Informatics in Medicine Unlocked 30 (2022) 100888

Contents lists available at ScienceDirect

Informatics in Medicine Unlocked

journal homepage: www.elsevier.com/locate/imu

Identifying molecular signatures and pathways shared between Alzheimer's and Huntington's disorders: A bioinformatics and systems biology approach

Nosin Ibna Mahbub [a,1], Md. Imran Hasan [b,1], Md Habibur Rahman [b,**], Feroza Naznin [c], Md Zahidul Islam [a], Mohammad Ali Moni [d,e,*]

[a] Dept. of Information and Communication Technology, Islamic University, Kushtia-7003, Bangladesh
[b] Dept. of Computer Science and Engineering, Islamic University, Kushtia-7003, Bangladesh
[c] Dept. of Computer Science and Engineering, Green University of Bangladesh, Dhaka, Bangladesh
[d] Dept. of Computer Science and Engineering, Pabna University of Science and Technology, Bangladesh
[e] Artificial Intelligence & Digital Health Data Science, School of Health and Rehabilitation Sciences, Faculty of Health and Behavioural Sciences, The University of Queensland, St Lucia, QLD 4072, Australia

ARTICLE INFO

Keywords:
Alzheimer's disease
Huntington's disease
Biomarker signatures
Differentially expressed genes
Protein-protein interaction
Protein-drug interactions

ABSTRACT

Alzheimer's disease and Huntington's disease are considered to be the most lethal illnesses that result in common human disorders. Alzheimer's disease (AD) is a progressive neurological illness distinguished by age-related dementia, mental abnormalities, and poor memory, among other symptoms. However, Huntington's disease (HD) is influenced by genetics as well as a generalized dysfunction of the motor system. Despite the fact that many similar genetic elements have been found in the literature as being interrelated between these two diseases, it is still unclear how people acquire infected with these two neurological disorders. Detecting biomarkers for Alzheimer's and Huntington's disease in brain tissue might help in drug development and treatment. The purpose of this research was to find brain cell transcripts that show levels of gene expression linked to the progression of Alzheimer's and Huntington's disease. A bioinformatics pipeline was used to study one RNA-Seq transcriptomic dataset and one microarray dataset, and 24 significant differentially expressed genes (DEGs) were discovered that were shared by two brain cell datasets. We uncovered disease-gene association networks and signaling pathways, as well as gene ontology (GO) investigations and hub protein identification, to determine the roles of these DEGs. The discovery of significant gene ontologies and molecular pathways increased our understanding of the pathophysiology of these two disorders, and the hub proteins B2M, HLA-A, HLA-E, HLA-B, HLA-C, HLA-F, CANX, HLA-DQA1, HLA-DRA, and HLA-DRB1 might be exploited to design therapeutic interventions. In neurological disorder subjects, we uncovered efficient hypothetical linkages between pathogenic processes in brain cells, implying that brain cells may be exploited to detect and monitor illness origin and development, as well as design pharmacological therapies.

1. Introduction

Alzheimer's disease (AD) is a chronic neurological illness that gradually robs patients of cognitive function and eventually leads to death [1]. AD is a neurological condition that is growing more common in the world's aging populations [2,3]. Age is the most notable hazard indicator for AD [4]. The condition is never visible in children, even when they have disease-causing mutations that cause to be over expressed from birth [5]. Because most people's memory degrades slightly with age, the distinction between typical age-related forgetfulness and the initial signs of AD can be blurry [2]. Its symptoms involved memory loss, linguistic difficulties, and erratic behavior. He investigated that affected brain after her death and identified many abnormal clumps (now recognized as amyloid plaques) and tangled bundles of fibers (now called neurofibrillary, or tau, tangles). These plaques and tangles in the brain are mostly considered to be among of Alzheimer's disease's most visible features [6]. Alzheimer's disease is estimated to affect 2.3 million people in the United States (range, 1.09–4.8 million)

* Corresponding author. Dept. of Computer Science and Engineering, Green University of Bangladesh, Dhaka, Bangladesh.
** Corresponding author.
E-mail addresses: habib@iu.ac.bd (M.H. Rahman), moni@pust.ac.bd, m.moni@uq.edu.au (M.A. Moni).
[1] Nosin Ibna Mahbub & Md. Imran Hasan contributed equally and hold joint first authorship of this work.

https://doi.org/10.1016/j.imu.2022.100888
Received 22 October 2021; Received in revised form 22 January 2022; Accepted 17 February 2022
Available online 31 March 2022

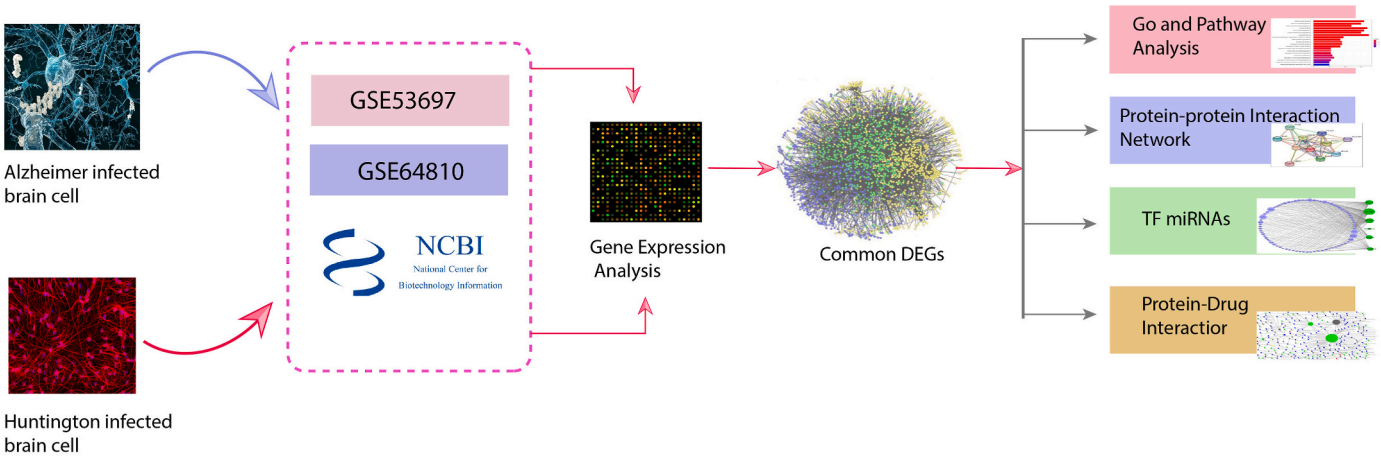

Fig. 1. Proposed methodology and the workflow of our work.

Table 1
A summary of the datasets included in this study, together with their geo-characteristics and quantitative measures.

Disease name	GEO accession	Total DEGs count	Up regulated DEGs count	Down regulated DEGs count
Alzheimer's Disease (AD)	GSE53697	308	100	208
Huntington's Disease (HD)	GSE64810	1655	894	761

[7]. The prevalence of AD increases every five years after age 60, rising from 1% for individuals between the ages of 60 and 64 to up to 40% for those aged 85 and above [7].

Huntington's disease (HD) is indeed a rare hereditary neurological disease characterized by uncontrolled excessive neurological movements and cognitive and emotional deficits [8,9]. Symptoms of HD often appear in middle age after infected people have had children, although the condition may appear at any point between infancy and senescence [10]. Huntington's, the disease-causing defective protein, has an enlarged CAG repetition, resulting in a polyglutamine sequence at the N-terminus that may be any length. There's evidence to support the theory that this extra segment provides a harmful functional benefit [10].

Huntington's disease (HD) and Alzheimer's disease (AD) are two neurological illnesses that overlap clinical characteristics associated with specific brain impairment. Both disorders are caused by misfolding and deposition of particular proteins that associate with mitochondria and disrupt with endoplasmic reticulum (ER)/mitochondria-contact sites [11]. The accumulation of data suggests that mitochondrial Ca2+ homeostasis dysfunction underpins the vulnerability to specific neuronal death reported in HD and AD [11]. Moss [12] proposed that transcription is interrupted in peripheral cells in HD via processes that are similar to those seen in the brain. The convergence of immunological upregulation in HD and AD implies a common pathogenic process incorporating macrophage phagocytosis and microglial synaptic pruning, and thus opens the possibility of addressing both diseases with similar treatment methods [12].

Microarray-based gene expression analysis is the most widely used and successful high-throughput technology for studying complex disease etiology. Human ovarian cancer (OC) gene expression profiling studies, on the other hand, are extremely rare. In this study, we attempted to investigate the differentially expressed genes (DEGs), gene network, pathways, and protein interactions that are specific to HD [13]. To interpret the biological significance of these changes in gene expression, we used an integrated bioinformatic analysis that expanded on traditional microarray analysis methods, such as Gene Ontology (GO) and pathway analysis, to build interaction networks that could identify novel prognostic markers and therapeutic targets [14]. Denggang Fu [15] claims Computational biology was used to study the potential molecular processes and tumor immune landscape of these IRGs. An examination of tumor-infiltrating lymphocytes and immune checkpoint molecules indicated a different immunological landscape in the high-risk and low-risk groups. In several articles, bioinformatics approaches have been applied to identify potential molecular biomarkers, pathway analysis, gene ontology, and drug targets [13–17].

