Briefings in Bioinformatics, 00(0), 2021, 1–23

doi: 10.1093/bib/bbaa365
Method Review

Bioinformatics and machine learning methodologies to identify the effects of central nervous system disorders on glioblastoma progression

Md Habibur Rahman, Humayan Kabir Rana, Silong Peng, Xiyuan Hu, Chen Chen, Julian M.W. Quinn and Mohammad Ali Moni

Corresponding author: Mohammad Ali Moni , Bone Biology Division, Garvan Institute of Medical Research, Darlinghurst, NSW, Australia and WHO Collaborating Centre on eHealth, School of Public Health and Community Medicine, Faculty of Medicine, The University of New South Wales, Sydney, Australia; E-mail: m.moni@unsw.edu.au.

Abstract

Glioblastoma (GBM) is a common malignant brain tumor which often presents as a comorbidity with central nervous system (CNS) disorders. Both CNS disorders and GBM cells release glutamate and show an abnormality, but differ in cellular behavior. So, their etiology is not well understood, nor is it clear how CNS disorders influence GBM behavior or growth. This led us to employ a quantitative analytical framework to unravel shared differentially expressed genes (DEGs) and cell signaling pathways that could link CNS disorders and GBM using datasets acquired from the Gene Expression Omnibus database (GEO) and The Cancer Genome Atlas (TCGA) datasets where normal tissue and disease-affected tissue were examined. After identifying DEGs, we identified disease-gene association networks and signaling pathways and performed gene ontology (GO) analyses as well as hub protein identifications to predict the roles of these DEGs. We expanded our study to determine the significant genes that may play a role in GBM progression and the survival of the GBM patients by exploiting clinical and genetic factors using the Cox Proportional Hazard Model and the Kaplan–Meier estimator. In this study, 177 DEGs with 129 upregulated and 48 downregulated genes were identified. Our findings indicate new ways that CNS disorders may influence the incidence of GBM progression, growth or establishment and may also function as biomarkers for GBM prognosis and potential targets for therapies. Our comparison with gold standard databases also provides further proof to support the connection of our identified biomarkers in the pathology underlying the GBM progression.

Md Habibur Rahman is an Assistant Professor at the Department of Computer Science and Engineering, Islamic University, Kushtia, Bangladesh. Recently, he has received Ph.D. degree from the Institute of Automation Chinese Academy of Sciences, Beijing, China. His research interest encompasses Bioinformatics, Health Informatics, Machine Learning, Deep Learning, and Medical Data Analysis.
Humayan Kabir Rana is a Senior Lecturer at the Department of Computer Science and Engineering, Green University of Bangladesh. His research interest includes bioinformatics, machine learning, digital image and signal processing.
Silong Peng received the Ph.D. degrees from the Institute of Mathematics Chinese Academy of Sciences, Beijing, China. Currently, he is a Professor at the Institute of Automation Chinese Academy of Sciences. His research interests include wavelets, multivariate signal processing, digital image processing, medical imaging and bioinformatics.
Xiyuan Hu received the Ph.D. degree in pattern recognition and intelligent systems from the Institute of Automation Chinese Academy of Sciences, Beijing, China. He is currently an Associate Professor at the Institute of Automation Chinese Academy of Sciences. His research interests include adaptive signal processing, digital image processing and compression.
Chen Chen received the Ph.D. degrees in computer science from the University of Copenhagen, Denmark. She is currently an Assistant Professor at the Institute of Automation Chinese Academy of Sciences. Her research interests include machine learning, pattern recognition and medical image analysis.
Julian M. W. Quinn received the Ph.D. from the University of Oxford, U.K., in 1992, then moved to Australia to work in bone and cancer biology at the St Vincent's Institute of Medical Research, Prince Henrys Institute (now Hudson Institute) and the Garvan Institute as a Senior Research Officer until 2017. He now works as a Surgical Research Officer at the Surgical Education and Research Training Institute at Royal North Shore Hospital in Sydney Australia. His interests include bone physiology and applications of biostatistics and bioinformatics.
Mohammad Ali Moni is a Research Fellow and Conjoint Lecturer at the University of New South Wales, Australia. He received his Ph.D. in clinical Bioinformatics and Machine Learning from the University of Cambridge. His research interest encompasses artificial intelligence, machine learning, data science, medical image processing and clinical bioinformatics.
Submitted: 26 August 2020; **Received (in revised form):** 25 October 2020

Key words: bioinformatics; machine learning; central nervous system disorders; glioblastoma; comorbidity; pathway; ontology; proteins; survival analysis.

INTRODUCTION

There is an increasing need to facilitate the management of individuals affected by multiple coexisting diseases, i.e. comorbidities that may interact and complicate clinical care [1]. Thus, the coexistence of two or more diseases in an individual raises important questions about their underlying linking etiological pathway and their impact on health care [2]. The interactions of comorbidities are relatively unexplored by biomedical and clinical bioinformatics analyses compared to that of the pathology of single diseases [3]. Comorbidity interactions of brain cancers with central nervous system (CNS) disorders have been noted by a series of epidemiological and clinical studies [4–10], although there is also some evidence of an inverse association between them [11, 12]. Indeed, among the CNS disorders, Alzheimer's disease (AD), epilepsy disease (ED), Huntington's disease (HD), multiple sclerosis (MS) and Parkinson's disease (PD) are neurodegenerative diseases (NDs) that are most often and heavily associated with the progression of brain cancer, particularly with glioblastoma (GBM) [4–10].

AD is a neurological disease characterized by its extracellular deposition of amyloid-β peptides which leads to the death of brain cells and loss of memory [13, 14]. Epidemiological investigations show that patients experiencing AD have a higher risk of developing GBM [15]. Despite AD and GBM are related to the alteration of the same molecular pathways, different studies have demonstrated an inverse association between AD and GBM both from a genetic and epigenetic point of view. Different components have also been proposed to be associated with immediate and converse comorbidities, for example, the climate, way of life or medications and genetic and molecular elements that could at least in part mediate these relationships [16]. Specifically, it has been exhibited that although GBM and AD share similar molecular pathways, substantial contrasts exist in their adjustment. Indeed, while quick cell proliferation and apoptotic cell capture are normal features of GBM, cell harm, and resulting in cell demise are basic consequences in AD [17]. ED [18] is a neurological disease associated with excitatory glutamatergic signaling and γ aminobutyric acid (GABA) [19] mediated impaired inhibitory signaling which leads to the incidence of GBM [20]. Little is known about the relationship and progression of ED and GBM. Glutamine synthetase (GS) inadequacy and high articulation of Lyn and Fyn kinases in the CNS may contribute to both ED and GBM. Greater levels of Fe^{3+} ions in intra-or peri-tumoral regions, because of small hemorrhages from pathological blood vessels, may likewise contribute to the progression of ED and GBM. Epidemiological examination indicates that this connection might be due to molecular-biological changes and isocitrate dehydrogenase (IDH) transformations [21]. Our insight about this relationship, however, is limited and current treatment is a long way from being fully effective [3]. Pathogenic variants of Huntingtin protein (mHTT) encoded by the HTT gene underlies HD development which involves neuronal loss and excitotoxicity in the brain [23] and contributes to the development of GBM [24]. Earlier investigations have indicated some risk factors are shared by HD and GBM, for example, both environmental and intrinsic biological factors, aging, patients with psychiatric symptoms (insomnia, depression), motor symptoms (chorea, ataxia) and apraxia are at risk of developing GBM [25, 26]. MS is a chronic ND with multifactorial pathogenesis that includes genetic and environmental factors [27, 28]. Although the etiology of MS is not clear, CD4+ T helper cells (Th_1 and Th_{17}) seem to play a role to cause GBM [29]. A recent epidemiological study reported that MS patients have an increased occurrence of GBM [30]. Consequently, the chance of common underlying molecular mechanisms in both MS and GBM is possible. Interestingly, recently identified genes associated with MS have functions (or putative functions) in the immune system, and alterations in innate immunity-related genetic regions have been related to the development of GBM. However, so far there is no adequate evidence to confirm this. It might be additionally speculated that variant epigenetic components, for example, DNA methylation and histone protein modifications could be associated with the pathogenesis of both MS and GBM. The molecular evidence of the relationship of human polyomavirus JC virus (JCV) of the Papova virus family is additionally revealed in the pathogenesis of these illnesses. PD [31] results in dopaminergic neurons loss that leads to the resting tremor, rigidity, hypokinesia and postural instability of PD [32] and is associated with the co-occurrence of GBM [33]. While PD and GBM are characterized by very different cellular pathologies, there is evidence of common pathogenic mechanisms affecting PD and GBM. This includes contrarily deregulated pro-survival and immune signaling, mitochondrial dysfunction and metabolic changes. There are inversely regulated common genes that are associated with the two diseases and these suggest some key pathways involving dysregulated cell proliferation and metabolism that PD and GBM have in common. Because of the complex nature of both PD and GBM etiology and pathogenesis, further investigations are needed to reveal to better understand and to compare both diseases and to clarify why similar inverse dysregulated cell pathways can prompt such different diseases. Inevitably, a better comprehension of the pathological mechanisms underlying PD and GBM will help to identify potentially shared medication that could be modulated to treat these diseases [34, 35].

