the genetic influence caused by associated factors and disorders for the Alzheimer's progression is demonstrated, studying gene expression profiling, PPI sub-network, gene ontologies and molecular pathways. An extensive study regarding phylogenetic and pathway analysis is conducted to reveal the genetic associations of the AD. Significance of these genes and pathways in AD processes were also validated with gold benchmarking datasets including Online Mendelian Inheritance in Man (OMIM) and dbGaP gene-disease associations databases.

II. Materials and Methods

A. Data

We have analyzed gene expression microarray datasets to identify the association of different factors with the AD at the molecular level. All the datasets used in this study were collected from the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus database (https://www.ncbi.nlm.nih.gov/geo/). Ten different datasets with accession numbers: GSE1297, GSE23343, GSE15524, GSE25941, GSE1786, GSE68231, GSE6573, GSE25220, GSE52553, and GSE4806 were analyzed for this study [24] [25] [26] [27] [28] [29] [30] [31] [32] [33]. The AD dataset (GSE1297) is obtained by gene expression profiling of hippocampal tissues on 31 separate microarrays from nine control subjects and 22 AD patients with varying severity. The type II diabetes dataset (GSE23343) contains gene expression data obtained through extensive analysis after conducting liver biopsies in humans. The source of the obesity dataset (GSE15524) is subcutaneous and omental adipose tissue analyzed through expression profiling of 20,000 probes in 28 tissue samples. The age dataset (GSE25941) is a global microarray data from skeletal muscle transcriptome of 28 different subjects. The sedentary lifestyle dataset (GSE1786) was obtained by expression profiling array from the vastus lateralis muscle using needle biopsies. The high-fat diet (HFD) dataset (GSE68231) is the expression data from human skeletal muscle identifying accumulation of intramyocellular lipid (IMCL). The high body fat (HBF) dataset (GSE6573) is an Affymetrix human gene expression array data from the abdominal fat tissue. The red meat dietary intervention dataset (GSE25220) is an Agilent-014850 whole human genome microarray data from human colon biopsies before and after participating in a high red-meat dietary intervention. The alcohol consumption dataset (GSE52553) is an Affymetrix human gene expression array data of Lymphoblastoid cells from 21 alcoholics and 21 control subjects. The smoking dataset (GSE4806) is a gene expression profiles of T-lymphocytes from smokers and non-smokers.

B. Method

Analyzing oligonucleotide microarray data for gene expression is known to be an effective and responsive approach to demonstrate the molecular assessors of human diseases. In this study, we used this methodology along with global transcriptome analysis to investigate the gene expression profiles of the AD with 8 risk factors and type II diabetes. To mitigate the problems involving messenger RNA (mRNA) data comparison using different platforms and experimental set-ups, we normalized each gene expression data for each disease using the Z-score (or zero mean) transformation for both disease and control state. Each sample of gene expression matrix was normalized using mean and standard deviation. The expression value of gene i in sample j represented by g_{ij} was transformed into Z_{ij} by computing

$$Z_{ij} = \frac{g_{ij} - mean(g_i)}{SD(g_i)} \quad (1)$$

where SD is the standard deviation. Comparing values of gene expression for various samples and diseases are made possible by this transformation.

Data were transformed using log_2 and differentially expressed genes for both disease and control states were obtained by

performing unpaired student t-test, and significant genes were identified by using threshold values. A threshold for p-value and absolute log Fold Change (logFC) values were set to at most 0.05 and at least 1.0 respectively. We built two infectome-diseasome relationships networks using Cytoscape (v3.5.1) [34] for both up-regulated and down-regulated genes focusing on the AD. Each node of the networks are either diseases or associative factors. These networks can also be considered as bipartite graphs where diseases or factors are connected when they share at least 1 differentially expressed gene.

We used the web-based visualization software STRING [35] for the construction and analysis of the Protein-Protein Interaction (PPI) network which was further analyzed by Cytoscape. An undirected graph representation was used for the PPI network, where the nodes indicate proteins and the edges symbolize the interactions between the proteins. We performed a topological analysis using Cyto-Hubba plugin [36] to identify highly connected proteins (i.e., hub proteins) in the network and the degree metrics were employed [37]. For further introspection into the metabolic pathways of the AD, we incorporated the pathway and gene ontology analysis on all the differentially expressed genes that were common between the AD and other risk factors and diseases using the web-based gene set enrichment analysis tool EnrichR [38]. In this analysis, the Gene Ontology (GO) Biological Process (BP) and KEGG pathway databases were selected as annotation sources. For statistical significance, the highest adjusted p-value was considered 0.05 to obtain enrichment results. Obtained GO and pathway were further analyzed by Cytoscape. Moreover, two gold bench mark validated datasets, OMIM (https://www.omim.org/) and dbGaP (https://www.ncbi.nlm.nih.gov/gap) were included in our study to validate the principle of our network based approach.

III. Results

A. Gene Expression Analysis

To identify the dysregulated genes due to the AD the gene expression patterns from hippocampal CA1 tissues of AD patients were analyzed and compared with normal subject using NCBI's GEO2R online tool (https://www.ncbi.nlm.nih.gov/geo/geo2r/?acc=GSE1297) [24]. 484 genes (p-value at most 0.05 and absolute log_2 fold change value at least 1.0) were found to be differentially expressed compared to healthy subjects where 336 genes were up-regulated and 148 genes were down-regulated.