According to JiahuiWan [16] because the most well-known molecular mechanism of lncRNA is to operate as a microRNA "sponge" that regulates the activity of mRNAs, lncRNAs are also known as competing endogenous RNAs (ceRNAs). S. Udhaya Kumar [17] focused on finding dysregulated molecular pathways and key genes that are differentially regulated in Familial hypercholesterolemia (FH), as well as possible genetic variables and putative underlying processes that raise the risk of atherosclerosis in FH individuals.

Although there is strong evidence that there are pathological and clinically significant connections between AD and HD but the relationship has not been thoroughly explored. As a result, the nature of these linkages is poorly known. Due to the intricacy of AD and HD's etiology, its biological basis, as well as the molecular systems that enable this association, are still unknown. Furthermore, bioinformatics research that investigated the association between AD and HD is still insufficient.

Two datasets were used in this research to discover the biological association between AD and HD. The datasets were obtained from the Gene Expression Omnibus (GEO) database, with reference codes GSE53697 and GSE64810 for AD and HD, respectively. At first, differentially expressed genes (DEGs) for datasets were identified, and then common DEGs genes for two diseases were found. In this case, the common DEGs serve as the key experimental genes for the whole research. Further experiment and analysis were carried out using these common DEGs, including:

- Ontological and functional enrichment analysis to dermine shared ontologies and pathways.
- The network of protein-protein interactions (PPIs) formed by the mutual DEGs.
- The transcriptional components of frequent DEGs have been identified.
- A network of protein-drug interactions is being developed to find prospective medications.
- The gene-disease association network will be used to uncover other diseases that share similar DEGs.

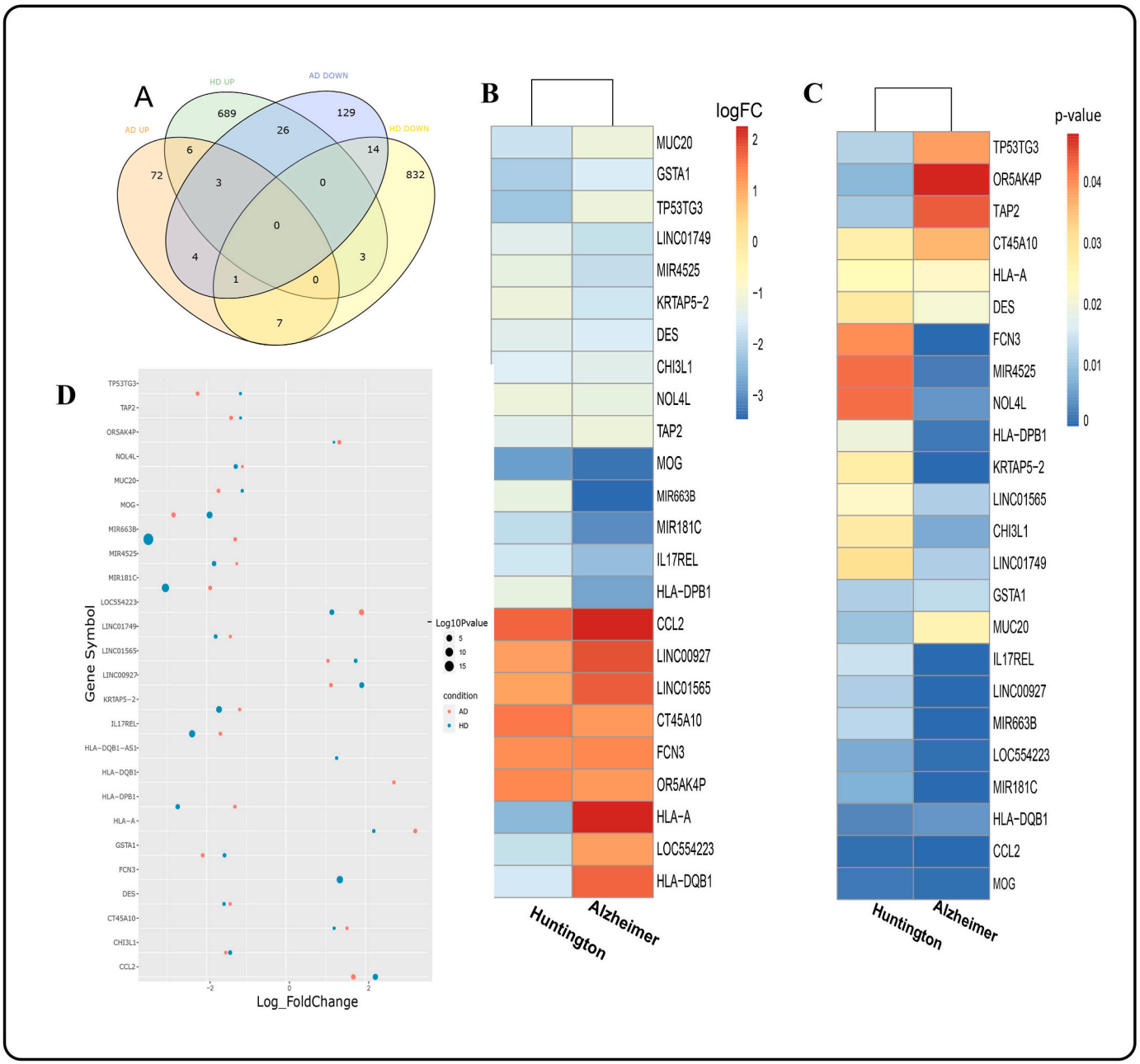

Fig. 2. Comparison of RNA-Seq analyses of Alzheimer's disease and Huntington's disease. **(A)** The Venn diagram depicts the number of shared important genes associated with AD and HD. **(B)** Heat map illustrating the log fold change for the genes shared by AD and HD. **(C)** Heat map illustrating the p-values for the genes shared by AD and HD. **(D)** The bubble figure depicts the joint log fold changes and p-values for the common genes shared by AD and HD.

So, the main objective of this study was to construct a workflow depending on bioinformatics methods for detecting potential associations between AD and HD. Determining the nature of these linkages may provide light on the molecular processes behind these illnesses and may ultimately contribute to the discovery of possible treatments that may result in the creation of disease-modifying drugs. Fig. 1 illustrates the sequential workflow of our research.

2. Materials and procedures

2.1. Datasets utilized in this research

The National Center for Biotechnology Information's (NCBI) GEO database (https://www.ncbi.nlm.nih.gov/geo/) [18] was used in this study for assessing the shared genomic interrelationships between AD and HD. We used an RNA-seq dataset for Alzheimer's disease and a microarray dataset for Huntington's disease. The Alzheimer dataset was (GEO accession ID: GSE53697) human brain tissue comprising eight advanced Alzheimer's disease brain instances and nine normal samples which was processed using the high-throughput sequencing functionality known as Illumina HiSeq (Homo sapiens) [19]. The Huntington's dataset (GEO accession ID: GSE64810) was compiled from human post-mortem BA9 brain tissue, which included 49 samples from neurologically normal individuals and 20 samples from Huntington's disease patients. We used a common platform GEO2R tool [20] for analysis the Huntington's dataset. The Dataset was contributed by Labadorf et al. [21]. Table 1 contains the summarized information of the datasets.

Table 2
Each common gene is listed with its p-value and logFC value.

Gene_symbol	logFC for AD	p-value for AD	logFC for HD	p-value for HD
HLA-A	3.261	5.93E-03	2.215	2.27E-02
LOC554223	1.906	1.40E-04	1.158	6.17E-04
CCL2	1.701	9.37E-04	2.255	1.12E-04
CT45A10	1.539	2.69E-02	1.215	3.61E-02
FCN3	1.331	3.99E-02	1.359	4.34E-07
LINC00927	1.136	1.18E-02	1.911	3.15E-04
LINC01565	1.06	2.21E-02	1.758	1.19E-02
HLA-DQB1	2.723	1.93E-02	1.279	2.78E-02
OR5AK4P	1.346	8.44E-03	1.21	4.79E-02
IL17REL	−1.654	1.47E-02	−2.366	1.35E-06
MIR4525	−1.241	4.23E-02	−1.813	2.10E-03
DES	−1.407	2.85E-02	−1.561	2.04E-02
NOL4L	−1.099	4.25E-02	−1.266	4.55E-03
HLA-DPB1	−1.287	2.01E-02	−2.73	1.89E-03
KRTAP5-2	−1.168	2.77E-02	−1.687	7.44E-06
TAP2	−1.383	1.10E-02	−1.145	4.39E-02
GSTA1	−2.099	1.17E-02	−1.55	1.30E-02
MUC20	−1.702	9.72E-03	−1.102	2.54E-02
MOG	−2.834	1.75E-03	−1.926	5.97E-06
CHI3L1	−1.519	2.82E-02	−1.407	6.71E-03
MIR663B	−1.281	1.29E-02	−3.47	5.95E-20
MIR181C	−1.912	7.51E-03	−3.039	1.05E-09
TP53TG3	−2.228	1.20E-02	−1.149	3.85E-02
LINC01749	−1.403	3.08E-02	−1.773	1.16E-02

2.2. Identification of DEGs and mutual DEGs between AD and HD

The purpose of differential expression analysis is to explore out which genes are expressed at various levels under different conditions [22]. These genes can provide biological insight into the processes that are impacted by the state of interest [23]. The datasets were processed in R language (version 3.6.1) and Bioconductor platforms to find DEGs in AD and HD based on their linked controls. Firstly, we normalized the gene expression data employing the log2 transform and statistical techniques. To control rate of false discovery we used "Limma" package from R programming language with Benjamini-Hochberg correction [24]. The important DEGs were determined using P-value less than 0.05 and a $|\log FC| > 1$. The common DEGs of GSE53697 and GSE64810 were obtained using the Jvenn online VENN analysis tool [25].