Primary brain tumors are neoplasms developing CNS cells and gliomas are the most common such tumors. These are mainly thought to derive from glial cells or glial progenitors, although this is not always the case [36]. According to their level of malignancy, gliomas have been ranked the scale from grade I to grade IV by World Health Organization (WHO) [37]. Grade IV glioma has more advanced features of malignancy among all grades of glioma. The most common and aggressively malignant tumors, WHO grade IV glioma known as GBM is an extremely heterogeneous cancer whose pathology is characterized by uncontrolled cellular proliferation and its abnormal growth makes it one of the major causes of cancer-related mortality and morbidity [38]. Patients suffering from GBM have a poor prognosis, short medial survival and low response to therapies [39]. GBM tissue samples can usually be clearly characterized by tumor-related protein, lipid, genetic or metabolic markers for diagnosis or treatment [40–42]. A range of molecular techniques has been employed to identify genetic biomarkers

and finding proof of mutations involves deep sequencing, high-resolution melting (HRM), immunohistochemistry, droplet digital PCR (ddPCR) in attempts to identify GBM variants that cannot be identified by histology or other current methods.

Algorithm: Pseodocode for Algorithm
Input: Microarray, RNA-Seq and TCGA datasets containing two different conditions; case samples, control samples. So, notionally we have X (dataset with S samples) $=X_{case} \cup X_{control}$
Output: Differentially expressed genes, Diseasome Network using Differentially Expressed Genes, Enriched Signaling Pathways, Enriched Ontological Pathways, Protein-Protein Interactions (PPIs), Biomarkers genes associated with survival and Survival Curve

1. Searching datasets from public repository with certain criteria
2. For each dataset $i = 1, 2, \ldots, N$:
    1. Load dataset
    2. Normalize datasets
    3. Design matrix model
    4. Convert datasets into expression set class
    5. Create design matrix: case vs control
    6. Fit LIMMA or DESeq2 model based on datasets for filtering the design model
    7. Calculate DEGs
        1. Adjust P-value and logFC
        2. Apply False discovery rate (FDR)
        3. List significant genes
        4. Create and save statistical table
3. Compare gene sets of two different diseases to identify common up and downregulated genes
    - Common upregulated gene set
    - Common downregulated gene set
4. For upregulated and downregulated genes between disease pair
    - Construct diseasome network for upregulated and downregulated DEGs
    - Construct PPIs network for identifying hub genes
    - Perform enrichment analysis for signaling pathways
    - Perform enrichment analysis for Ontological pathways
    - List enriched signaling pathways
    - List enrich ontological pathways
    - Plot signaling pathways
    - Tabulate ontological pathways
5. Preprocessing clinical and gene expression data for same patient from TCGA datasets
6. Regression analysis
    - Fit Cox Proportional Hazard Model for Univariate and Multivariate analysis
    - List biomarkers geens associated with survival
7. Fit PL estimator for Survival Curve
8. Results
    - List of DEGs
    - Gene-Disease association (diseasome) networks
    - PPIs network for hub proteins
    - Signaling pathways
    - Ontological pathways
    - Biomarkers genes associated with survival
    - Survival Curve for biomarkers genes
9. Result comparison with Gold Standard Databases and Literature
    - Consistent result with Gold standard databases and literature

---

Proteomic analysis by mass spectrometry techniques offers some promise of identifying tumor-associated proteins and post-translational alterations that may have pathogenic importance and may find new CNS malignant tumor markers [43, 44]. Proteomic mass spectrometry permits the proteomic profiling of healthy and pathological samples, with the potential to relate protein alterations to a specific disease state [45] to develop new clinical diagnostic methods. A promising approach for GBM diagnosis involves the investigation of biological fluids using liquid biopsies that may characterize particular neoplasms [46].

GBM is very hard to analyze and often untreatable. Existing clinical methodologies for the detection of GBM are insufficient and strictly subject to results acquired utilizing proteomic, lipid biopsy and molecular techniques. The progression of brain tumors, particularly in the beginning phases, might not give any clear and early clinical manifestations. In addition, the utilization of neuroimaging strategies is not possible to be used for mass screening and often does not generally permit adequate identification of the presence and the malignancy of a brain tumor.

One of the principal issues of GBM management is the lack of effective analytic methodologies. In the absence of effective pharmacological and surgical treatments, the identification of early indicative and prognostic biomarkers is of key significance to improve the survival rate of patients and to develop new personalized treatments. To achieve early identification and characterization of a GBM, further study of genetic, proteomic and metabolic changes that are typical of GBM progression is needed to improve early diagnostic methods. However, a single 'omic data analysis alone does not appear to be adequate to successfully characterize the features of tumor progression. We thus utilized our systematic methodologies of genetic, and multi-omics datasets could identify new factors that improve the identification and characterization of GBM tumors and their progression.

Although the causal links between NDs and GBM cancer comorbidity have not yet been elucidated, there are some genetics and environmental stressors as risk factors common to both NDs and GBM [47]. Increasing cellular oxidative stress [48] and inflammation [49] have been reported as potential common etiologies in both disorders. The release of glutamate is one type of causal link between these diseases [50]. Studies suggested that glutamate excitotoxicity is involved in numerous NDs such as AD, ED, HD, MS and PD, which results in the long-term progressive neuronal loss [51]. Although the etiology of NDs and GBM remains unclear and both cells release glutamate and are characterized by abnormal behavior, it remains a challenge to identify and understand the effect of NDs on the progression of GBM and factors that affect the survival of GBM cancer patients. To address these problems, the main objectives of this study were to develop an integrated framework based on the bioinformatics and machine learning model to examine the role of

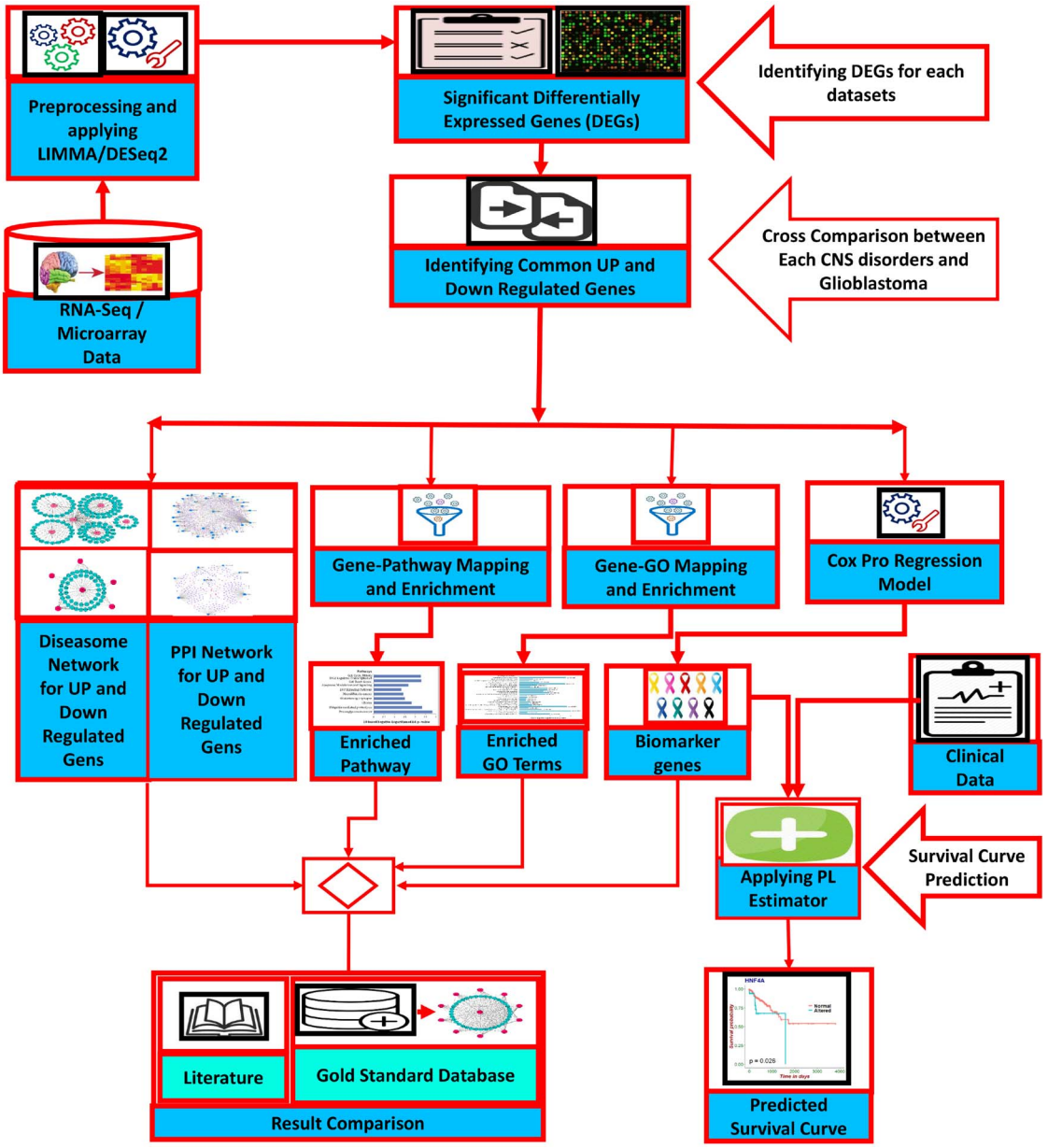


FIG. 1. Overview of the proposed bioinformatics and machine learning approach.

NDs on brain cancer and how they cause the incidence and progression of GBM by affecting molecular pathways and genes altered in GBM.

Firstly, our bioinformatics frameworks examine gene expression profiles to unravel shared differentially expressed genes (DEGs) that are altered in NDs and GBM and then filtering shared DEGs through disease-gene association (diseasome) networks, signaling pathway, ontological pathway, hub protein identification from protein–protein interactions to predict the function of these DEGs. However, there are few effective and efficient machine learning methods to identify cancer biomarker genes that are dysregulated in both NDs disorder and GBM and affect cancer patient survival. We employed the Cox Proportional Hazard model [52] and product limit (PL) estimator [53] to investigate the effect of clinical and genetic factors utilizing The Cancer Genome Atlas (TCGA) data that play a significant role in the survival of cancer patients by survival analysis. Finally, we have also compared our findings that our network analysis has identified through the use of gold benchmark databases dbGaP, OMIM, OMIM Expanded, and literature review which provide further proof to support the connection of our identified genes in the pathology underlying the GBM progression.