In order to investigate the relationship of the AD with 8 risk factors and type II diabetes, we performed several steps of statistical analysis for mRNA microarray data regarding each risk factors and disease. Thus, we selected the most significant over and under regulated genes for each risk factor and disease. Our analysis identified a large number of disregulated genes (958 in ageing, 1405 in alcohol consumption, 824 in HBF, 739 in HFD, 381 in obesity, 482 in dietary red meat, 800 in sedentary lifestyle, 400 in smoking and 1438 in diabetes).

The common over- and under- expressed genes between the AD and other risk factors and diseases were also detected through a cross-comparative analysis. The findings demonstrated that the AD shares total 35, 34, 18, 15, 13, 10, 8, 7 and 4 significant genes with type II diabetes, alcohol consumption, sedentary lifestyle, ageing, HFD, obesity, smoking, HBF and dietary red meat respectively. Two infectome-diseasome associations networks centered on the AD were built using Cytoscape to identify statistically significant associations among these risk factors and diseases. Network shown in Fig. 1 interprets the association among up- regulated genes and another network shown in Fig. 2 depicts relations between among down regulated genes. Notably, 3 significant genes, HLA-DRB4, IGH and IGHA2 are commonly up regulated among the AD, type II diabetes and alcohol consumption; 2 significant genes IGHD and IGHG1, are commonly up regulated among the AD, type II diabetes, alcohol consumption and sedentary lifestyle. It is noteworthy that, relatively higher number of differentially expressed genes is obtained between the AD and type II diabetes, whereas the AD and dietary red meat share only 4 differentially expressed genes.

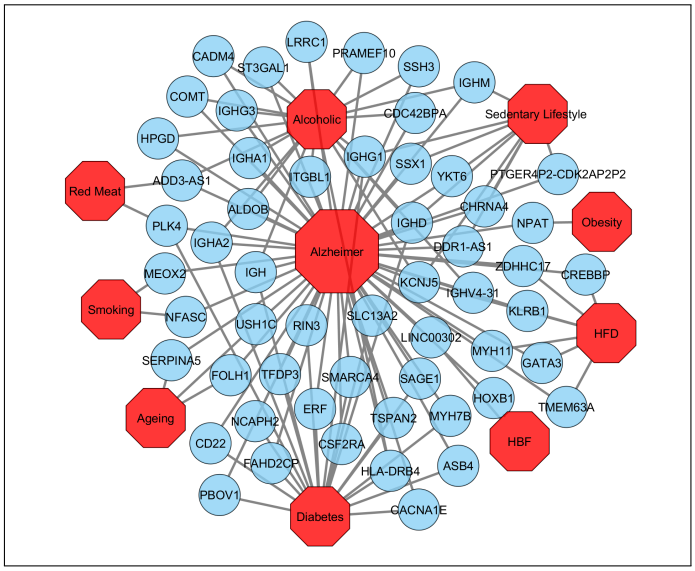

Fig. 1. Diseasome network of the AD with type II diabetes, ageing, sedentary lifestyle, HFD, HBF, dietary red meat, alcohol consumption, obesity and smoking. Red-colored octagon-shaped nodes represent categories of factors and or disease, and round-shaped sky blue-colored nodes represent up-regulated genes that are common for the AD with the other risk factors and or diseases. A link is placed between a risk factor or disease and gene if alteration of that gene leads to the specific disorder.

B. Protein-Protein Interaction Network Analysis

The PPI network was constructed using all the distinct 108 from total 144 differentially expressed genes that were common among the AD and other risk factors and diseases (Fig. 3). Each node in the network represents a protein and an edge indicates the interaction between two proteins. The network is also grouped into 9 clusters representing risk factors and diseases to depict the protein belongings. It is notable that, KCNJ5 protein belongs to the maximum 3 clusters indicating that it is the common among the AD, alcohol consumption, HFD and sedentary lifestyle and interacts with other proteins

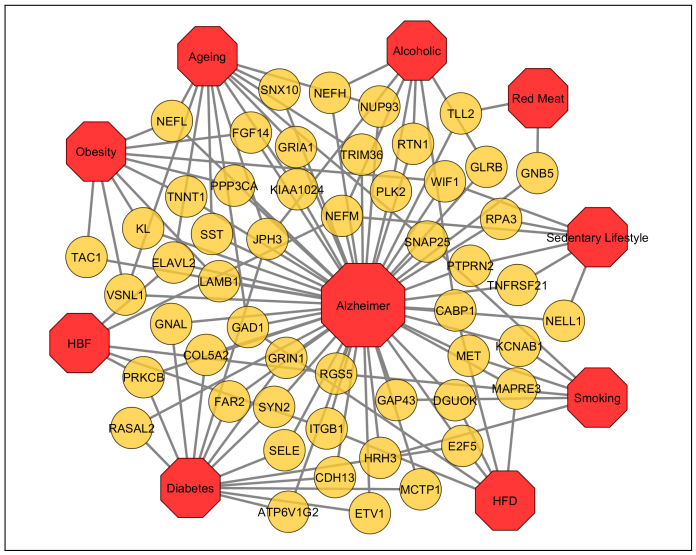

Fig. 2. Diseasome network of the AD with type II diabetes, ageing, sedentary lifestyle, HFD, HBF, dietary red meat, alcohol consumption, obesity and smoking. Red-colored octagon-shaped nodes represent categories of factors and or disease, and round-shaped yellow-colored nodes represent down-regulated genes that are common for the AD with the other risk factors and or diseases. A link is placed between a risk factor or disease and gene if alteration of that gene leads to the specific disorder.

from different clusters. However, each of the proteins PLK4, PRKCB, E2F5, GAD1, VSNL1, RGS5, ITGB1, CABP1 and NEFM belong to two clusters, and interact with other proteins in the network. For topological analysis, a simplified PPI network was constructed using Cyto-Hubba plugin to show 10 most significant hub proteins (Fig. 4), which are CREBBP, PRKCB, ITGB1, GAD1, GNB5, PPP3CA, CABP1, SMARCA4, SNAP25 and GRIA1. The identified most significant hub protein CREBBP (CREB binding protein) plays major role during the evolution of central nervous system. Alteration of CREBBP activity is evidenced to be responsible in the AD progression [39]. These hub proteins could be the targeted proteins for the drug development.