2.3. Assessment of gene ontologies and pathway enrichment

Gene set enrichment evaluation is a crucial experimental endeavor that attempts to identify basic biological observation such as biological processes or chromosome locations correlated with various interconnected diseases [26]. Gene ontology and pathway enrichment evaluations were used to deduce the relevant biological concepts and signaling pathways underlying frequent DEGs. The research utilized EnrichR (https://amp.pharm.mssm.edu/Enrichr), a commonly used online platform for gene set enrichment [27]. The three types of gene ontology (GO) and functional process are biological process, cellular component, and molecular function. To find common pathways between AD and HD, we used the KEGG (Kyoto Encyclopedia of Genes and Genomes), WikiPathways, Reactome, and BioCarta databases as sources of pathway annotations. For all analyses, a significant margin was estimated to be a p-value < 0.05.

2.4. Determining transcription factors and miRNAs that interact with mutual DEGs

In order to control gene transcription, transcription factors attach to specific DNA patterns. Transcription factor attachment regions are shorter DNA patterns (5–20 bp in length) that are selectively associated by one or several transcription factors [28]. Identifying transcription factor binding locations and predicting their roles remain tough computational biology tasks.

We utilized the NetworkAnalyst tool for identifying topologically plausible TFs binding to shared DEGs in the JASPAR database. JASPAR is the freely accessible database of TF profiles from multiple species across six taxonomic groups [29]. The associations of miRNAs with their specific genes were investigated in order to detect miRNAs that try to attach to a gene expression in order to inhibit protein production [30]. The primary databases for experimentally validated miRNA-target associations are Tarbase [31] and mirTarbase [32]. Using Network Analyst's topological analysis, we discovered important miRNAs from Tarbase and mirTarbase.

2.5. Network assessment of protein–protein interactions (PPIs)

In all organisms, PPIs are critical for cellular functions and biological processes. The exploration of protein interactions will contribute to a deeper understanding of infection pathways, as well as the identification of multiple medication drugs and treatment optimization [32]. The PPI network of proteins generated by mutual DEGs was constructed using the STRING Protein-Protein Interaction database (version 11.0) (https://string-db.org/) [33] to reflect how our specified DEGs, as well as proteins, communicate physically and functionally with each other. The annotation of protein interactions in STRING (https://string-db.org/) varies according to levels of medium confidence [33]. We set the lowest score confidence criterion to produce the PPI network sharply due to the small number of common DEGs. The Network Analyst web resource was used to conduct network analysis [34].

2.6. Assessment of protein-drug interactions

DrugBank (v5.0) is a fascinating internet-based database of comparative drug records. Simultaneously, it offers information on the impact of drugs on protein expression [35]. We utilized NetworkAnalyst to conduct protein-drug interactions in order to find possible interactions between our common DEGs and medicines in the DrugBank dataset [34].

2.7. Gene–disease association assessment

DisGeNET is a standardized database for gene–disease association that integrates relationships from multiple sources involving different biomedical aspects of illnesses. It highlights the growing understanding of human genetic disorders [36]. We have used network analyst to examine the gene-disease interaction in order to identify diseases and chronic complications correlated with common DEGs [34].

3. Result

3.1. Detection of DEGs and mutual DEGs between AD and HD

To evaluate the relationship between AD and HD, we used the NCBI's human RNA-seq dataset for AD and microarray data for HD. The investigations on the RNA-seq and microarray datasets were conducted using two packages from R programming language called the DESeq2 and limma, with the Benjamin-Hochberg false discovery rate. We examined significant DEGs depending on p-values lower than 0.05 and $|\log FC|$ higher than 1. In the AD dataset, we identified 308 significant DEGs, with 100 and 208 DEGs being substantially up-regulated and down-regulated, respectively. In HD, we obtained 1655 significant DEGs in the similar manner that 894 DEGs were up-regulated and 761 DEGs were down-regulated (DEGs along with the p-value and logFC values for both the datasets added as supplementary files). We utilized the Jvenn utility to do a cross-comparison in order to find shared DEGs between AD and HD. As a result, we discovered that AD and HD share 24 common DEGs. Fig. 2A depicts the overall cross-comparison between two databases in order to obtain common DEGs between AD and HD. A heat map

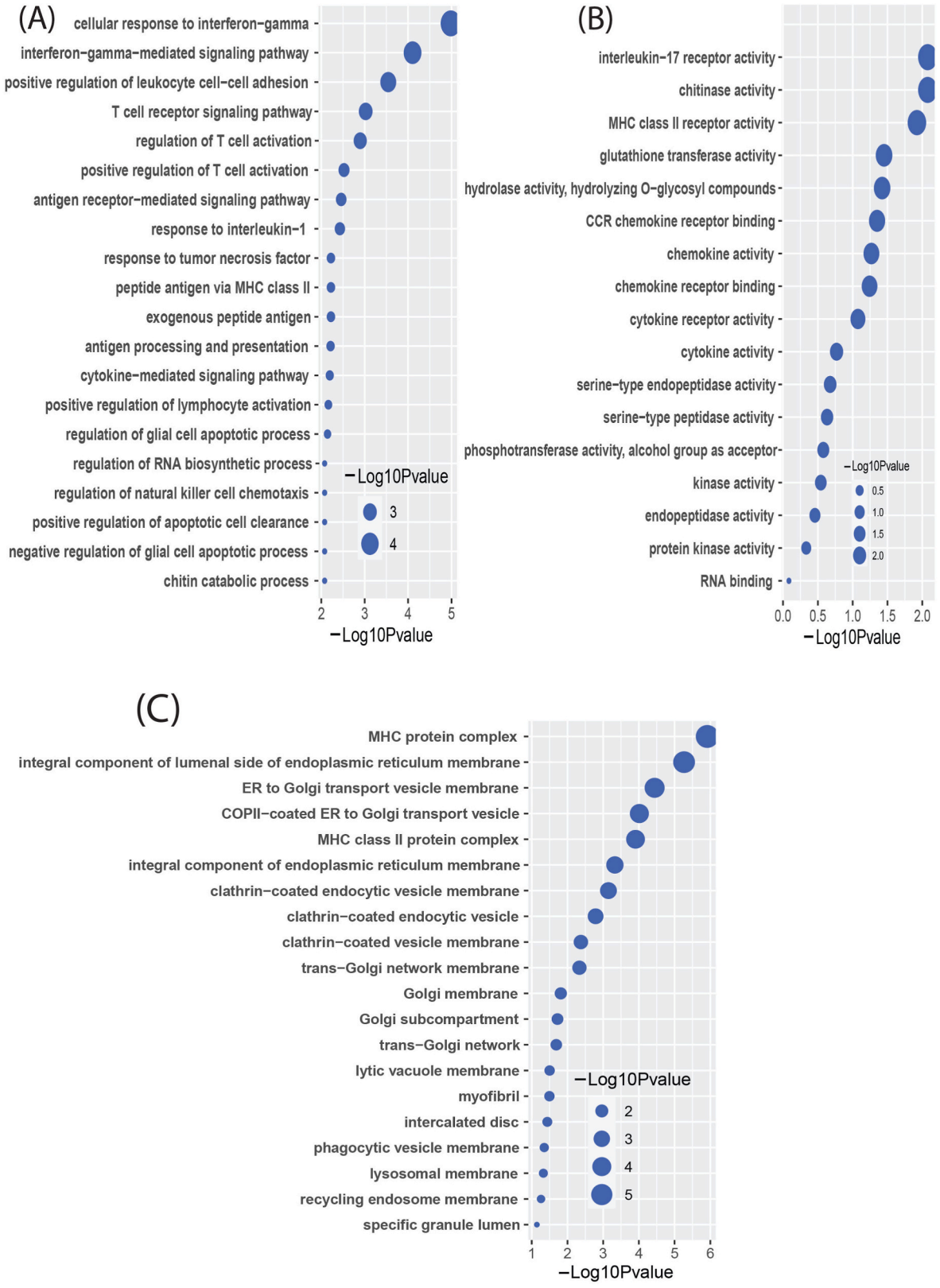

Fig. 3. The bubble plot of ontological analysis of shared DEGs between AD and HD performed by the Enricher online tool: here, **(A)** biological processes, **(B)** molecular function, and **(C)** cellular component.