MATERIALS AND METHODS

Literature search and datasets

We systematically surveyed research of epidemiological and clinical studies reported on the progression of brain cancer in patients with NDs published until October 2019. We originally studied epidemiological, clinical and neuroimaging studies to identify brain cancer linked to NDs, i.e. brain cancers that are

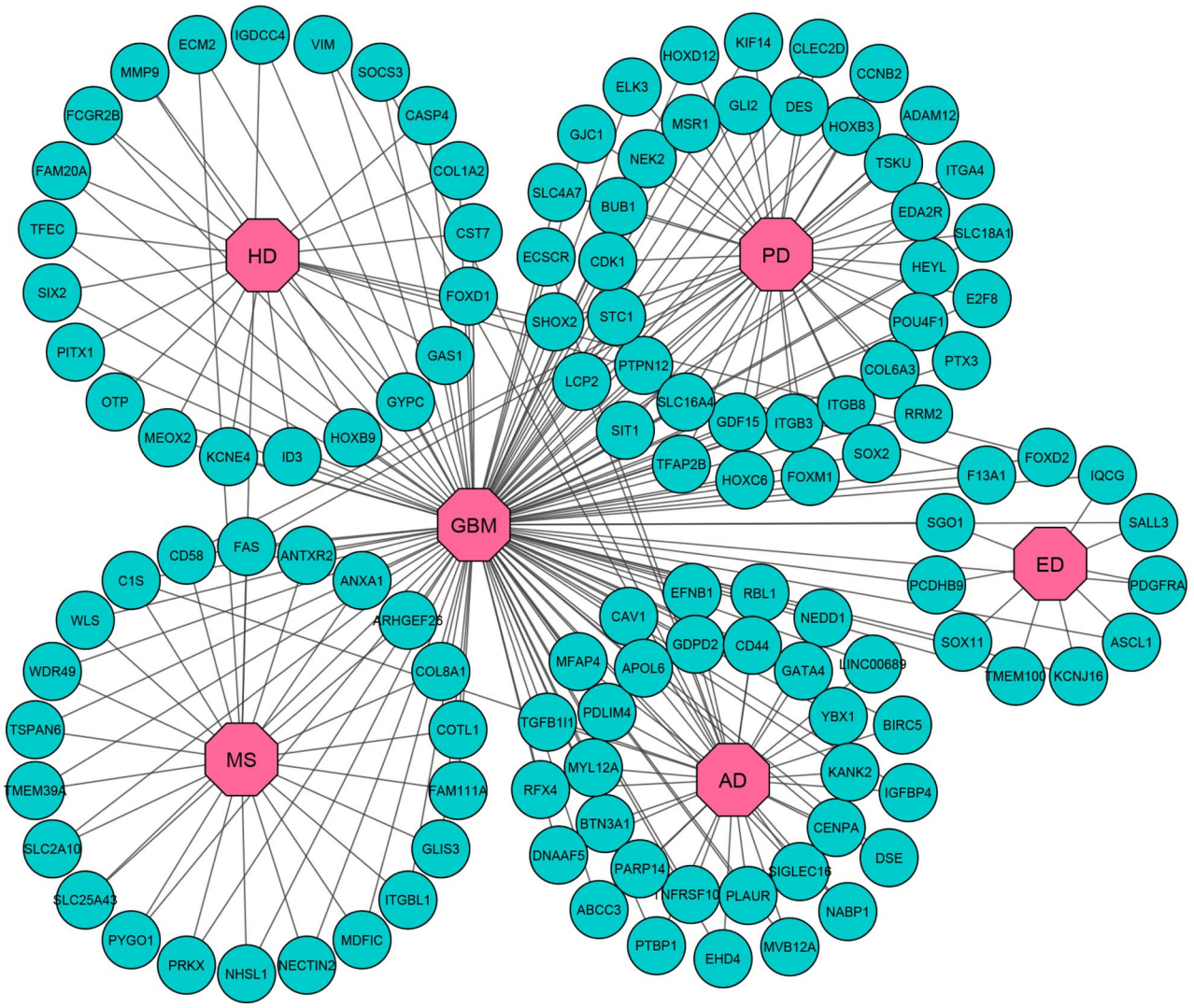


Fig. 2. Diseasome network for upregulated and downregulated DEGs shared between CNS disorders and GBM. The node legends are: red color nodes for diseases and round-shaped blue color nodes for genes.

Table 1. Statistics of the DEGs for selected CNS disorders and GBM cancer.

Disease Name	GEO Number with reference	Tissues	Platform	Control Samples	Case Samples	Sigt. Genes	UP Reg. Genes	Down Reg. Genes
AD	GSE28146 [56]	Hippocampal CA1 gray matter	Affymetrix Human Genome U133 Plus 2.0 Array	8	22	451	307	144
ED	GSE32534 [57]	Brain	Affymetrix Human Genome U133 Plus 2.0 Array	5	5	635	351	284
HD	GSE64810 [58]	Post mortem brain	Illumina HiSeq 2000	49	20	1211	684	527
MS	GSE52139 [59]	Brain	Affymetrix Human Genome U133 Plus 2.0 Array	8	8	820	342	478
PD	GSE19587 [60]	Post mortem brain	Affymetrix Human Genome U133A 2.0 Array	10	12	1052	794	258
GBM	GSE59612 [61]	Brain Tumor	Illumina HiSeq 2000	17	75	2577	1578	999

affected by the presence of NDs. Brain cancer is linked to various neurological diseases, among which we selected for study are AD, ED, HD, MS and PD. We first obtained some datasets that are freely available in [54] and the European Bioinformatics Institute (EBI) [55]. In this work, we searched datasets based on particular criteria (as indicated) for each disease. We thus retrieved a number of datasets from public repositories, among them some are RNA-seq and some are microarray data. We examined datasets from public resources and we collected RNA-seq and microarray datasets but most were discarded as they did not conform to our

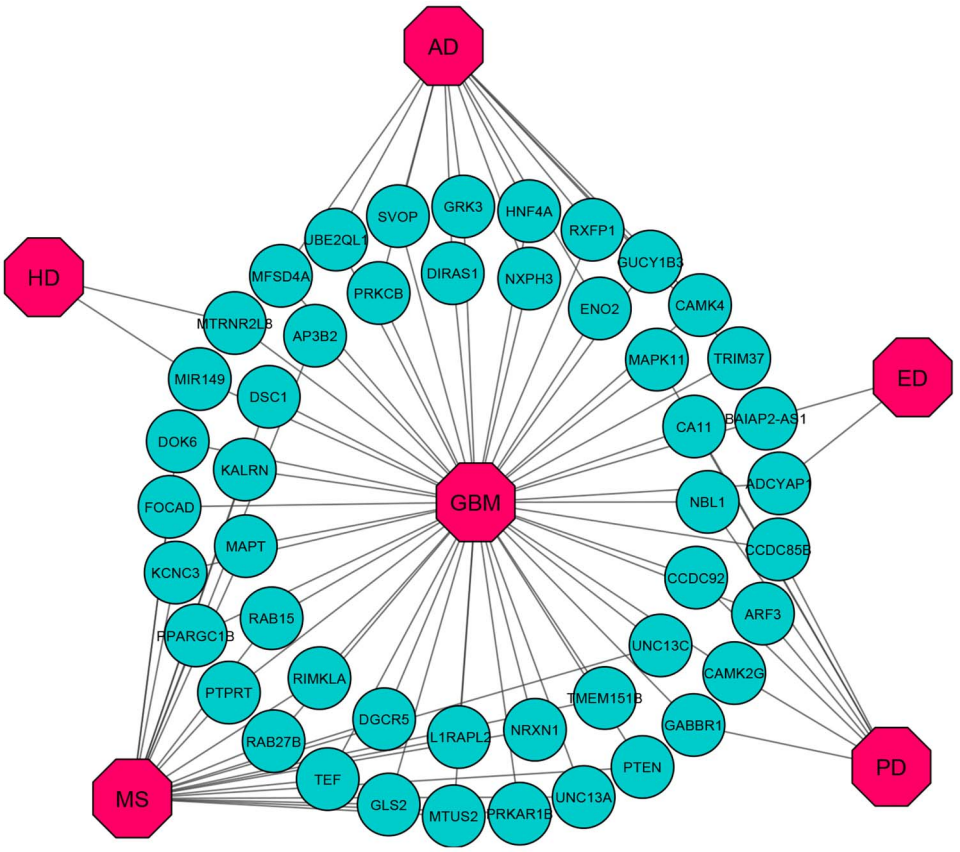


Fig. 2. continued

study criteria. The datasets that include a very low sample size (cutoff sample size of at least 10), missing control and case conditions, datasets of replicated samples, unexpected formatting and datasets that are not from human organisms were discarded by checking. We found that the selected datasets are suitable comparing to the other available datasets and appropriate for our study. We filtered the datasets to choose for this work those that show minimal bias and noise. DEGs that were found in more than one disease are termed here 'common' DEGs for those diseases. Further filtering was applied in the pipeline to the processed data (i.e. o common DEGs identified) which also reduces bias and noise.