C. Pathway and Functional Correlation Analysis

In order to identify the molecular pathways associated with the AD and predicted links to the affected pathways, we performed pathway analysis on all the differentially expressed genes that were common among the AD and other risk factors and diseases using the KEGG pathway database (http://www.genome.jp/kegg/pathway.html) and the web-based gene set enrichment analysis tool EnrichR [38]. Total 115 pathways were found to be over-represented among several groups. Notably, nine significant pathways that are related to the nervous system were found which are Long-term potentiation (hsa04720), Synaptic vesicle cycle (hsa04721), Retrograde endocannabinoid signaling (hsa04723), Glutamatergic synapse (hsa04724), Cholinergic synapse (hsa04725), Serotonergic synapse (hsa04726), GABAergic synapse (hsa04727), Dopaminergic synapse (hsa04728), and Long-term depression (hsa04730). These pathways along with some other common pathways found are shown in Table I. A gene and pathway association is analyzed by constructing a network for the resulted pathways using Cytoscape (Fig. 5). Besides, we dug up the over-represented ontological groups by performing gene biological process ontology enrichment analysis using EnrichR on the commonly dysregulated genes between the AD and other risk factors and diseases. Total 215 significant gene ontology groups including peripheral nervous system neuron development (GO:0048935), neurotransmitter transport (GO:0006836), neuromuscular synaptic transmission (GO:0007274), peripheral nervous system development (GO:0007422), negative regulation of neurological system process (GO:0031645), regulation of neurotransmitter secretion (GO:0046928), regulation of neuronal synaptic plasticity (GO:0048168), autonomic nervous system development (GO:0048483), sympathetic nervous system development (GO:0048485), neuromuscular process controlling balance (GO:0050885), neuron apoptotic process (GO:0051402), regulation of neurotransmitter transport (GO:0051588) and neuroepithelial cell differentiation (GO:0060563) were observed (see Table II). A gene and gene ontology association network is constructed for the obtained gene ontology using Cytoscape (Fig. 6).

IV. Discussion

In this study, we investigated the molecular mechanism of the AD and its genetic association with other risk factors and diseases. For this purpose, we conducted analysis in gene expression of AD patients, molecular key pathways, gene ontologies and PPIs. These analyses through network-based approach can unfold novel relationships between the AD and other susceptibility/risk factor. Findings could be very potential and have not been possessed by any previous individual studies. Our outcomes identified several significant genes that yield an opportunity to identify therapeutic targets for the AD. Besides this, our analysis also identified and characterized various biological functions related to these genes.

Our gene expression analysis showed that the AD is strongly associated with type II diabetes (35 genes), alcohol consumption (34 genes), sedentary lifestyle (18 genes) and ageing (15 genes) as they share the maximum dysregulated genes. We constructed and analyzed the PPI network to have a better understanding of the central mechanism behind the AD. For this reason, to construct a PPI network around the differentially expressed genes for our study, we have combined the results of statistical analyses with the protein interactome network. For finding central proteins (i.e., hubs), topological analysis strategies were employed. These identified Hubs proteins might be considered as candidate biomarkers or potential drug targets. From the PPI network analysis, it is observed that 10 hub genes (CREBBP, PRKCB, ITGB1, GAD1, GNB5, PPP3CA, CABP1, SMARCA4, SNAP25 and GRIA1) are involved in the AD.

Besides, disease- related genes play a vital role in the human interactomes via the pathways. In this study, we identified nine significant pathways that are associated with the nervous