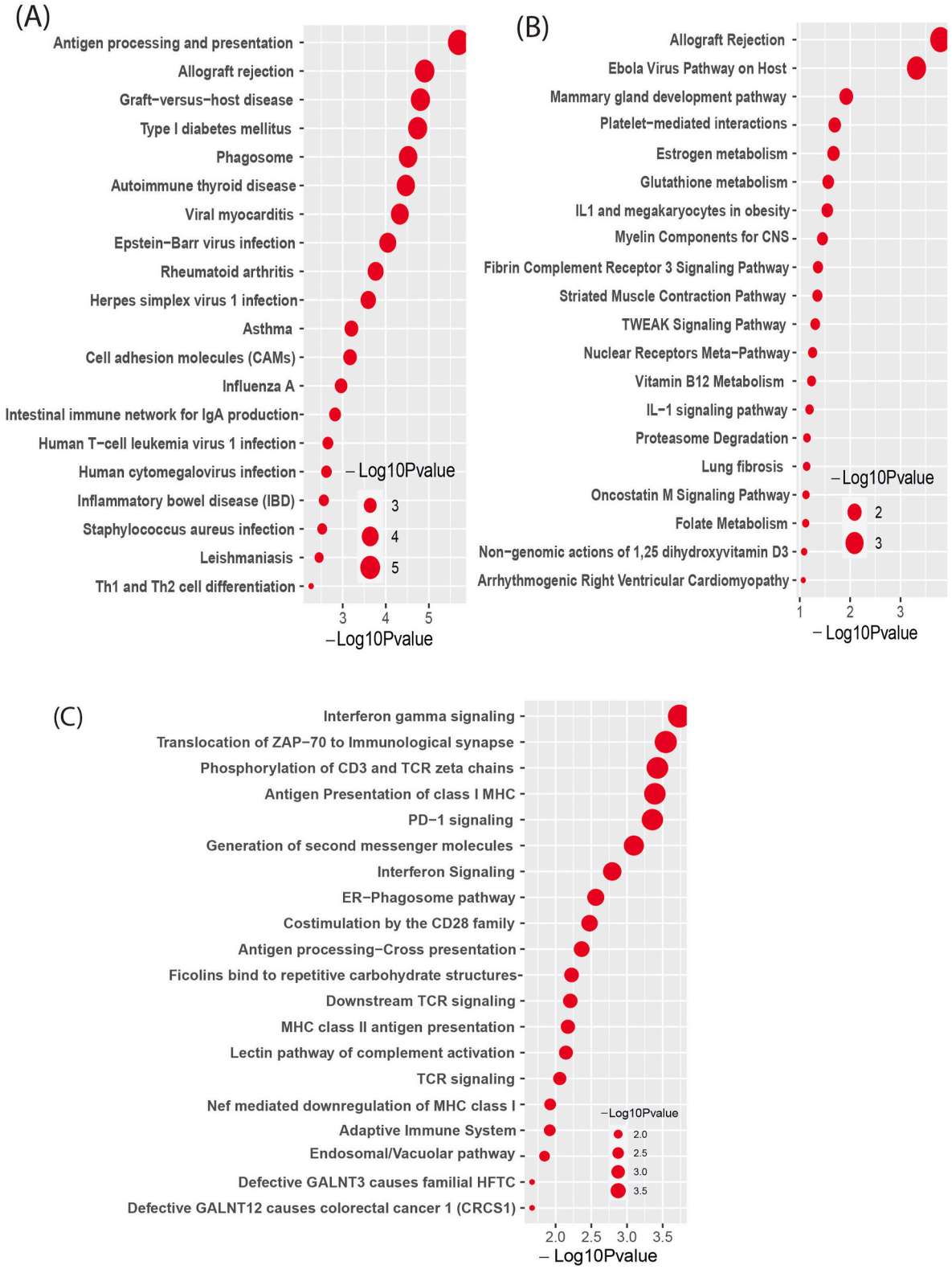

Fig. 4. The bubble plot of pathway enrichment analysis of shared DEGs between AD and HD performed by the Enricher online tool: here, **(A)** KEGG pathway**, (B)** wikipathway, **(C)** reactome pathway.

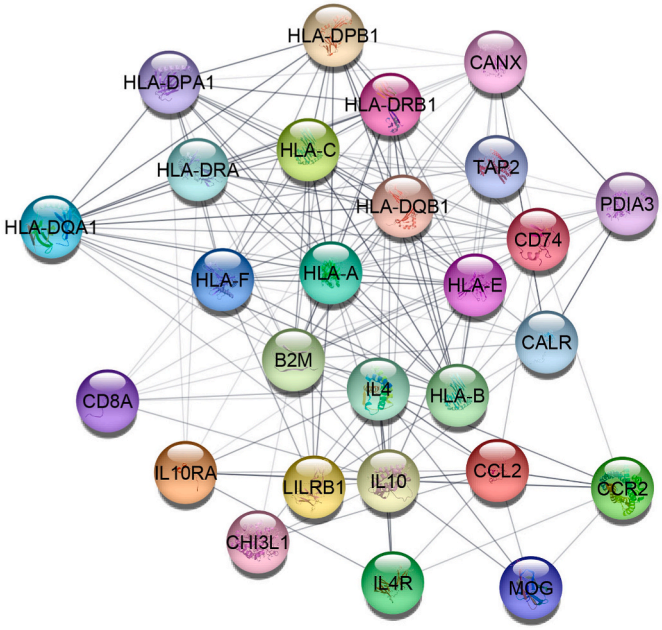

Fig. 5. PPI network of DEGs shared by AD and HD. The circle nodes in the diagram represent DEGs, and the edges represent node interactions. String was used to create the PPI network, which was then visualized in Cytoscape.

and a bubble plot Fig. 2B–D depict the distinct transcriptional signature caused by AD and HD. Two genes (CCL2 and MOG) are often dysregulated in AD and HD. These findings indicate that, while AD and HD have similar transcriptomic profiles, the CCL2 and MOG genes are shared by both diseases. In Table 2 shows, each common gene is listed with its p-value and logFC value.

3.2. Gene ontology and pathway enrichment assessment

Enrichr was used to conduct gene ontology and pathway enrichment analysis to determine the clinical function and enriched pathways illustrated in this investigation that associated DEGs. Gene ontology takes into account gene functions and their components in order to incorporate complete quantifiable knowledge assets. Additionally, ontology and annotation are designed to facilitate the execution of a complicated biological structure model, which is often employed in biomedical activities [37]. The gene ontology study was performed in 3 areas (biological process, cellular component, and molecular function), using annotations derived from the GO database. The bubble graph in Fig. 3(A-C) displays the overall ontology assessment for each group. The main significant ontology was cellular response to interferon−gamma, interferon−gamma−mediated signaling pathway, positive regulation of leukocyte cell−cell adhesion, T cell receptor signaling pathway, interleukin−17 receptor activity, chitinase activity, MHC class II receptor activity, MHC protein complex, integral component of lumenal side of endoplasmic reticulum membrane, COPII−coated ER to Golgi transport vesicle, MHC class II protein complex etc.

Pathways assessment demonstrates how the organism responds to the inherent changes. It acts as a model technique for demonstrating how different diseases interact through fundamental molecular or biological processes [38]. Four worldwide databases, KEGG, WikiPathways, Reactome, and BioCarta, were used to extract the most affected pathways of common DEGs between AD and HD. Fig. 4 depicted the