We filtered six different microarray and RNA-seq datasets with accession numbers GSE28146, GSE32534, GSE64810, GSE52139, GSE19587 and GSE59612 [56–61]. The AD dataset (GSE28146) is a microarray dataset which consists of only CA1 hippocampal gray matter extracted by laser capture methods from snap-frozen brain tissue [56] with 8 control and 22 disease-affected samples. The ED dataset (GSE32534) is a microarray dataset derived from peritumoral neocortex tissue samples, which consists of five ED subjects and five control subjects [57]. The HD dataset (GSE64810) is an mRNA expression dataset from human prefrontal cortex tissue with 20 HD subjects and 49 neuropathologically normal controls samples extracted by next-generation high-throughput sequencing [58]. The MS dataset (GSE52139) is derived from RNA extraction and hybridization on Affymetrix microarrays derived from the spinal cord, which consists of eight MS subjects and eight control subjects [59]. The PD dataset (GSE19587) is a microarray dataset collected from the human postmortem brain having 12 PD subjects and 10 control samples extracted by Affymetrix U133A Plus 2.0 arrays [60]. The GBM dataset (GSE59612) is an RNA-seq dataset, obtained from radiographically localized biopsies during glioma resection surgeries, which is a study of 75 GBM and 17 control brain tissue samples [61].

To evaluate the patient's survival for the dysregulated significant genes overlapped between NDs and GBM, we acquired clinical and RNA-seq data for GBM from the cBioPortal [62, 63]. GBM clinical datasets have 592 samples with 70 features, RNA-seq gene expression data samples have 397 gene mutation for 575 cases [64]. We used clinical and genetic factors to evaluate the survival of brain cancers (GBM) patients.

Data preprocessing and identification of DEG

Microarray and RNA-seq datasets based gene expression analysis are sensitive methods to study global gene expression and to identify possible molecular pathways that are activated in human tissues affected by disorders [65]. We can mine such data to identify biomarker genes that are associated with GBM cancer progression and cancer patient survival. It is a challenge to achieve this in complex disease prognostic studies but it can result in a method for making more accurate prognoses [66]. All these microarrays and RNA-seq-based datasets were generated by comparing the transcriptome profile of a diseased tissue against control (non-diseased) tissues. As the generated data are from different sources and cell types, we performed the preprocessing of our data using the quintile normalization and Z-score transformation [67] to keep away from complications. We then applied Linear Models for Microarray (limma)

[68] and DESeq [69] implemented in programming language R to identify the dysregulated DEGs for each dataset. To alleviate the effect of outliers, we employed $\log_2$ transformation, and then the shared dysregulated DEGs between the two diseased were identified. We then filtered the statistically significant dysregulated DEGs that fulfilled the criteria of an adj. P-value < 0.01 and the absolute value of a $\log_2$ fold change of 1.0. We used the Benjamini–Hochberg (BH) and Bonferroni correction methods to adjust P-values. Gene expression dysregulation can be expressed mathematically as follows:

$$DEG_i = \begin{cases} UP - regulated & \text{if adj. P-value } < 0.01 \ \& \ \text{logFC } \geq 1.0, \\ DOWN - regulated & \text{if adj. P-value } < 0.01 \ \& \ \text{logFC } \leq -1.0. \end{cases} \quad (1)$$

Here, DEG_i is the i*th* DEG which has either upregulated or downregulated value based on the criteria stated in the above equation. We employed the multilayer topological and neighborhood-based benchmark approach to represent the disease-gene association. We built up disease-gene association (diseasome) network where the network node is either a gene or disease node. In the diseasome network, a disease is connected with other diseases if the disease shares at least one or more than one dysregulated DEGs. We consider that D is disease sets and G is DEGs sets, then diseasome networks attempt to represent gene $g \in G$ is associated with disease $d \in D$. If G_i and G_j are dysregulated genes sets associated with diseases D_i and D_j, respectively, then the number of overlapping dysregulated DEGs (n_{ij}^g) linked with both diseases D_i and D_j are computed by the following equation [70–72]:

$$n_{ij}^g = N(G_i \cap G_j) \quad (2)$$

Following [65], the common neighbors are identified and the edge score for the neighboring node pair is calculated as follows:

$$E(i,j) = \frac{N(G_i \cap G_j)}{N(G_i \cup G_j)} \quad (3)$$

where G indicates the set of network nodes, E indicates the set of all edges between nodes.

Pathway and functional enrichment analysis

To provide additional insights into the new molecular signaling mechanisms and biological significance and look over how the factor that contributes to the generation of a trait from the NDs tissues and relates with expression alterations of the GBM gene, we performed gene set enrichment analysis via Enrichr [73]. For pathway enrichment analyses, we employed KEGG [74], Reactome [75], Wiki [76] and Biocarta databases [77] to identify molecular pathways revealed by the DEGs common between CNS disorders and GBM. The enrichment analyses of the DEGs common between the CNS disorders and GBM revealed Gene ontology (GO) terms considering the GO domain; biological process (BP) by ontological enrichment analysis [78]. Statistical parameter, adjusted P-value < 0.01 was selected for enrichment analysis.

Protein–protein interaction analysis

For the study of the protein–protein interaction (PPI), we utilized the STRING database [79] for upregulated and downregulated DEGs via a visualization software, Network Analyst [80]. The highest confidence score of 900 was used as a cutoff value for interaction. The hubs proteins [81] were selected by setting the topological parameter, a degree greater than 15. The distance D_s between a pair of protein (i, j) in the PPI is computed by the following equations:

$$D_s(i,j) = \frac{2|N_i \cap N_j|}{|N_i| \cup |N_j|} \quad (4)$$

where N_i indicates the set of neighbors i and N_j indicates the set of neighbor set j.

Survival analysis

GBM is a complex cancer disease that is caused by abnormalities and affects the expression patterns of genes. As we aim to predict the survival of GBM cancer patients using clinical and gene expression data [62], we acquired clinical and RNA-seq data for GBM cancer (Glioblastoma (TCGA, PanCancer Atlas) from the cBioPortal [62] where both clinical and RNA-seq data are available for 585 GBM patients.

In the RNA-seq data, we identified the normal tissue samples and tumor tissue samples by checking the TCGA barcode. We performed the transformation of the RNA-seq data using Z-scores transformation for each gene expression value. We calculated the Z-score value for RNA-seq data as follows:

$$Z = \frac{\text{Value for gene X in brain cancer - Mean value for gene X in Normal}}{\text{Standard deviation of the value for gene X in reference}} \quad (5)$$

We applied this transformation to determine the altered and normal (unaltered) expression value. Differentially expressed altered samples are divided into overexpress and underexpressed. We, therefore, determined the altered and normal samples by setting the threshold value as follows:

$$Z \geq 2 \succ \text{Overexpress} \quad (6)$$

$$Z \leq -2 \succ \text{Underexpress} \quad (7)$$

$$2 < Z > -2 \succ \text{Normal (Unaltered)} \quad (8)$$

To predict the effect of clinical and genetic factors that affect the relative risk of patient's survival for biomarker genes that are common to NDs and the GBM, we applied the standard Cox Proportional Hazards Model for univariate and multivariate analysis [52].

The Cox Proportional Hazards Model is written as follows:

$$h_{Cox}(t|X_i) = h_o(t)\exp(\beta^T X_i) \quad (9)$$

Here $h_{Cox}(t|X_i)$ is the expected conditional hazard function at time t for a subject i with covariate information given that vector X_i, $h_o(t)$ is the covariate information independent baseline hazard function and β indicates the corresponding regression coefficients vector for the covariates. The estimated regression

coefficient from the fitted model is used to calculate the hazard ratio (HR) to identify which covariates affect cancer patient survival. If x_r is a covariate, then it s hazard ratio is calculated by exponentiating the corresponding regression coefficient, β_r such as $\exp(\beta_r)$. To determine survival status, PL estimator known as Kaplan–Meier estimator is defined as follows [53]:

$$\hat{S}_f(t_j) = \prod_{i=1}^{j} \left(1 - \frac{d_j}{n_j}\right) \tag{10}$$

Here $\hat{S}_f(t_j)$ indicates estimated survival function at time t_j and d_j indicates the number of events at time t_j, and n_j indicates subjects number at time t_j. After estimating the survival function, we use the Log-rank test for the comparisons of altered versus normal group of genes associated with patient's survival. The expressions of the null hypothesis are written as follows:

$$H_o : S_{altered}(t) = S_{\text{unaltered}}(t) \tag{11}$$

$$H_A : S_{altered}(t) \neq S_{\text{unaltered}}(t) \tag{12}$$

Here the function H_o is the same for altered (overexpress or underexpressed) and unaltered (normal) groups of genes and H_A is not the same for altered and unaltered groups of genes.

Statistical analyses

The chosen gene expression datasets and their matric information was firstly downloaded and converted into a class for differential gene expression analysis. After reviewing the sample records (GSM) manually for sample classification, we constructed design models (patients, controls). This created a design model which was then filtered using LIMMA/DESeq2. Using a threshold of adjusted P-value and absolute log Fold Change (logFC) values of at most 0.01 and at least 1.0, respectively, we identified DEGs. After the comparison between the datasets from the two diseases, we obtained all the upregulated and downregulated DEGs. We then constructed upregulated and downregulated diseasome networks as well as upregulated and downregulated PPI networks and we performed enrichment analysis to identify the signaling pathway and ontological pathway. Then, we prepare clinical and genetic factors for the same patients. Subsequently, we fitted a Cox Proportional Hazard Model for univariate and multivariate analysis to identify cancer biomarkers genes associated with cancer patient survival. Finally, we fitted a PL estimator to construct a survival curve for biomarker DEGs and compared our results with those from gold standard databases and literature.