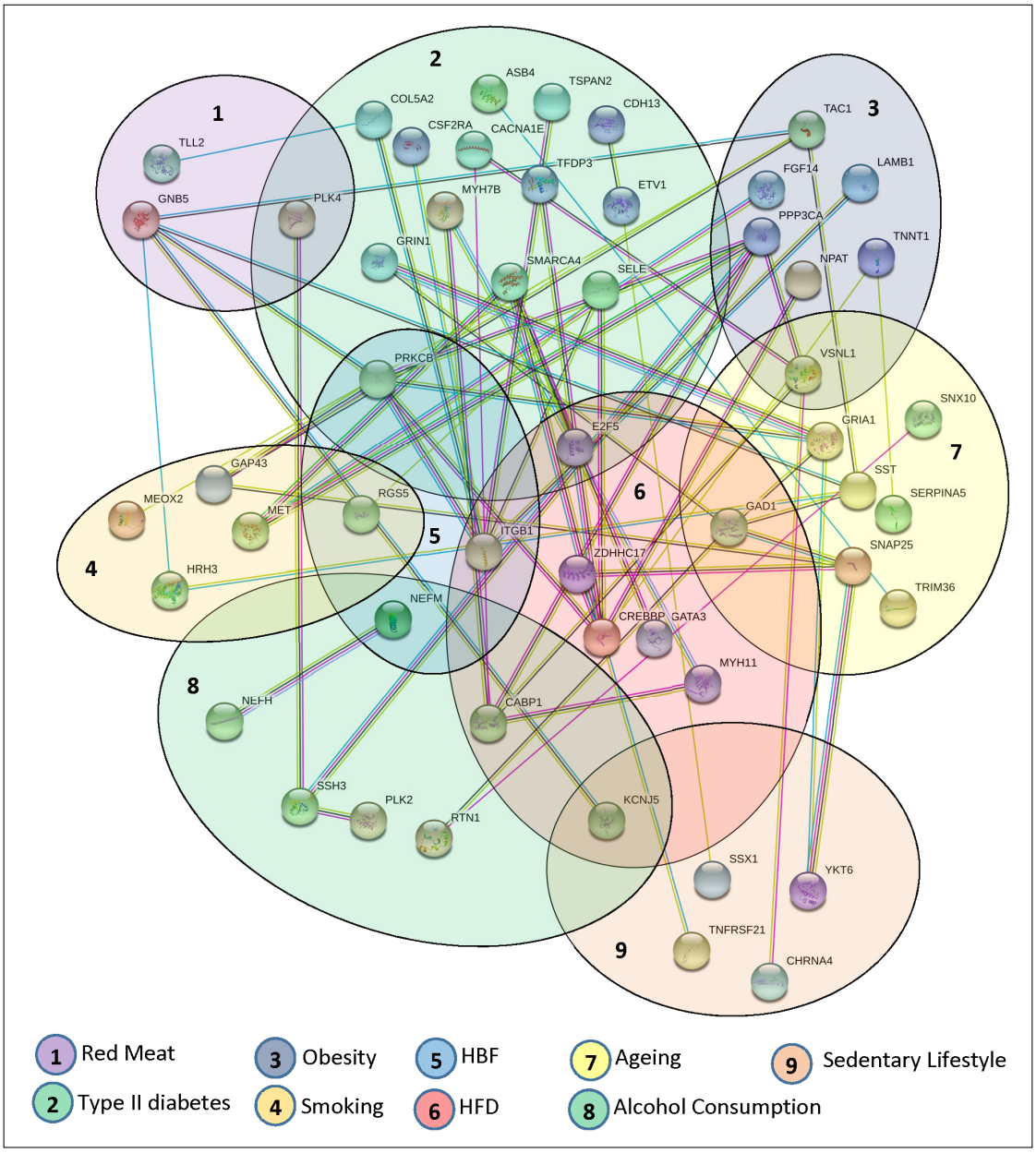

Fig. 3. Protein-Protein interaction network of commonly dysregulated genes among AD and other risk factors and diseases. Each cluster indicates the gene belongings.

system which are Long-term potentiation, Synaptic vesicle cycle, Retrograde endocannabinoid signaling, Glutamatergic synapse, Cholinergic synapse, Serotonergic synapse, GABAergic synapse, Dopaminergic synapse, and Long-term depression. Our study also identified several gene ontology groups including peripheral nervous system neuron development, neurotransmitter transport, neuromuscular synaptic transmission, peripheral nervous system development, negative regulation of neurological system process, regulation of neurotransmitter secretion, regulation of neuronal synaptic plasticity, autonomic nervous system development, sympathetic nervous system development, neuromuscular process controlling balance, neuron apoptotic process, regulation of neurotransmitter transport and neuroepithelial cell differentiation, which are closely related to the nervous system.

We have also analyzed the differentially expressed genes of each risk factor and type II diabetes with OMIM and dbGaP databases using EnrichR to validate our identified results using the valid gold benchmark gene-disease associations. Table III shows the genes of each risk factor/disease that are resulted to be associated with the AD. These results corroborate that, the differentially expressed genes of 8 risk factors and type II diabetes are responsible for the AD. As a whole, our findings compensate a major gap of about AD biology. It will also open up an entry point to establish a mechanic link between the AD and various risk factors and diseases.

TABLE I
SOME SIGNIFICANT KEGG PATHWAYS THAT ARE RELATED TO THE NERVOUS SYSTEM AND COMMON AMONG THE AD AND OTHER RISK FACTORS AND DISEASES.

KEGG ID	Pathway	Genes in pathway	Risk factor/ disease
hsa04720	Long-term potentiation	GRIA1, PRKCB, GRIN1, CREBBP, PPP3CA	Ag, T2D, HBF, HFD, Ob
hsa05014	Amyotrophic lateral sclerosis (ALS)	GRIA1, NEFL, NEFM, NEFH, PPP3CA	Ag, AC, HBF, Ob, SL
hsa04728	Dopaminergic synapse	KCNJ5, COMT, GNAL, PRKCB, GNB5	AC, T2D, HBF, RM
hsa05031	Amphetamine addiction	GRIA1, PRKCB, GRIN1, PPP3CA	Ag, T2D, HBF, Ob
hsa04662	B cell receptor signaling pathway	PRKCB, CD22, PPP3CA	T2D, HBF, Ob
hsa04713	Circadian entrainment	PRKCB, GRIN1, GNB5	T2D, HBF, RM
hsa04724	Glutamatergic synapse	PRKCB, GRIN1, GNB5	T2D, HBF, RM
hsa04940	Type I T2D mellitus	GAD1, PTPRN2	Ag, HFD, SL
hsa05100	Bacterial invasion of epithelial cells	ITGB1, MET	HBF, HFD, Sm
hsa05140	Leishmaniasis	HLA-DRB4, PRKCB, ITGB1	T2D, HBF, HFD
hsa05146	Amoebiasis	GNAL, PRKCB, LAMB1,	T2D, HBF, Ob
hsa00250	Alanine, aspartate and glutamate metabolism	FOLH1, GAD1	Ag, HFD
hsa04014	Ras signaling pathway	PRKCB, RASAL2, GRIN1, GNB5	T2D, RM
hsa04310	Wnt signaling pathway	PRKCB, PPP3CA, WIF1	HBF, Ob
hsa04360	Axon guidance	ITGB1, MET	HBF, Sm
hsa04370	VEGF signaling pathway	PRKCB, PPP3CA	HBF, Ob
hsa04512	ECM-receptor interaction	ITGB1, LAMB1	HBF, Ob
hsa04514	Cell adhesion molecules (CAMs)	HLA-DRB4, SELE, CD22, ITGB1	T2D, HBF
hsa04520	Adherens junction	CREBBP, MET	HFD, Sm
hsa04530	Tight junction	MYH7B, PRKCB	T2D, HBF
hsa04721	Synaptic vesicle cycle	SNAP25, SNAP25	Ag, Sm
hsa04723	Retrograde endocannabinoid signaling	PRKCB, GNB5	HBF, RM
hsa04725	Cholinergic synapse	PRKCB, GNB5	HBF, RM
hsa04726	Serotonergic synapse	PRKCB, GNB5	HBF, RM
hsa04727	GABAergic synapse	PRKCB, GNB5	HBF, RM
hsa04730	Long-term depression	GRIA1, PRKCB	Ag, HBF
hsa04911	Insulin secretion	PRKCB, SNAP25	HBF, Sm
hsa04933	AGE-RAGE signaling pathway in diabetic complications	PRKCB, SELE	T2D, HBF
hsa04961	Endocrine and other factor-regulated calcium reabsorption	PRKCB, KL	HBF, Ob
hsa05032	Morphine addiction	PRKCB, GNB5	HBF, RM
hsa05033	Nicotine addiction	GRIA1, CHRNA4	Ag, SL