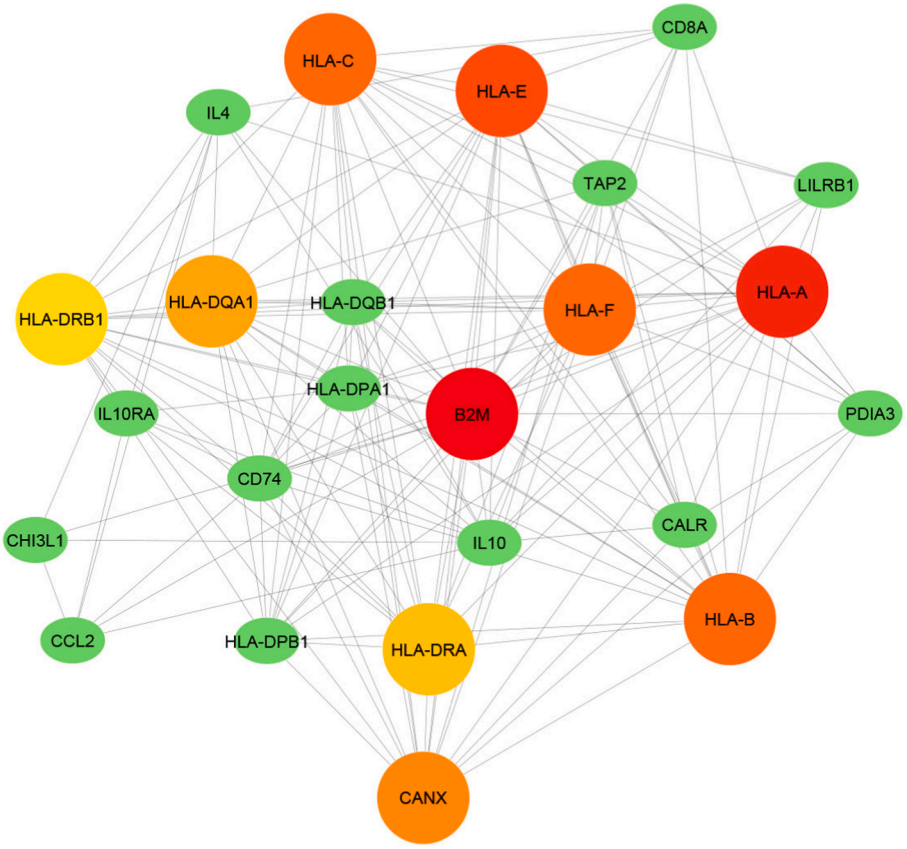

Fig. 6. Identification of hub genes from the PPI network utilizing the Cytohubba plugin in Cytosacpe. To extract hub genes, the new MCC protocol of the Cytohubba plugin was used. The largest nodes in this diagram represent the top ten hub genes and their intermolecular interactions. The network is made up of 24 nodes and 159 edges.

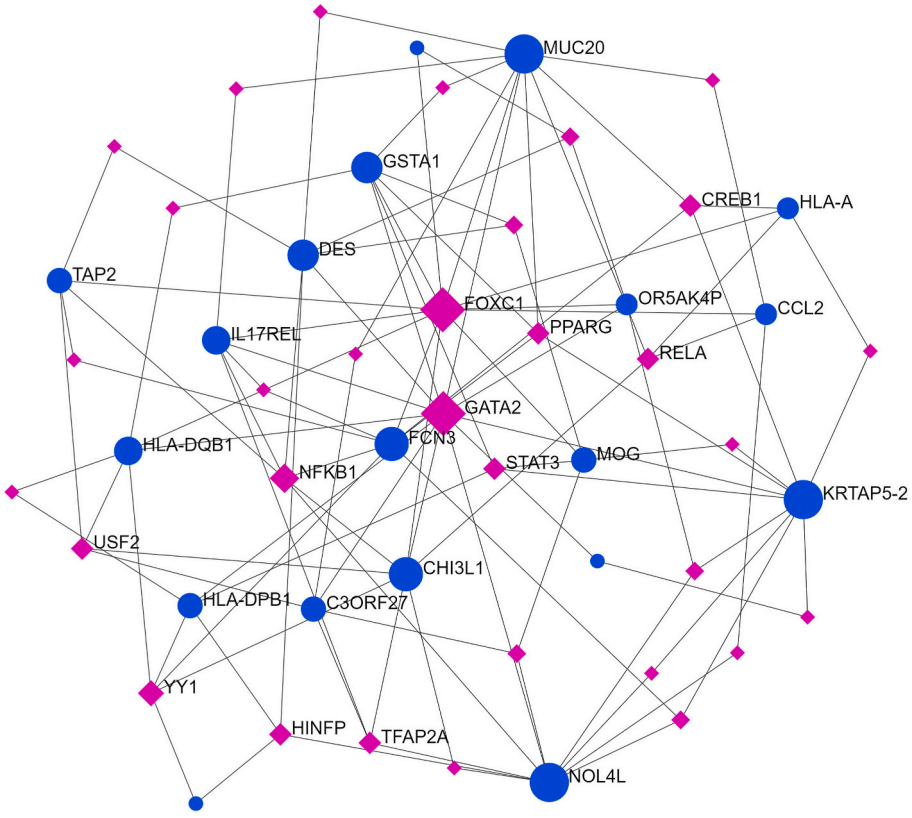

Fig. 7. DEGs-TFs Common DEG interaction network created via Network Analyst. The nodes identified by circles are TFs, while the nodes represented by diamonds associate with TFs.

pathway enrichment analysis in bar graphs. Here, Antigen processing and presentation, Allograft rejection, Graft−versus−host disease, Ebola Virus Pathway on Host, Mammary gland development pathway, Interferon gamma signaling, Translocation of ZAP−70 to Immunological synapse, Phosphorylation of CD3 and TCR zeta chains were the significant pathway.

3.3. Hub protein identification

In order to anticipate typical DEG connections and attachment pathways, we examined the STRING PPI network and displayed it in Cytoscape. The PPI network of frequent DEGs shown in Fig. 5 contains 27 nodes and 173 edges.

At the same time, several interconnected nodes in a PPI network are recognized as hub genes. The leading 10 DEGs were identified as the most influential genes based on a study of the PPI network utilizing the Cytohubba plugin in Cytoscape. The hub genes are B2M, HLA-A, HLA-E, HLA-B, HLA-C, HLA-F, CANX, HLA-DQA1, HLA-DRA, and HLA-DRB1. This hub genes can be important biomarkers, leading to novel therapeutic methods for diseases under investigation. Since hub genes have potential, we built a submodule network (Fig. 6) with the help of the Cytohubba plugin to better understand their close communication and proximity.

3.4. Identifying transcription factors and miRNAs that interact with mutual DEGs

A network-based method was applied to decipher the regulatory TFs, and miRNAs of hub proteins, and the DEGs-TFs and DEGs-miRNA linkages networks were examined to identify transcriptional and post-transcriptional regulatory fingerprints of similar DEGs. The associations involving DEGs and TFs are illustrating in Fig. 7. The correlations between DEGs and miRNAs also appear in Fig. 8. Five transcription factors, MUC20, KRTAP5-2, NOL4L, GSTA1, and CCL2, and ten microRNAs, mir-10b-5p, mir-941, mir-107, mir-330-3p, mir-26a-5p, mir-34a-5p, mir-26b-5p, mir-124-3p, mir-101-3p, and mir-148-3p, were obtained from both interaction networks.

3.5. Identification of candidate drugs

The objective was to find candidate drugs that could potentially affect AD and HD while also investigating the protein-drug interaction. Analysis of protein-drug interactions is essential to understand the characteristics necessary to sensitive receptors [39]. The interaction analysis between the protein and drug revealed the drug's interaction with a hub protein. Fig. 9 shows two drug molecules, called Pyroglutamic Acid and Beta-D-Glucose associate with the hub proteins of KRTAP5-2.

3.6. Identification of gene-disease association

The assumption that various disorders can be associated or linked with each other is that they generally have one or more common genes [40]. Disorder-specific therapeutic interface techniques attempt uncovering the connection between genes and disorders. According to Network Analyst's study of the gene-disease relationship, Schizophrenia, Chemical and Drug Induced Liver Injury, Asthma, Multiple Sclerosis, Celiac Disease, and Brain Ischeia disorders are the most