Overview of proposed integrated bioinformatics and machine learning approach

We designed and developed a multistage quantitative framework as an integrated pipeline based on bioinformatics and machine learning methodologies as shown in Figure 1 and we also provided a representative picture of the algorithm used for the selection of studies and datasets here analyzed which is shown in Pseudocode for Algorithm.

RESULTS

Gene expression analysis

To look over the effects of NDs that influence the progression of brain cancer, we used the gene expression microarray and RNA-seq data collected from NCBI and EBI. By using our proposed method, the top significant DEGs was identified by choosing $<$ 0.01 and the absolute logFC of 1 which is summarized in Table 1.

We identified overlapping DEGs between NDs and GBM, termed as biomarker genes by matching upregulated DEGs of NDs with the upregulated genes of GBM and downregulated DEGs of NDs with the downregulated genes of GBM. In this way, 49 biomarker genes were identified between AD and GBM, 11 biomarker genes were identified between ED and GBM, 22 biomarker genes were identified between HD and GBM, 47 biomarker genes were identified between MS and GBM and 48 biomarker genes were identified between PD and GBM. We performed hypergeometric test and Jaccard index tests for the DEG's genes in order to establish their role as predictive diagnostic biomarkers for brain tumors, or other CNS disorders which are included in Table A in the supplementary file (S1). To reveal the significant associations between GBM and NDs, we built up two separated diseasome networks for upregulated and downregulated genes that are common between GBM and NDs as shown in Figure 2.

Pathway and functional association analysis

The relation between complex diseases through underlying molecular mechanisms are understood by signaling pathways [83]. The pathway enrichment analysis is a technique to determine what pathways are activated in common between diseases [84]. After identifying biomarker genes overlapping between NDs and GBM, we performed pathway enrichment analysis using EnrichR [73]. Signaling pathways enrichment analysis of the commonly dysregulated gene between NDs and GBM were performed using KEGG, Reactome, Wiki and BioCarta pathway databases. We emphasized more on pathways by considering four pathways datasets and by manually curation we identified a number of 9, 8, 7, 6 and 10 enriched significant signaling pathways common between AD and GBM, ED and GBM, HD and GBM, MS and GBM and PD and GBM, respectively, having an adjusted P-value of below 0.01 as shown in Figure 3.

Gene ontological analysis

GO is a comprehensive conceptual model to represent gene and gene product functions [85]. For the identified biomarker genes seen in NDs cells and GBM, we performed GO terms enrichment analysis to identify the most significant ontological pathways common between NDs and GBM employing a comprehensive gene set enrichment analyzing tool, Enrichr [73] using GO; BP. Notably, we performed functional annotation through GO enrichment analysis and by manually curation we found 13, 14, 13, 15 and 15 ontological pathways between AD and GBM, ED and GBM, HD and GBM, MS and GBM and PD and GBM, respectively, considering adjusted P-value of below 0.01 as shown in Table 2.

PPI analysis

A PPI is the physical interactions of proteins in the cell. To understand interactions in the cell physiology, we performed the analysis of the protein–protein interaction of DEGs using STRING via Network Analyst [79]. We built up the PPI network using upregulated and downregulated DEGs to reveal the hub proteins

Table 2. Enriched ontological pathways common between GBM and each NDs/CNS disorder **(a)** AD **(b)** ED **(c)** HD **(d)** MS and **(e)** PD.

GO ID	Ontological Pathway	Genes in the pathway	Adj. P-value
	(a) Ontological pathways common between GBM and AD.		
GO:2000045	regulation of G1/S transition of mitotic cell cycle	KANK2;RBL1	5.68E-03
GO:2001242	regulation of intrinsic apoptotic signaling pathway	CAV1;PLAUR	5.92E-03
GO:0019221	cytokine-mediated signaling pathway	SOCS3;CD44	1.89E-02
GO:0043011	myeloid dendritic cell differentiation	CAMK4	2.42E-02
GO:0042098	T cell proliferation	BTN3A1	2.66E-02
GO:0043067	regulation of programmed cell death	KANK2;BIRC5	2.79E-02
GO:1903900	regulation of viral life cycle	MVB12A	2.90E-02
GO:1902165	regulation of intrinsic apoptotic signaling pathway	CD44	2.90E-02
GO:0070102	interleukin-6-mediated signaling pathway	SOCS3	3.38E-02
GO:0042127	regulation of cell proliferation	KANK2;TGFB1I1	3.42E-02
GO:0030856	regulation of epithelial cell differentiation	CAV1	4.32E-02
GO:0034114	regulation of heterotypic cell-cell adhesion	CD44	4.56E-02
GO:0000188	inactivation of MAPK activity	CAV1	4.79E-02
	(b) Ontological pathways common between GBM and ED.		
GO:0050767	regulation of neurogenesis	SOX11;ASCL1	4.32E-04
GO:0045664	regulation of neuron differentiation	SOX11;ASCL1	9.35E-04
GO:0033674	positive regulation of kinase activity	PDGFRA;ADCYAP1	1.68E-03
GO:0060253	negative regulation of glial cell proliferation	SOX11	3.30E-03
GO:0060563	neuroepithelial cell differentiation	SOX11	5.49E-03
GO:0014033	neural crest cell differentiation	SOX11	8.22E-03
GO:0030858	positive regulation of epithelial cell differentiation	TMEM100	1.04E-02
GO:0043410	positive regulation of MAPK cascade	PDGFRA;ADCYAP1	1.05E-02
GO:0021782	glial cell development	SOX11	1.09E-02
GO:0051960	regulation of nervous system development	ASCL1	1.37E-02
GO:0010001	glial cell differentiation	SOX11	1.37E-02
GO:0048145	regulation of fibroblast proliferation	PDGFRA	2.50E-02
	(c) Ontological pathways common between GBM and HD.		
GO:0019221	cytokine-mediated signaling pathway	MT2A;MMP9	2.58E-04
GO:2001235	positive regulation of apoptotic signaling pathway	MMP9;S100A9	3.77E-03
GO:0030198	extracellular matrix organization	COL1A2;MMP9	1.46E-02
GO:0070098	chemokine-mediated signaling pathway	CCL2;PPBP	1.93E-02
GO:0030182	neuron differentiation	HAND2;ID3	1.95E-02
GO:0034350	regulation of glial cell apoptotic process	CCL2	2.44E-02
GO:0043067	regulation of programmed cell death	KANK2;GAS1	2.44E-02
GO:0034351	negative regulation of glial cell apoptotic process	CCL2	2.84E-02
GO:0045687	positive regulation of glial cell differentiation	CXCR4	4.03E-02
GO:0032874	positive regulation of stress-activated MAPK cascade	HAND2;FCGR2B	4.27E-02
GO:1901215	negative regulation of neuron death	SIX1;CCL2	4.56E-02
GO:0048713	regulation of oligodendrocyte differentiation	CXCR4	4.81E-02
GO:0001558	regulation of cell growth	MSX1;S100A9	4.97E-02
	(d) Ontological pathways common between GBM and MS.		
GO:0051966	regulation of synaptic transmission, glutamatergic	UNC13A;NRXN1	5.02E-03
GO:0021955	central nervous system neuron axonogenesis	PTEN	1.63E-02
GO:0001774	microglial cell activation	MAPT	1.63E-02
GO:0070302	regulation of stress-activated protein kinase signaling cascade	FAS	1.86E-02
GO:0030307	positive regulation of cell growth	UNC13A;MAPT	2.23E-02
GO:0043068	positive regulation of programmed cell death	FAS;KALRN	2.24E-02
GO:0002858	regulation of natural killer cell mediated cytotoxicity directed against tumor cell target	NECTIN2	2.33E-02
GO:0014002	astrocyte development	MAPT	2.33E-02
GO:0048854	brain morphogenesis	PTEN	2.56E-02
GO:0016055	Wnt signaling pathway	PTEN;WLS	2.71E-02
GO:0051895	negative regulation of focal adhesion assembly	PTEN	3.24E-02
GO:0030334	regulation of cell migration	ANXA1;PTEN	3.80E-02

(Continued)

Table 2. Continued

GO ID	Ontological Pathway	Genes in the pathway	Adj. P-value
GO:0030856	regulation of epithelial cell differentiation	PRKX	4.15E-02
GO:0042981	regulation of apoptotic process	GLS2;FAS;KALRN	4.16E-02
GO:0006536	glutamate metabolic process	GLS2	4.82E-02
	(e) Ontological pathways common between GBM and PD.		
GO:0030198	extracellular matrix organization	ITGB3; ADAM12	2.63E-06
GO:0045664	regulation of neuron differentiation	SOX2; HEYL	4.47E-06
GO:0043523	regulation of neuron apoptotic process	TFAP2B; TGFB2	9.24E-05
GO:0045666	positive regulation of neuron differentiation	TCF3; NBL1	1.46E-04
GO:0051726	regulation of cell cycle	CCNB2; TGFB2	8.22E-04
GO:0043406	positive regulation of MAP kinase activity	MAPK11; CDK1	2.14E-03
GO:0042127	regulation of cell proliferation	TFAP2B; BTG3;	4.51E-03
GO:0050678	regulation of epithelial cell proliferation	SOX2; TGFB2	5.95E-03
GO:1901215	negative regulation of neuron death	KIF14; SIX1	8.81E-03
GO:0007346	regulation of mitotic cell cycle	BTG3; CDK1	1.02E-02
GO:0046578	regulation of Ras protein signal transduction	TGFB2; KIF14	1.10E-02
GO:0045597	positive regulation of cell differentiation	MSR1; MAPK11	1.75E-02
GO:0001558	regulation of cell growth	TGFB2; CCDC85B	1.96E-02
GO:0007399	nervous system development	SHOX2; CELSR1	2.90E-02
GO:0045685	regulation of glial cell differentiation	CDK1	3.02E-02

considering the topological parameters, degree greater than 15 and the interaction level 900 as confidence score as shown in Figure 4.