Ag=Ageing, T2D=Type II Diabetes, Ob=Obesity, RM=Red Meat, AC=Alcohol Consumption, SL=Sedentary Lifestyle, Sm=Smoking.

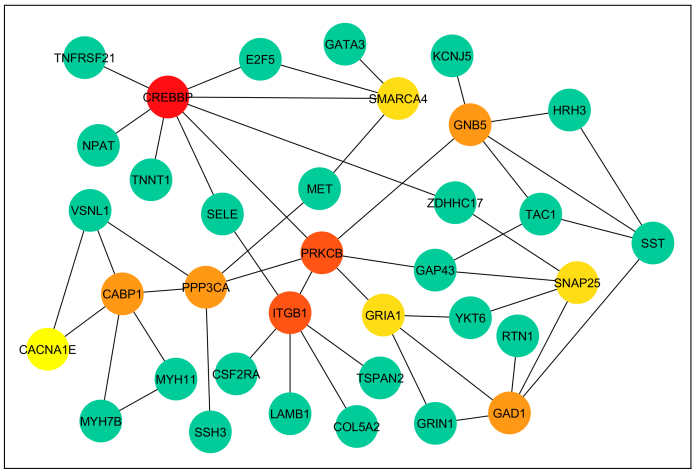

Fig. 4. The simplified PPI network of the commonly dysregulated genes among between AD and other risk factors and diseases. The for 10 most significant hub proteins are marked as red, orange and yellow respectively.

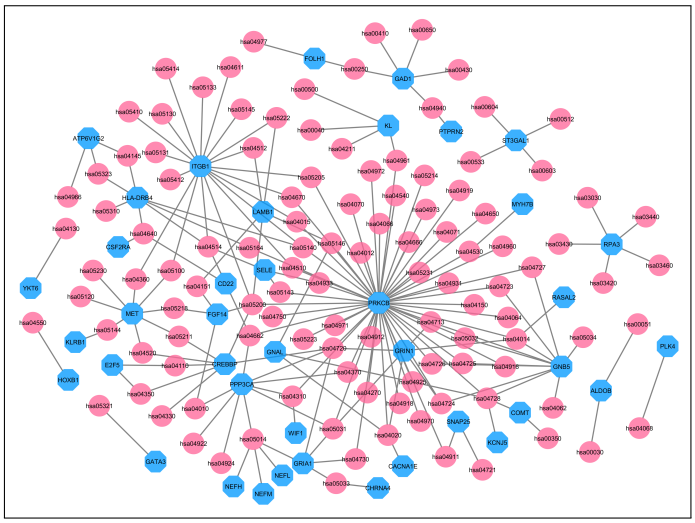

Fig. 5. The gene and pathway association network for all pathways obtained for the commonly dysregulated genes between the AD and other risk factors and diseases. Sky blue-colored octagon-shaped nodes represent genes, and round-shaped pink-colored nodes represent pathway (KEGG Id). A link is placed when a gene belongs to a pathway.

V. CONCLUSIONS

In this study, genomic data is considered to identify the genetic association of various diseasome relationships with the AD. Our findings elicit that the network methods can illustrate disease progression that yields a potential advancement towards having better insight into the origin and development of the AD. Detecting the complex relationship of various risk factors with the disease may disclose novel and useful information for having a better understanding of overall mechanism as well as planning remedy strategy for the AD. Making genomic-based recommendations for accurate disease diagnosis and effective treatment can be boosted by the approaches demonstrated in this study. This enhancement may eventually lead genomic information-based personalized

TABLE II
SIGNIFICANT GO ONTOLOGIES THAT ARE RELATED TO NERVOUS SYSTEM, AND COMMON BETWEEN THE AD AND OTHER RISK FACTORS AND DISEASES.