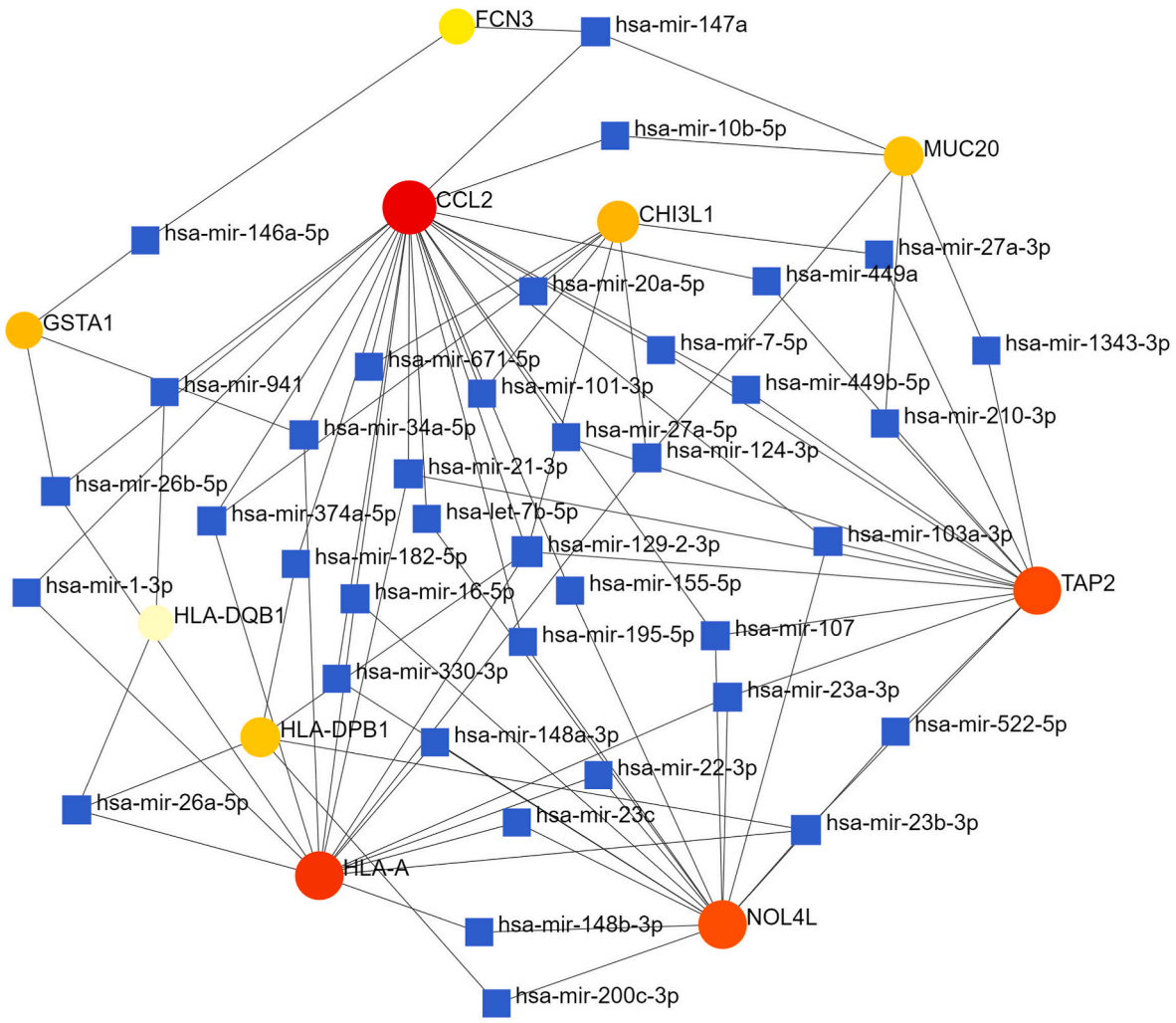

Fig. 8. DEGs-miRNA Common DEG interaction network created via NetworkAnalyst. Circular nodes symbolize DEGs, and square circle structure reflects miRNAs.

synchronized to our identified hub genes. Fig. 10 depicts the connection involving genes and disease.

4. Discussion

We explored gene activity data from Alzheimer's and Parkinson's disease patients in this experiment. We utilized bioinformatics pipelines to describe genes that are dysregulated in both illnesses and may serve as potential treatment candidates or diagnostic biomarkers. Microarray datasets for HD and RNA-seq datasets for AD were employed to examine candidate biomarker genes. The statistical study of the AD and HD transcriptomics showed 24 DEGs with identical expression variations in the two diseases. Gene Ontology (GO) and pathway evaluation were utilized to obtain understanding into the biological significance of these shared genes in the pathogenesis of AD and HD. The Gene Ontology (GO) model is a comprehensive theoretical framework in the field of gene expression that specifies gene activities and interactions. It progresses incrementally via the collection of scientific knowledge regarding gene functions and regulation, which is contingent on various ontological categories and language connections across classes [41]. Enrichr was utilized for gene ontology exploration of common genes across three divisions (biological, cellular, molecular) and an ontological annotation source was used for the GO database. According to GO [42], biological mechanisms are instances of molecular interactions.

The cell part is the cellular form in which the gene controls its activity, and the molecular notion refers to molecular activities. The pathway analysis is a new strategy that explores and reveals how biologically or molecularly complicated disorders are linked. The pathway is the optimal way to achieve the responses of an organism that are caused by internal changes [43]. Gene ontology and the pathway study revealed many mechanisms involved with neurodegenerative diseases. However, this integrative study has succeeded in identifying additional major proteins or hub proteins shared between these two illnesses. It suggests other research paths, such as possible preventative treatments for novel uses.

PPIs experiments are often performed to reach essential illness-related signaling molecules and pathways which might amplify disease facets [44]. Therefore, we conduct a PPI investigation to determine critical hub proteins. B2M, HLA-A, HLA-E, HLA-B, HLA-C, HLA-F, CANX, HLA-DQA1, HLA-DRA, and HLA-DRB1 were identified as 10 hub proteins. In this study, we identified the respective proteins encoded between AD and HD by common DEGs.

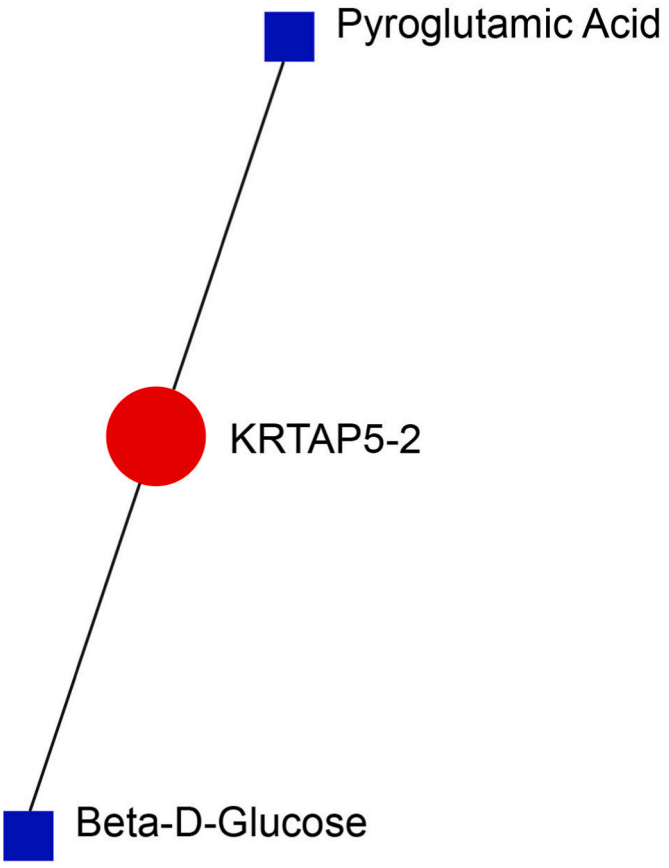

Fig. 9. Interactions between proteins and drugs. A hub protein's association with its drugs is noticed.

In [45] mentioned, only HD information about gene expression and miRNA is provided. Whereas in our study, we examined the hub gene, pathway analysis, gene ontology, potential drug target, and gene disease association with miRNA and TFs both HD and AD.

Kang revealed that TEMRA cells from Alzheimer's disease patients had proinflammatory (IFN- and TNF) and cytotoxic activities (NKG7, GZMA, and B2M) [46]. Huntington's expression was determined using 2CT with -2-microglobulin (B2M) as the reference in Ref. [47]. HLA-A alleles were shown to be linked with a higher chance of obtaining AD in a Chinese population studied by the authors of [48]. According to the authors' results, a subpopulation of brain cells known as angiogenic endothelial cells is produced in Alzheimer's patients. Increased amounts of angiogenic expansion elements are present in these angiogenic epithelial units and recipients (EGFL7, FLT1, and VWF), as well as antigen-presenting components (i.e., B2M and HLA-E) [49]. HLA-A and HLA-C were the most prevalent HLA class - I genotypes in both AD patients and the overall population [50]. HLA-F also implicated in neurotransmission, giving fresh insights into the underlying process [51]. The authors discovered dramatically changed expression of HLA-DRA and IPMK in Alzheimer disease brains compared to normal brains using a gene expression data set [52]. Genome-wide association studies (GWAS) have identified rs9271192, a single-nucleotide polymorphism (SNP) within HLA-DRB1, as a potential marker for AD in Caucasians [53]. A common biomarker is attempting to suppress the elevated signals associated with AD and HD targeted therapy. Gene regulation involves both transcriptional and post-transcriptional stages of regulatory control. Bioinformatics methods were used to identify the most important genes, miRNAs, DEG-miRNA interactions, and their associated pathways in AD and HD [54]. As a result, our defined genes and microRNAs may be used in future molecular studies of AD and HD. To classify DEG regulatory molecules, we examined the important regulatory molecules TF and miRNAs. Regarding DEGs-TF interaction, we established the top five transcription factors, which are MUC20, KRTAP5-2, NOL4L, GSTA1, and CCL2. According to Xiaoyu Dong [55], the miRNA network revealed that hsa-miR-4488, hsa-miR-196a-5p, and hsa-miR-549a had a high degree and may be involved in HD etiology and possible treatment targets. Messenger RNAs are targeted by microRNAs, which are small (22 nt) RNA molecules that disrupt their synthesis. Thus, miRNAs control DEGs in this manner. By providing a comprehensive view of the regulatory mechanism networks underlying TFs, the current research establishes potential molecule targets for genetic counseling and prenatal diagnosis of TFs [55]. According to Zhang [56], future advancements and challenges will be discussed, including more powerful bioinformatics approaches and high-throughput technologies for TF and miRNA target prediction, as well as the integration of multilevel networks. Nowadays, miRNAs are gaining prominence as biomarkers in a variety of complex diseases, including cancer. Mir-10b-5p, mir-941, mir-107, mir-330-3p, mir-26a-5p, mir-34a-5p, mir-26b-5p, mir-124-3p, mir-101-3p, and mir-148-3p are the top 10 regulatory miRNAs identified in our study.