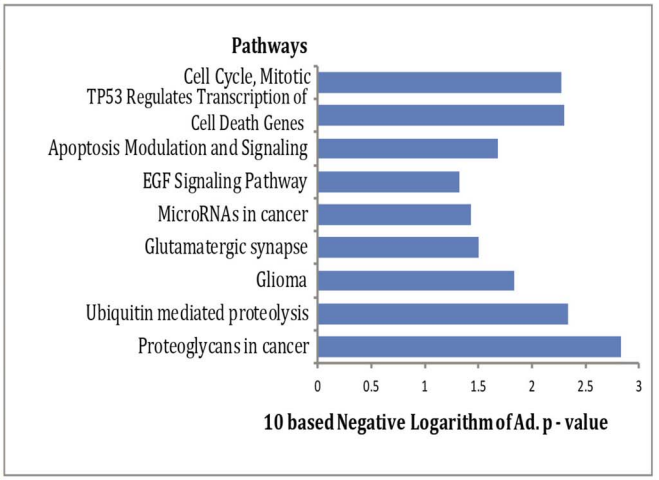
The PPI network of upregulated DEGs had 980 nodes and 1454 edges with a topological parameter, a degree greater than 15, while the PPI network of downregulated DEGs had 287 nodes and 317 edges with a degree greater than 15. More interactions than expected were observed for PPI networks of both upregulated and downregulated DEGs due to medium stringency score (900). We performed the topological analysis and detected hub genes using degree matrices (degree greater than 15) for both the upregulated and downregulated DEGs networks using visualization software: Network Analyst. A total of 30 hub proteins (CDK1, CCNB1, CAV1, CENPA, YBX1, BUB1, SOCS3, WNT5A, RBL1, CCNB2, CDC45, NDC80, EFNB1, GATA4, BIRC5, ESPL1, UBC, TCF3, CXCR4, GLI2, CD44, PDGFRA, PTBP1, LCP2, PRKX, SGO1, WLS, NEK2, TGFB2, FAS) were identified for upregulated DEGs and eight hub proteins (PTEN, MAPK11, HNF4A, MAPT, PRKCB, PRKAR1B, KALRN, CAMK2G) for downregulated DEGs were identified. The specified topological properties of the top hub genes from the PPI network was found for upregulated and downregulated PPI networks using the Cytoscape tool [86]. We have provided the specified topological parameters as Table B and Table C in the supplementary file (S1) for upregulated and downregulated PPI networks. Among the parameter list, the degree is the most fundamental trait of a node in a network which is characterized as the number of contiguous connections, for example, the number of communications that interface one protein to its neighbors. Betweenness centrality is a proportion of a vertex's impact on the data stream for each set of vertices, expecting that information flows mainly through the shortest pathways between them. Closeness centrality is an approach to identify objects that can viably send data through a network. A hub protein's closeness centrality determines its normal distance to every other node. Nodes with a high score of closeness have the shortest lengths to all the other nodes. An average clustering coefficient is a function of the level of the average clustering of hub nodes in a PPI network. The topological coefficient is the quantitative measurement of the degree to which a node connects neighbors with other nodes. A topological coefficient of 0 is estimated to hub nodes that have one or no neighbors. The identified hub proteins are potential biomarkers which may lead to new GBM therapeutic targets and may play significant key roles in signal transduction during the progression of GBM.

Survival analysis

To analyze survival, we employed both gene expression and clinical data acquired from TCGA and measured the risk of survival of the DEGs associated with GBM patients. Consequently, we determine significant genes that are involved in the pathological processes of GBM progression and the survival of GBM cancer patients. In our study, the survival function for the significant genes of two groups (altered and unaltered) that are common between NDs and the GBM is estimated by applying the Cox Proportional Hazard Model and PL estimator. Both the univariate and multivariate regression is fitted in this study and the regression results of the most important statistically significant DEGs which are common between GBM and our selected NDs of CNS disorder are shown in Table 3, Table 4, Table 5, Table 6 and Table 7, respectively.

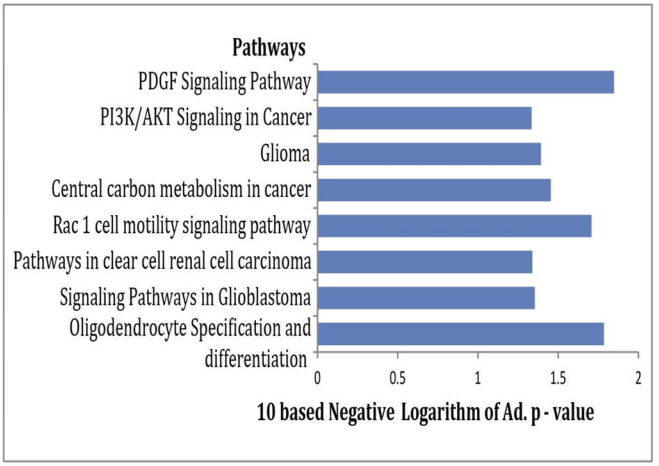
Based on the nonzero regression coefficients, we found 15 genes between AD and GBM, 3 genes between ED and GBM, 10 genes between HD and GBM, 10 genes between MS and GBM and 16 genes between PD and GBM from the RNASeq data as the most significant differentially expressed common biomarker genes by picking out a threshold of P-value ≤ 0.05. In Table 3, Table 4, Table 5, Table 6 and Table 7, the positive sign of the regression coefficients indicates that the hazard (risk of death) is higher and thus prognosis worse. The P-value of

Pathway	Genes in the pathway	Adjusted p-valu
Proteoglycans in cancer	PRKCB, CAV1, PLAUR, CD44	1.47E-03
Ubiquitin mediated proteolysis	SOCS3, UBE2QL1, TRIM37	4.60E-03
Glioma	PRKCB, CAMK4	1.46E-02
Glutamatergic synapse	GRK3, PRKCB	3.18E-02
MicroRNAs in cancer	PRKCB, VIM, CD44	3.68E-02
EGF Signaling Pathway	PRKCB	4.79E-02
Apoptosis Modulation and Signaling	TNFRSF10C, BIRC5	2.10E-02
TP53 Regulates Transcription of Cell Death Genes	TNFRSF10C, BIRC5	4.98E-03
Cell Cycle, Mitotic	NEDD1, RBL, PRKCB, BIRC5, CENPA	5.32E-03



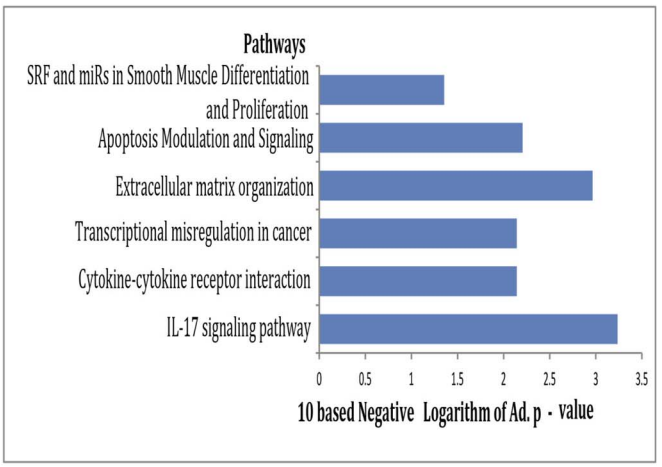
(a) Pathways associated with significantly common differentially expressed genes of the AD with Glioblastoma.

Pathway	Genes in the pathway	Adjusted p-value
Oligodendrocyte Specification and differentiation	ASCL1	1.64E-02
Signaling Pathways in Glioblastoma	PDGFRA	4.42E-02
Pathways in clear cell renal cell carcinoma	PDGFRA	4.58E-02
Rac 1 cell motility signaling pathway	PDGFRA	1.96E-02
Central carbon metabolism in cancer	PDGFRA	3.52E-02
Glioma	PDGFRA	4.05E-02
PI3K/AKT Signaling in Cancer	PDGFRA	4.63E-02
PDGF Signaling Pathway	PDGFRA	1.42E-02



(b) Pathways associated with significantly common differentially expressed genes of the ED with Glioblastoma.

Pathway	Genes in the pathway	Adjusted p-value
IL-17 signaling pathway	CCL2, MMP9, S100A9	5.81E-04
Cytokine-cytokine receptor interaction	IL1R2, CXCR4, CCL2, PPBP	7.21E-03
Transcriptional misregulation in cancer	HOXA10, IL1R2, SIX1, MMP9	7.21E-03
Extracellular matrix organization	COL1A2, MFAP2, NID1, FMOD	1.08E-03
Apoptosis Modulation and Signaling	IL1R2, CASP4, TNFRSF11B	6.23E-03
SRF and miRs in Smooth Muscle Differentiation and Proliferation	NKX2-5	4.42E-02

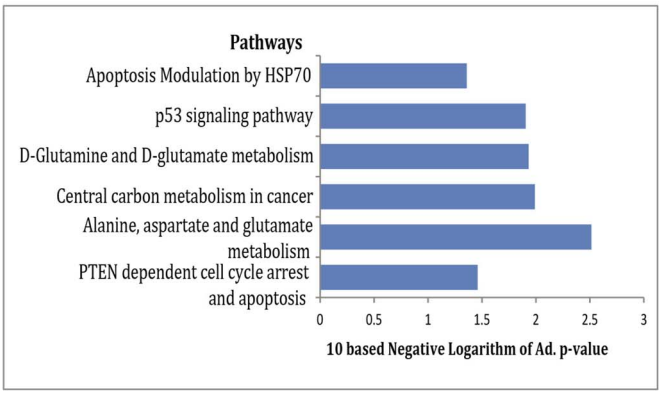


(c) Pathways associated with significantly common differentially expressed genes of the HD with Glioblastoma.