GO ID	Pathway	Genes in pathway	Risk factors/ disease
GO:0045110	Intermediate filament bundle assembly	NEFL, NEFM, NEFH	Ag, AC, HBF, Ob, SL
GO:0002455; GO:0006909; GO:0006911; GO:0006958; GO:0050851; GO:0050853; GO:0050864; GO:0050871; GO:0051251	Humoral immune response mediated by circulating immunoglobulin; Phagocytosis; Phagocytosis, engulfment; Complement activation, classical pathway; Antigen receptor-mediated signaling pathway; B cell receptor signaling pathway; Regulation of B cell activation; Positive regulation of B cell activation; Positive regulation of lymphocyte activation	IGHG3, IGHM, IGHG1, IGHV4-31, IGHD, IGHA1, IGHA2	AC, T2D, SL
GO:0006836	Neurotransmitter transport	SNAP25	Ag, Sm
GO:0050890	Cognition	CHRNA4, HRH3	SL, Sm
GO:0050885; GO:0060563	Neuromuscular process controlling balance; Neuroepithelial cell differentiation	USH1C	Ag
GO:0046928; GO:0048168; GO:0048935; GO:0051588	Regulation of neurotransmitter secretion; Regulation of neuronal synaptic plasticity; Peripheral nervous system neuron development; Regulation of neurotransmitter transport	MCTP1	T2D
GO:0006968	Cellular defense response	ITGB1	HFD
GO:0007274	Neuromuscular synaptic transmission	CHRNA4	SL
GO:0007422	Peripheral nervous system development	NFASC	Sm

Ag=Ageing, T2D=Type II Diabetes, Ob=Obesity, RM=Red Meat, AC=Alcohol Consumption, SL=Sedentary Lifestyle, Sm=Smoking.

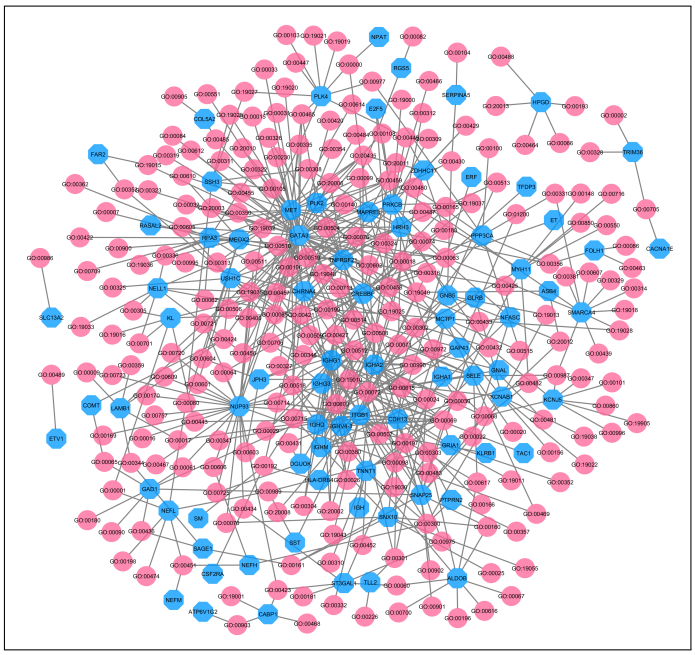

Fig. 6. The gene and gene ontology association network for all the gene ontologies obtained for commonly dysregulated genes between the AD and other risk factors and diseases. Sky blue-colored octagon-shaped nodes represent genes, and round-shaped pink-colored nodes represent gene ontology (GO Term). A link is placed when a gene belongs to an ontology.

TABLE III
GENE-DISEASE ASSOCIATION ANALYSIS OF DIFFERENTIALLY EXPRESSED GENES OF 8 RISK FACTORS AND TYPE II DIABETES WITH AD USING OMIM AND DBGAP DATABASES.

Risk factor/ disease	Genes	Adjusted p-value
Ageing	PLAG1, HMGA2, DNAH11, CR1, VSNL1, DCHS2, F13A1, DISC1, SLC28A1	4.34E-02
Alcohol Consumption	GNAQ, GNAS, ARNT, APBB2, ADCY2, ADCY1, IGF1R, MS4A6A, RTN1, ATXN1, PIEZO2, ST3GAL1	4.72E-01
Type II Diabetes	AKAP13, HNF4A, HMGA2, BUB1, IGF1R, CR1, DIAPH3, TENM4, CADPS, NEDD9, NPAS3	4.66E-01
HBF	PTGIR, THRA, HMGA2, ADCY2, PSEN1, DRD1, BUB1, DBT, PIEZO2	1.18E-01
HFD	AKAP13, CREBBP, HNF4A, HMGA2, DNAH11, COL22A1, PIEZO2, RORA, GFRA2, CD33	5.17E-01
Obesity	RYR2, RBFOX1, VSNL1, NR2F1, HMGA2	1.94E-01
Red Meat	HFE, HMGA2, ADCY2, F13A1	4.41E-01
Sedentary Lifestyle	TFAP2A, HSP90AA1, CREB1, THRA, HFE, APOE, TSHR, RYR2, DIAPH3, DCHS2, DBT, CLU	1.01E-01
Smoking	A2M, OPRD1, SMAD1, ACE, BUB1, CNTNAP2	8.45E-02

medicine for the more precise insight into disease mechanism, and thus an opportunity to determine accurate detection, treatment and remedy for the AD disease.