The study of the gene-disease interaction network identified the comorbid disorders linked to the hub genes. The most synchronized diseases to our identified hub genes are Schizophrenia, Chemical and Drug Induced Liver Injury, Asthma, Multiple Sclerosis, Celiac Disease, and Brain Ischemia.

Finally, in order to find new drugs that target the hub proteins, we explored protein-drug interactions and discovered that Pyroglutamic Acid and Beta-D-Glucose drug molecules associate with the hub protein KRTAP5-2. More investigation is required to determine the significance of these medicines in the diagnosis of AD and HD.

5. Conclusion

In this research, we employed a bioinformatics method to examine gene expression transcriptomic profiles in order to identify possible biomarkers that could clarify critical pathobiological pathways influencing AD and HD. We used overlap, core connection, and gene filtering to identify the shared responsive gene between AD and HD.

Following that, we determined signaling pathways and gene ontology processes and then presented a PPI network for shared genes. We utilized transcriptional analysis to determine DEG-miRNA associations as well as protein-protein interactions. The protein–drug and protein–chemical association networks illustrate how pharmacological and chemical substances interact with certain genes. As a result, our methodology will help to push the decision-making process ahead in the field of personalized healthcare. Despite our best efforts, this research contains flaws. The sample size for certain illness studies may be inadequate to describe all of the essential disease-associated genes required to determine frequent DEGs. As such, additional investigation may be necessary to properly evaluate the biological significance of the probable intended possibilities reported in this research.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

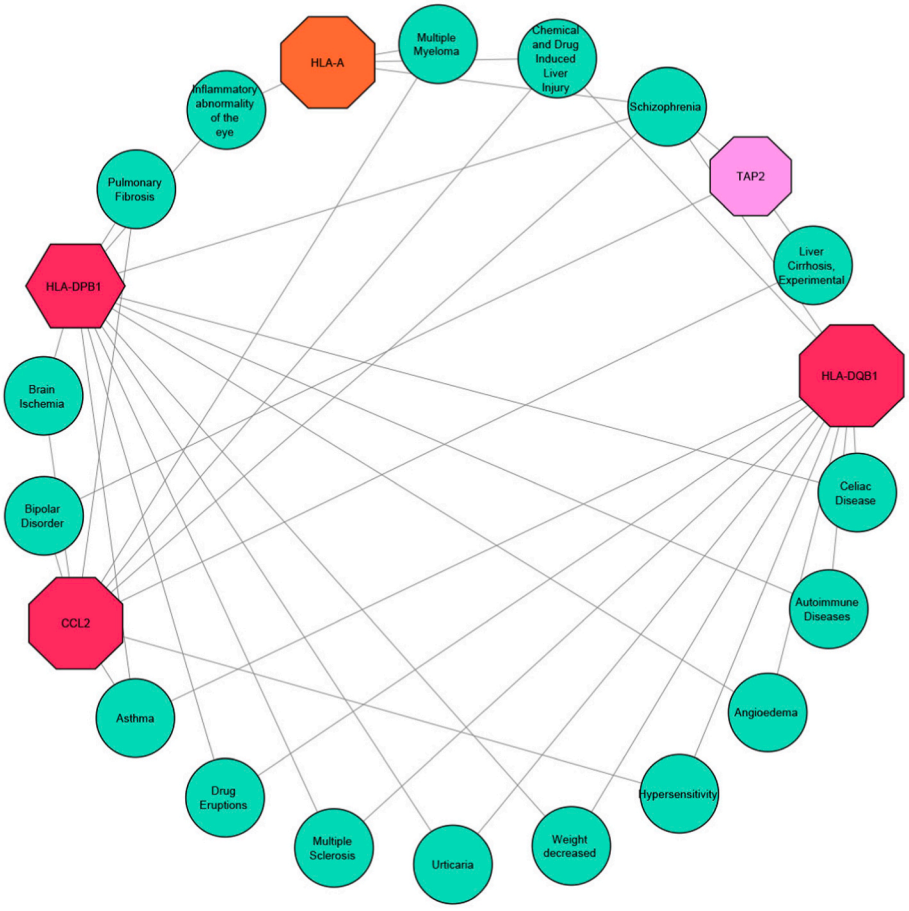

Fig. 10. The network of gene-disease associations depicts common DEG-related diseases. The disease represented by the rectangular node and the following gene symbols is represented by the octogen node.

Author statement

Nosin Ibna Mahbub, Md. Imran Hasan: Analysis and interpretation of the data, Writing– original draft, Writing– review & editing. Md Habibur Rahman: Conception and design, Writing–original draft, Writing– review & editing. Feroza Naznin: Writing–review & editing. Md Zahidul Islam: Writing–review & editing. Mohammad Ali Moni: Writing–review & editing. Md Habibur Rahman and Mohammad Ali Moni supervised the whole project. All authors checked and approved the submissions.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

We would like to thank the team members of the Bioinformatics Lab headed by Dr. Md. Habibur Rahman who helped us and provided valuable insight into the research.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.imu.2022.100888.

References

[1] Association A, et al. 2017 alzheimer's disease facts and figures. Alzheimer's Dementia 2017;13:325–73.
[2] Mucke L. Alzheimer's disease. Nature 2009;461:895–7.
[3] Rahman MH, Rana HK, Peng S, Kibria MG, Islam MZ, Mahmud SH, Moni MA. Bioinformatics and system biology approaches to identify pathophysiological impact of covid-19 to the progression and severity of neurological diseases. Comput Biol Med 2021:104859.
[4] Ferri CP, Prince M, Brayne C, Brodaty H, Fratiglioni L, Ganguli M, Hall K, Hasegawa K, Hendrie H, Huang Y, et al. Global prevalence of dementia: a delphi consensus study. Lancet 2005;366:2112–7.
[5] Ayton S, Lei P, Bush AI. Metallostasis in alzheimer's disease. Free Radic Biol Med 2013;62:76–89.
[6] Castellani RJ, Rolston RK, Smith MA. Alzheimer disease. Disease-a-month: DM 2010;56:484.
[7] Cummings JL, Cole G. Alzheimer disease. JAMA 2002;287:2335–8.
[8] Finkbeiner S. Huntington's disease. Cold Spring Harbor perspectives in biology 2011;3:a007476.
[9] Roos RA. Huntington's disease: a clinical review. Orphanet J Rare Dis 2010;5:1–8.
[10] Walker FO. Huntington's disease. Lancet 2007;369:218–28.
[11] Naia L, Ferreira IL, Ferreiro E, Rego AC. Mitochondrial ca2+ handling in huntington's and alzheimer's diseases–role of er-mitochondria crosstalk. Biochem Biophys Res Commun 2017;483:1069–77.
[12] Moss DJH, Flower MD, Lo KK, Miller JR, van Ommen G-JB, AC't Hoen P, Stone TC, Guinee A, Langbehn DR, Jones L, et al. Huntington's disease blood and brain show a common gene expression pattern and share an immune signature with alzheimer's disease. Sci Rep 2017;7:1–12.
[13] Kumar SU, Kumar DT, Siva R, Doss C, Zayed H. Integrative bioinformatics approaches to map potential novel genes and pathways involved in ovarian cancer. Front Bioeng Biotechnol 2019;7:391.
[14] Udhaya Kumar S, Thirumal Kumar D, Siva R, George Priya Doss C, Younes S, Younes N, Sidenna M, Zayed H. Dysregulation of signaling pathways due to differentially expressed genes from the b-cell transcriptomes of systemic lupus erythematosus patients–a bioinformatics approach. Front Bioeng Biotechnol 2020;8:276.