FIG. 3. Enriched signaling pathways common between GBM and each CNS disorder (a) AD, (b) ED, (c) HD, (d) MS and (e) PD.

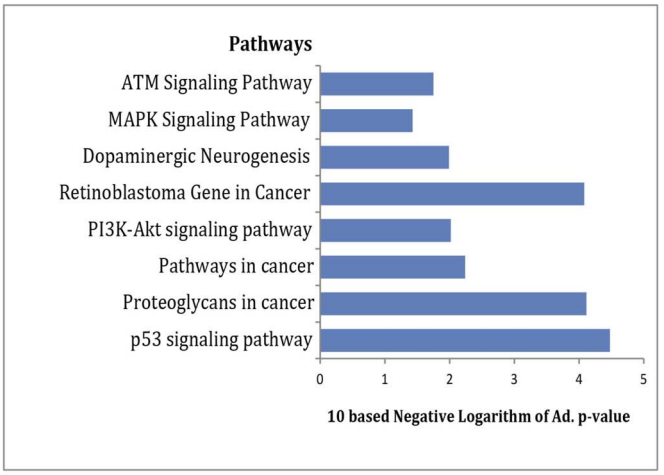
the survival curve is used to indicate the difference between altered versus an unaltered group of genes of the survival pattern. The survival curve comparing altered and unaltered groups for each significant gene that are common between GBM and NDs obtained from the PL estimator function is shown as in Figure 5. We have provided two (Figure 5(b) and Figure 5(c)) of three subfigure of Figure 5 as supplementary (S1) to reduce redundancy.

Pathway	Genes in the pathway	Adjusted p-value
PTEN dependent cell cycle arrest and apoptosis	PTEN	3.47E-02
Alanine, aspartate and glutamate metabolism	GLS2, RIMKLA	3.06E-03
Central carbon metabolism in cancer	GLS2, PTEN	1.02E-02
D-Glutamine and D-glutamate metabolism	GLS2	1.17E-02
p53 signaling pathway	PTEN, FAS	1.24E-02
Apoptosis Modulation by HSP70	FAS	4.37E-02



(d) Pathways associated with significantly common differentially expressed genes of the MS with Glioblastoma.

Pathway	Genes in the pathway	Adjusted p-value
p53 signaling pathway	CCNB2, CCNB1, RRM2, CDK1, FAS	3.34E-05
Proteoglycans in cancer	TGFB2, ITGB3, WNT5A, FAS, HOXD10	7.69E-05
Pathways in cancer	HEYL, EDNRA, TGFB2, WNT5A, FAS	5.74E-03
PI3K-Akt signaling pathway	ITGA4, ITGB3, COL4A6, ITGB8, COL6A3	9.58E-03
Retinoblastoma Gene in Cancer	CCNB2, CCNB1, CDC45, RRM2, CDK1	8.29E-05
Dopaminergic Neurogenesis	SOX2, GLI2	1.02E-02
MAPK Signaling Pathway	MAPK11, TGFB2, PLA2G4A, FAS	3.73E-02
ATM Signaling Pathway	CCNB1, CDK1	1.77E-02



(e) Pathways associated with significantly common differentially expressed genes of the PD with Glioblastoma.

FIG. 3. continued.

From Figure 5, we see that 54 genes are identified as the most significant genes which are associated with the survival of GBM patients. Among them, gene number 1–15 between AD and GBM, 16–18 between ED and GBM, 19–28 between HD and GBM, 29–38 between MS and GBM and 39–54 between PD and GBM associated with survival of the GBM patients. In the survival curve of each gene, the altered expression of genes indicates that a patient is less likely to survive compared to a person in a non-altered group. From both univariate and multivariate analyses, we observe that the P-value lower than 0.05 indicates the significant difference between altered versus unaltered group of genes associated with GBM patient survival. Moreover, the determination of the joint role of important clinical and genetic factors have important entailment in identifying influential genes that are associated with the survival of GBM patients. In Figure 5, it is also noted that the normal gene expression is indicated by the red line and the altered gene expression is indicated by the blue line in the survival curve as shown in Figure 5.

Results comparison with Gold benchmark databases and literature

We have developed a pipeline framework based on bioinformatics and machine learning tools to identify biomarkers for which CNS disorders influence the progression of GBM. As far as we know, there are no such methods to identify the effect of CNS disorders in GBM progression. Thus, to compare our results obtained from our pipeline based on well-established bioinformatics and machine learning tools, there is no other criterion against which it can be compared. There are many other cross-checking tests for the obtained results but they are too time-consuming for the results of the test to be clinically useful, and those tests may not be possible to perform on a patient. A benchmarking study consists of a robust and comprehensive evaluation of the results of well-established bioinformatics and machine learning algorithms. Benchmarking studies aim to rigorously compare the results of different methods with gold-standard benchmark datasets to determine the strengths of results from proposed methods to provide recommendations regarding suitable choices of analysis methods. Gold standard benchmarking refers to the criteria by which scientific evidence is evaluated and use accepted standards that we can look to as an accurate and reliable reference for the comparison of our results. In clinical bioinformatics and medicine, we often refer to dbGap, OMIM and OMIM Expanded as the gold standard database for comparing our results. These studies use gold-standard data sets to serve as a reference and well-defined scoring metrics to assess the performance and accuracy of our results from a variety of analytical tasks and data types. To assess our results,

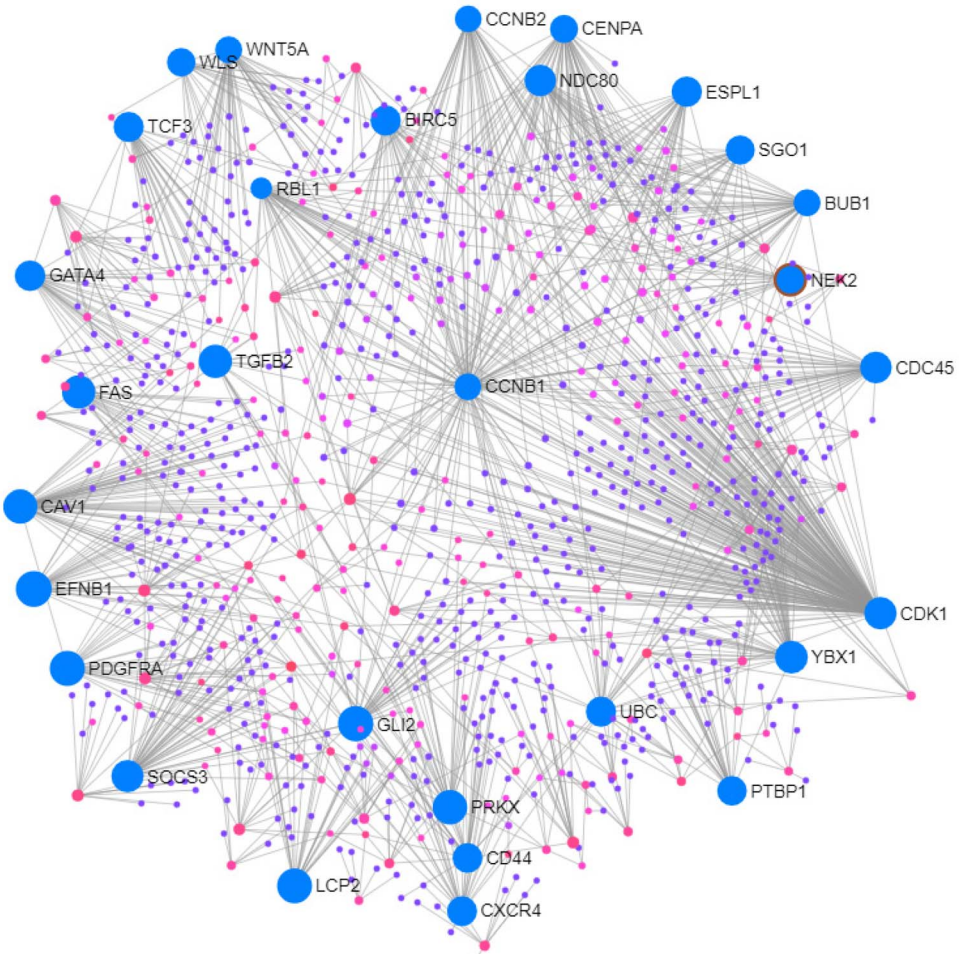


FIG. 4. Hub proteins from protein-protein interactions analysis.

we used the dysregulated DEGs (identified by the pipeline of the selected CNS disorders) as input to the online gene set enrichment analyzing tool, EnrichR [73], and obtained enriched genes and their associated disease from the stated three gold-standard benchmark databases. The statistical parameter, a P-value of threshold <0.05 was chosen for several steps of analysis to select genes associated with each cancer.