REFERENCES

[1] B. Duthey, "Background paper 6.11: Alzheimer disease and other dementias," *A Public Health Approach to Innovation*, pp. 1–74, 2013.
[2] A. Serrano-Pozo, M. P. Frosch, E. Masliah, and B. T. Hyman, "Neuropathological alterations in alzheimer disease," *Cold Spring Harbor perspectives in medicine*, vol. 1, no. 1, p. a006189, 2011.
[3] A. Association *et al.*, "2017 alzheimer's disease facts and figures," *Alzheimer's & Dementia*, vol. 13, no. 4, pp. 325–373, 2017.
[4] S. C. Waring and R. N. Rosenberg, "Genome-wide association studies in alzheimer disease," *Archives of neurology*, vol. 65, no. 3, pp. 329–334, 2008.

[5] J. Lindsay, D. Laurin, R. Verreault, R. Hébert, B. Helliwell, G. B. Hill, and I. McDowell, "Risk factors for alzheimers disease: a prospective analysis from the canadian study of health and aging," *American journal of epidemiology*, vol. 156, no. 5, pp. 445–453, 2002.

[6] M. Kivipelto, T. Ngandu, L. Fratiglioni, M. Viitanen, I. Kåreholt, B. Winblad, E.-L. Helkala, J. Tuomilehto, H. Soininen, and A. Nissinen, "Obesity and vascular risk factors at midlife and the risk of dementia and alzheimer disease," *Archives of neurology*, vol. 62, no. 10, pp. 1556–1560, 2005.

[7] R. Mayeux and Y. Stern, "Epidemiology of alzheimer disease," *Cold Spring Harbor perspectives in medicine*, p. a006239, 2012.

[8] J. Janson, T. Laedtke, J. E. Parisi, P. OBrien, R. C. Petersen, and P. C. Butler, "Increased risk of type 2 diabetes in alzheimer disease," *Diabetes*, vol. 53, no. 2, pp. 474–481, 2004.

[9] M. C. Morris, D. A. Evans, J. L. Bienias, C. C. Tangney, D. A. Bennett, N. Aggarwal, J. Schneider, and R. S. Wilson, "Dietary fats and the risk of incident alzheimer disease," *Archives of neurology*, vol. 60, no. 2, pp. 194–200, 2003.

[10] K. J. Anstey, H. A. Mack, and N. Cherbuin, "Alcohol consumption as a risk factor for dementia and cognitive decline: meta-analysis of prospective studies," *The American journal of Geriatric psychiatry*, vol. 17, no. 7, pp. 542–555, 2009.

[11] L. Tilley, K. Morgan, and N. Kalsheker, "Genetic risk factors in alzheimer's disease." *Molecular Pathology*, vol. 51, no. 6, p. 293, 1998.

[12] D. Altshuler, M. J. Daly, and E. S. Lander, "Genetic mapping in human disease," *science*, vol. 322, no. 5903, pp. 881–888, 2008.

[13] A. Rzhetsky, D. Wajngurt, N. Park, and T. Zheng, "Probing genetic overlap among complex human phenotypes," *Proceedings of the National Academy of Sciences*, vol. 104, no. 28, pp. 11 694–11 699, 2007.

[14] D.-S. Lee, J. Park, K. Kay, N. A. Christakis, Z. Oltvai, and A.-L. Barabási, "The implications of human metabolic network topology for disease comorbidity," *Proceedings of the National Academy of Sciences*, 2008.

[15] K.-I. Goh, M. E. Cusick, D. Valle, B. Childs, M. Vidal, and A.-L. Barabási, "The human disease network," *Proceedings of the National Academy of Sciences*, vol. 104, no. 21, pp. 8685–8690, 2007.

[16] I. Feldman, A. Rzhetsky, and D. Vitkup, "Network properties of genes harboring inherited disease mutations," *Proceedings of the National Academy of Sciences*, vol. 105, no. 11, pp. 4323–4328, 2008.

[17] K. Lage, E. O. Karlberg, Z. M. Størling, P. I. Olason, A. G. Pedersen, O. Rigina, A. M. Hinsby, Z. Tümer, F. Pociot, N. Tommerup *et al.*, "A human phenome-interactome network of protein complexes implicated in genetic disorders," *Nature biotechnology*, vol. 25, no. 3, p. 309, 2007.

[18] S. Suthram, J. T. Dudley, A. P. Chiang, R. Chen, T. J. Hastie, and A. J. Butte, "Network-based elucidation of human disease similarities reveals common functional modules enriched for pluripotent drug targets," *PLoS computational biology*, vol. 6, no. 2, p. e1000662, 2010.

[19] M. A. Moni and P. Liò, "Network-based analysis of comorbidities risk during an infection: Sars and hiv case studies," *BMC bioinformatics*, vol. 15, no. 1, p. 333, 2014.

[20] M. A. Moni and P. Lio', "Genetic profiling and comorbidities of zika infection," *The Journal of infectious diseases*, vol. 216, no. 6, pp. 703–712, 2017.

[21] L. Bertram, C. M. Lill, and R. E. Tanzi, "The genetics of alzheimer disease: back to the future," *Neuron*, vol. 68, no. 2, pp. 270–281, 2010.

[22] M. W. Logue, M. Schu, B. N. Vardarajan, J. Buros, R. C. Green, R. C. Go, P. Griffith, T. O. Obisesan, R. Shatz, A. Borenstein *et al.*, "A comprehensive genetic association study of alzheimer disease in african americans," *Archives of neurology*, vol. 68, no. 12, pp. 1569–1579, 2011.

[23] S. Seshadri, A. L. Fitzpatrick, M. A. Ikram, A. L. DeStefano, V. Gudnason, M. Boada, J. C. Bis, A. V. Smith, M. M. Carrasquillo, J. C. Lambert *et al.*, "Genome-wide analysis of genetic loci associated with alzheimer disease," *Jama*, vol. 303, no. 18, pp. 1832–1840, 2010.