[15] Fu D, Zhang B, Yang L, Huang S, Xin W. Development of an immune-related risk signature for predicting prognosis in lung squamous cell carcinoma. Front Genet 2020;11:978.
[16] Wan J, Jiang S, Jiang Y, Ma W, Wang X, He Z, Wang X, Cui R. Data mining and expression analysis of differential lncrna adamts9-as1 in prostate cancer. Front Genet 2020;10:1377.
[17] Udhaya Kumar S, Thirumal Kumar D, Bithia R, Sankar S, Magesh R, Sidenna M, George Priya Doss C, Zayed H. Analysis of differentially expressed genes and molecular pathways in familial hypercholesterolemia involved in atherosclerosis: a systematic and bioinformatics approach. Front Genet 2020;11:734.
[18] Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, et al. Ncbi geo: archive for functional genomics data sets—update. Nucleic Acids Res 2012;41:D991–5.
[19] Scheckel C, Drapeau E, Frias MA, Park CY, Fak J, Zucker-Scharff I, Kou Y, Haroutunian V, Ma'ayan A, Buxbaum JD, et al. Regulatory consequences of neuronal elav-like protein binding to coding and non-coding rnas in human brain. Elife 2016;5:e10421.
[20] Lim HG-M, Lee Y-CG. A cross-platform comparison of affymetrix, agilent, and illumina microarray reveals functional genomics in colorectal cancer progression. In: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE; 2019. p. 252–5.
[21] Labadorf A, Hoss AG, Lagomarsino V, Latourelle JC, Hadzi TC, Bregu J, MacDonald ME, Gusella JF, Chen J-F, Akbarian S, et al. Rna sequence analysis of human huntington disease brain reveals an extensive increase in inflammatory and developmental gene expression. PLoS One 2015;10:e0143563.
[22] Rahman MH, Peng S, Chen C, Moni MA, et al. Genetic effect of type 2 diabetes to the progression of neurological diseases. bioRxiv 2018:480400.
[23] Anjum A, Jaggi S, Varghese E, Lall S, Bhowmik A, Rai A. Identification of differentially expressed genes in rna-seq data of Arabidopsis thaliana: a compound distribution approach. J Comput Biol 2016;23:239–47.
[24] Rahman MH, Peng S, Hu X, Chen C, Rahman MR, Uddin S, Quinn JM, Moni MA. A network-based bioinformatics approachto identify molecular biomarkers for type 2 diabetes that are linked to the progression of neurological diseases. Int J Environ Res Publ Health 2020;17:1035.
[25] Bardou P, Mariette J, Escudié F, Djemiel C, Klopp C. jvenn: an interactive venn diagram viewer. BMC Bioinf 2014;15:1–7.
[26] Mahmud SH, Al-Mustanjid M, Akter F, Rahman MS, Ahmed K, Rahman MH, Chen W, Moni MA. Bioinformatics and system biology approach to identify the influences of sars-cov-2 infections to idiopathic pulmonary fibrosis and chronic obstructive pulmonary disease patients. Briefings Bioinf 2021.
[27] Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res 2016;44. W90–W97.
[28] Wang G, Wang F, Huang Q, Li Y, Liu Y, Wang Y. Understanding transcription factor regulation by integrating gene expression and dnasei hypersensitive sites. BioMed Res Int 2015;2015.
[29] Khan A, Fornes O, Stigliani A, Gheorghe M, Castro-Mondragon JA, Van Der Lee R, Bessy A, Cheneby J, Kulkarni SR, Tan G, , et alJaspar. Update of the open-access database of transcription factor binding profiles and its web framework. Nucleic Acids Res 2018.
[30] Hsu S-D, Lin F-M, Wu W-Y, Liang C, Huang W-C, Chan W-L, Tsai W-T, Chen G-Z, Lee C-J, Chiu C-M, et al. mirtarbase: a database curates experimentally validated microrna–target interactions. Nucleic Acids Res 2011;39:D163–9.
[31] Sethupathy P, Corda B, Hatzigeorgiou AG. Tarbase: a comprehensive database of experimentally supported animal microrna targets. Rna12 2006:192–7.
[32] Rain J-C, Selig L, De Reuse H, Battaglia V, Reverdy C, Simon S, Lenzen G, Petel F, Wojcik J, Schächter V, et al. The protein–protein interaction map of helicobacter pylori. Nature 2001;409:211–5.
[33] Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, et al. String v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res 2019;47:D607–13.
[34] Xia J, Gill EE, Hancock RE. Networkanalyst for statistical, visual and network-based meta-analysis of gene expression data. Nat Protoc 2015;10:823–44.
[35] Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, Sajed T, Johnson D, Li C, Sayeeda Z, et al. Drugbank 5.0: a major update to the drugbank database for 2018. Nucleic Acids Res 2018;46:D1074–82.
[36] Piñero J, Bravo À, Queralt-Rosinach N, Gutiérrez-Sacristán A, Deu-Pons J, Centeno E, García-García J, Sanz F, Furlong LI. Disgenet: a comprehensive platform integrating information on human disease-associated genes and variants. Nucleic Acids Res 2016:gkw943.
[37] Consortium GO. Expansion of the gene ontology knowledgebase and resources. Nucleic Acids Res 2017;45:D331–8.
[38] Wittig U, De Beuckelaer A. Analysis and comparison of metabolic pathway databases. Briefings Bioinf 2001;2:126–42.
[39] Mahmud SH, Chen W, Liu Y, Awal MA, Ahmed K, Rahman MH, Moni MA. Predtis: prediction of drug–target interactions based on multiple feature information using gradient boosting framework with data balancing and feature selection techniques. Briefings Bioinf 2021.
[40] Hasan MI, Rahman MH, Islam MB, Islam MZ, Hossain MA, Moni MA. Systems biology and bioinformatics approach to identify blood based signatures molecules and drug targets of patient with covid-19. Inform Med Unlock 2021:100840.
[41] Rahman MH, Rana HK, Peng S, Hu X, Chen C, Quinn JM, Moni MA. Bioinformatics and machine learning methodologies to identify the effects of central nervous system disorders on glioblastoma progression. Briefings Bioinf 2021. bbaa365.
[42] Rahman MH, Peng S, Hu X, Chen C, Rahman MR, Uddin S, Quinn JM, Moni MA. A network-based bioinformatics approach to identify molecular biomarkers for type 2 diabetes that are linked to the progression of neurological diseases. Int J Environ Res Publ Health 2020;17:1035.
[43] Podder NK, Rana HK, Azam MS, Rana MS, Akhtar MR, Rahman MR, Rahman MH, Moni MA. A system biological approach to investigate the genetic profiling and comorbidities of type 2 diabetes. Gene Rep 2020;21:100830.
[44] Rahman MH, Peng S, Hu X, Chen C, Uddin S, Quinn JM, Moni MA. Bioinformatics methodologies to identify interactions between type 2 diabetes and neurological comorbidities. IEEE Access 2019;7:183948–70.
[45] Dong X, Cong S. Bioinformatic analysis of microrna expression in huntington's disease. Mol Med Rep 2018;18:2857–65.
[46] Kang J-S, Liu P-P. Human herpesvirus 4 and adaptive immunity in alzheimer's disease. In: Signal transduction and targeted therapy. 5; 2020. p. 1–2.
[47] Gabert J, Beillard E, Van der Velden V, Bi W, Grimwade D, Pallisgaard N, Barbany G, Cazzaniga G, Cayuela J-M, Cave H, et al. Standardization and quality control studies of 'real-time'quantitative reverse transcriptase polymerase chain reaction of fusion gene transcripts for residual disease detection in leukemia–a europe against cancer program. Leukemia 2003;17:2318–57.
[48] Ma SL, Tang NLS, Tam CWC, Lui VWC, Suen EWC, Chiu HFK, Lam LCW. Association between hla-a alleles and alzheimer's disease in a southern Chinese community. Dement Geriatr Cognit Disord 2008;26:391–7.
[49] Lau S-F, Cao H, Fu AK, Ip NY. Single-nucleus transcriptome analysis reveals dysregulation of angiogenic endothelial cells and neuroprotective glia in alzheimer's disease. Proc Natl Acad Sci Unit States Am 2020;117:25800–9.
[50] Huang P, Yang Y-H, Chang Y-H, Chang S-L, Chou M-C, Lai C-L, Liu C-K, Chen H-Y. Association of early-onset alzheimer's disease with germline-generated high affinity self-antigen load. Transl Psychiatry 2020;10:1–6.
[51] Kawalia SB, Raschka T, Naz M, de Matos Simoes R, Senger P, Hofmann-Apitius M. Analytical strategy to prioritize alzheimer's disease candidate genes in gene regulatory networks using public expression data. J Alzheim Dis 2017;59:1237–54.
[52] Jiang Q, Hu Y, Liu G. Association of alzheimer disease susceptibility variants and gene expression in the human brain. JAMA Neurol 2016;73. 1255–1255.
[53] Lu R-C, Yang W, Tan L, Sun F-R, Tan M-S, Zhang W, Wang H-F, Tan L. Association of hla-drb1 polymorphism with alzheimer's disease: a replication and meta-analysis. Oncotarget 2017;8:93219.
[54] Moradifard S, Hoseinbeyki M, Ganji SM, Minuchehr Z. Analysis of microrna and gene expression profiles in alzheimer's disease: a meta-analysis approach. Sci Rep 2018;8:1–17.
[55] You G, Zu B, Wang B, Fu Q, Li F. Identification of mirna–mrna–tfs regulatory network and crucial pathways involved in tetralogy of fallot. Front Genet 2020;11: 552.
[56] Zhang H-M, Kuang S, Xiong X, Gao T, Liu C, Guo A-Y. Transcription factor and microrna co-regulatory loops: important regulatory motifs in biological processes and diseases. Briefings Bioinf 2015;16:45–58.