[24] E. M. Blalock, J. W. Geddes, K. C. Chen, N. M. Porter, W. R. Markesbery, and P. W. Landfield, "Incipient alzheimer's disease: microarray correlation analyses reveal major transcriptional and tumor suppressor responses," *Proceedings of the National Academy of Sciences*, vol. 101, no. 7, pp. 2173–2178, 2004.

[25] R. E. MacLaren, W. Cui, H. Lu, S. Simard, and K. Cianflone, "Association of adipocyte genes with asp expression: a microarray analysis of subcutaneous and omental adipose tissue in morbidly obese subjects," *BMC medical genomics*, vol. 3, no. 1, p. 3, 2010.

[26] U. Raue, T. A. Trappe, S. T. Estrem, H.-R. Qian, L. M. Helvering, R. C. Smith, and S. Trappe, "Transcriptome signature of resistance exercise adaptations: mixed muscle and fiber type specific profiles in young and old adults," *Journal of applied physiology*, vol. 112, no. 10, pp. 1625–1636, 2012.

[27] S. Radom-Aizik, S. Hayek, I. Shahar, G. Rechavi, N. Kaminski, and I. Ben-Dov, "Effects of aerobic training on gene expression in skeletal muscle of elderly men." *Medicine and science in sports and exercise*, vol. 37, no. 10, pp. 1680–1696, 2005.

[28] S. Kakehi, Y. Tamura, K. Takeno, Y. Sakurai, M. Kawaguchi, T. Watanabe, T. Funayama, F. Sato, S.-i. Ikeda, A. Kanazawa *et al.*, "Increased intramyocellular lipid/impaired insulin sensitivity is associated with altered lipid metabolic genes in muscle of high responders to a high-fat diet," *American Journal of Physiology-Endocrinology and Metabolism*, vol. 310, no. 1, pp. E32–E40, 2015.

[29] H. Misu, T. Takamura, H. Takayama, H. Hayashi, N. Matsuzawa-Nagata, S. Kurita, K. Ishikura, H. Ando, Y. Takeshita, T. Ota *et al.*, "A liver-derived secretory protein, selenoprotein p, causes insulin resistance," *Cell metabolism*, vol. 12, no. 5, pp. 483–495, 2010.

[30] F. Herse, R. Dechend, N. K. Harsem, G. Wallukat, J. Janke, F. Qadri, L. Hering, D. N. Muller, F. C. Luft, and A. C. Staff, "Dysregulation of the circulating and tissue-based renin-angiotensin system in preeclampsia," *Hypertension*, vol. 49, no. 3, pp. 604–611, 2007.

[31] D. G. Hebels, K. M. Sveje, M. C. de Kok, M. H. van Herwijnen, G. G. Kuhnle, L. G. Engels, C. B. Vleugels-Simon, W. G. Mares, M. Pierik, A. A. Masclee *et al.*, "N-nitroso compound exposure-associated transcriptomic profiles are indicative of an increased risk for colorectal cancer," *Cancer letters*, vol. 309, no. 1, pp. 1–10, 2011.

[32] J. N. McClintick, A. I. Brooks, L. Deng, L. Liang, J. C. Wang, M. Kapoor, X. Xuei, T. Foroud, J. A. Tischfield, and H. J. Edenberg, "Ethanol treatment of lymphoblastoid cell lines from alcoholics and non-alcoholics causes many subtle changes in gene expression," *Alcohol*, vol. 48, no. 6, pp. 603–610, 2014.

[33] P. Büttner, S. Mosig, and H. Funke, "Gene expression profiles of t lymphocytes are sensitive to the influence of heavy smoking: a pilot study," *Immunogenetics*, vol. 59, no. 1, pp. 37–43, 2007.

[34] M. E. Smoot, K. Ono, J. Ruscheinski, P.-L. Wang, and T. Ideker, "Cytoscape 2.8: new features for data integration and network visualization," *Bioinformatics*, vol. 27, no. 3, pp. 431–432, 2010.

[35] D. Szklarczyk, J. H. Morris, H. Cook, M. Kuhn, S. Wyder, M. Simonovic, A. Santos, N. T. Doncheva, A. Roth, P. Bork *et al.*, "The string database in 2017: quality-controlled protein–protein association networks, made broadly accessible," *Nucleic acids research*, p. gkw937, 2016.

[36] S.-H. Chen, C.-H. Chin, H.-H. Wu, C.-W. Ho, M.-T. Ko, and C.-Y. Lin, "cyto-hubba: A cytoscape plug-in for hub object analysis in network biology," in *20th International Conference on Genome Informatics*, 2009.

[37] B. Calimlioglu, K. Karagoz, T. Sevimoglu, E. Kilic, E. Gov, and K. Y. Arga, "Tissue-specific molecular biomarker signatures of type 2 diabetes: an integrative analysis of transcriptomics and protein–protein interaction data," *Omics: a journal of integrative biology*, vol. 19, no. 9, pp. 563–573, 2015.

[38] M. V. Kuleshov, M. R. Jones, A. D. Rouillard, N. F. Fernandez, Q. Duan, Z. Wang, S. Koplev, S. L. Jenkins, K. M. Jagodnik, A. Lachmann *et al.*, "Enrichr: a comprehensive gene set enrichment analysis web server 2016 update," *Nucleic acids research*, vol. 44, no. W1, pp. W90–W97, 2016.

[39] C. Rouaux, J.-P. Loeffler, and A.-L. Boutillier, "Targeting creb-binding protein (cbp) loss of function as a therapeutic strategy in neurological disorders," *Biochemical pharmacology*, vol. 68, no. 6, pp. 1157–1164, 2004.