Among the several diseases with their enriched genes, we choose only cancer-related diseases for constructing the disease-gene association network. Our selected cancer such as GBM is enriched with it's associated genes from the mentioned three gold-standard benchmark databases. Our enriched disease-gene association network is constructed using the list of cancer which is shown in Figure 6. This comparison confirms that significant genes we identified in CNS disorders as having known disease associations. Systematic benchmarking of our results based on gold standard data outperforms the existing results in a standardized way for this disease comorbidity which strengthens confidence in the data we obtain using our computational methodology. Also, we observe that our identified genes are associated with GBM progression by studying the literature to identify which of these genes have been clinically used as biomarkers for GBM. Specifically:

- Reifenberger G et. al [87] showed the CD44 gene associated with the incidence of GBM.
- Backes C et. al [88] previously found a link between FAM20A, RRM2 genes and GBM incidence.
- Crespo I et. al [89] and Appin CL et al [90] identified a link between the PDGFRA gene and GBM.
- Crespo I et. al [89] identified the ASCL1 gene associated with the progression of GBM.
- Cheng W et. al [91] showed the FCGR2B gene associated with the progression of GBM.
- Crespo I et al [89] and Cheng W et. al [91] identified the MMP9 gene associated with the progression of GBM.
- Li QJ et. al [92] and Dunn GP et. al [93] have shown the PTEN gene associated with the progression of GBM.
- Bo L et. al [94] identified the CDK1 gene associated with GBM incidence.
- Cheng F et. al [95] identified the SOX2 gene associated with the progression of GBM.

Therefore, it suggested that CNS's disorders may influence the progression of GBM.

DISCUSSION

This study aimed to establish the competency of the pipeline of bioinformatics and machine learning approach to identify the effects of NDs on GBM cancer comorbidity along with survival analysis and to gain deeper insights into GBM associated biomarkers that are essential for prognosis, diagnosis and treatment by investigating genetics and molecular mechanisms at multistage. To achieve our objectives, we applied this approach

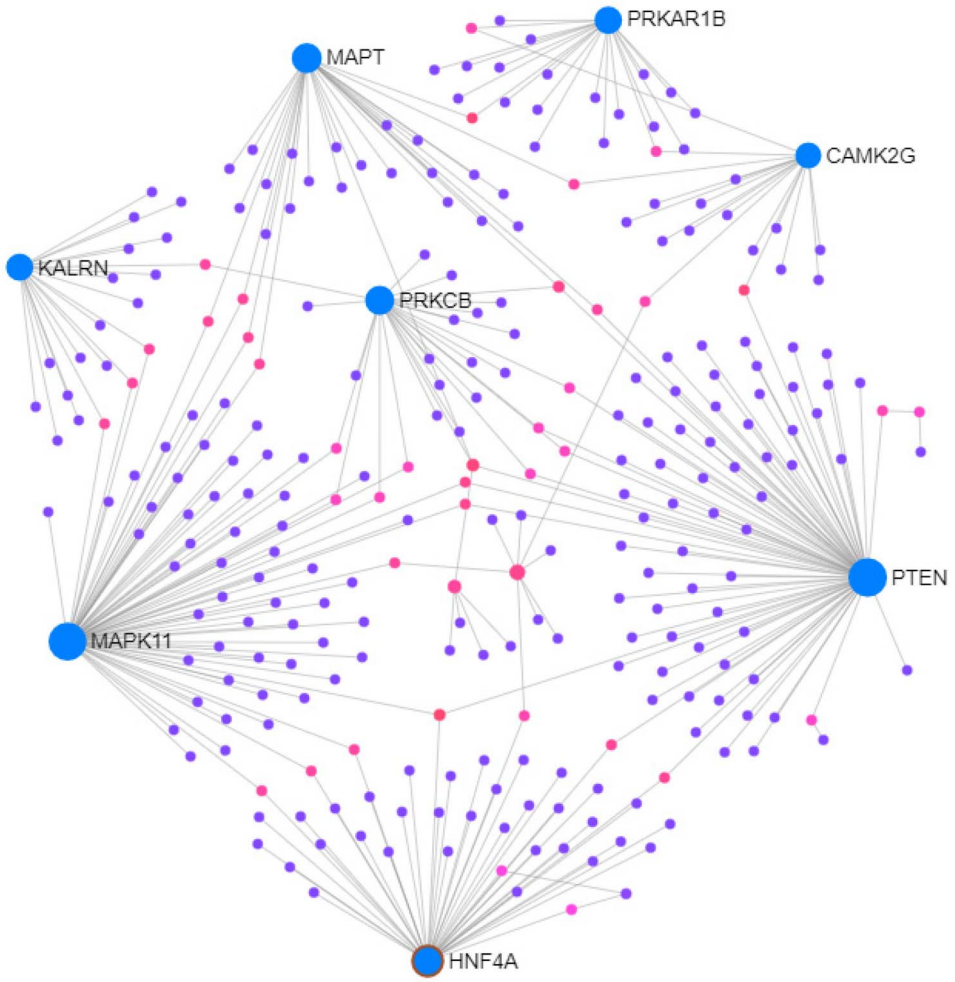
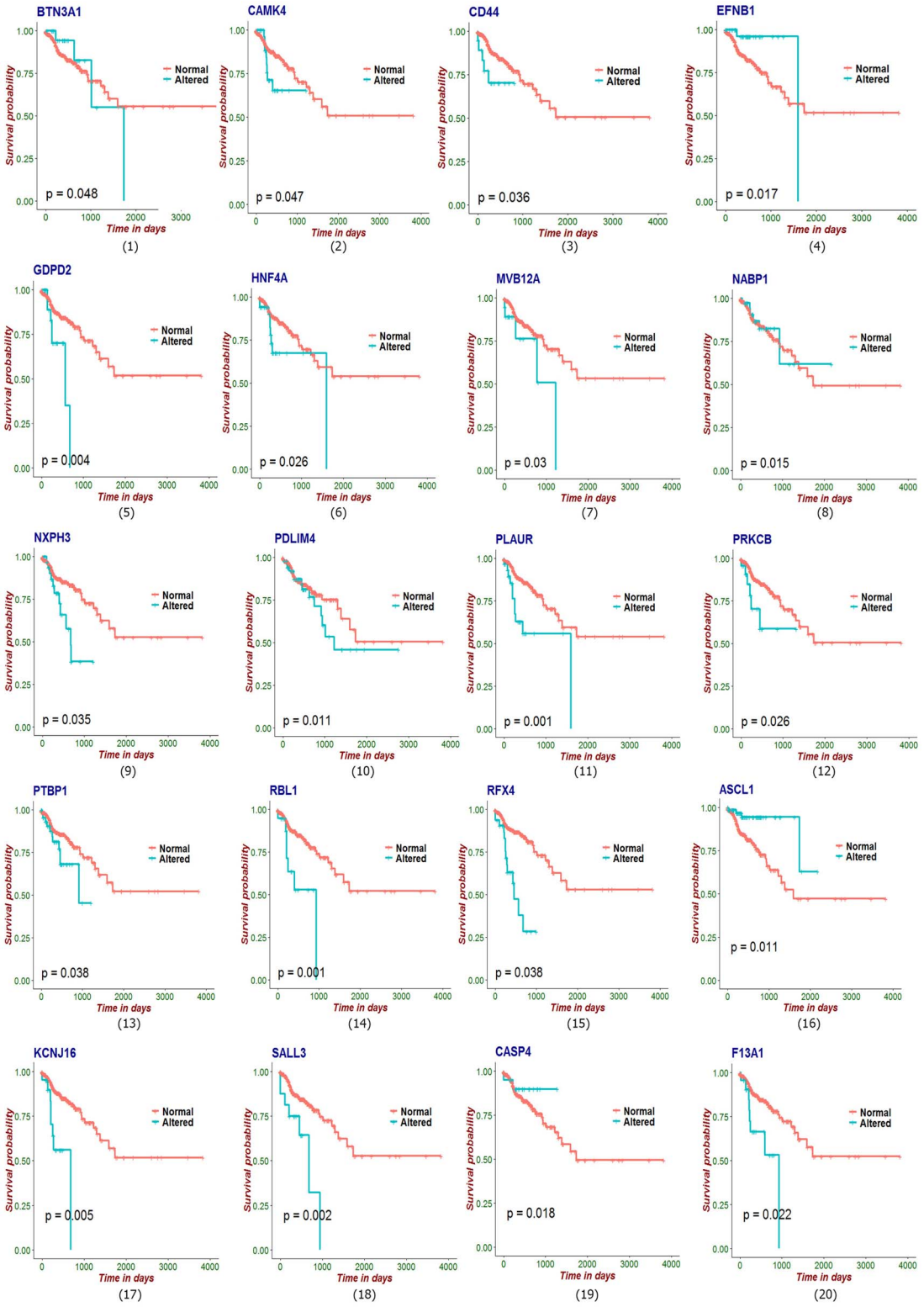


FIG. 4. continued.

to examining microarray and RNA-seq gene expression data from NDs and GBM each with control individual samples to identify the comorbidity progression of GBM.

Our proposed pipeline identified the significant DEGs between NDs and GBM suggesting NDs may influence the behavior of GBM. Based on upregulated and downregulated DEGs common between NDs and GBM, we constructed two separate up and down disease-gene association (diseasome) networks which show strong evidence that NDs can indeed interact with GBM as shown in Figure 2. The functional annotation via enrichment analysis provides new knowledge about complex diseases by determining common pathways between NDs and GBM. Our identified molecular and ontological pathways revealed by the dysregulated genes of GBM with NDs disorder is shown in Figure 3 and Table 2. Our identified molecular and ontological pathways indicated how NDs may affect GBM cancer and influence the progression of GBM. The multi-omics analysis involving protein–protein interaction revealed the hub genes that are shown in Figure 4 involved in key signaling pathways regulating important molecular pathways in the pathobiology of GBM and suggest that NDs affect GBM cancer. NDs of the CNS, for example, AD shares 49 significant DEGs with GBM, ED shares 11 significant DEGs with GBM, HD shares 26 significant DEGs with GBM, MS shares 47 significant DEGs with GBM and PD shares 51 significant DEGs with GBM as shown in Figure 2. These data provide significant proof to support the involvement of these genes with the GBM pathogenesis and the noxious effect of NDs/CNS disorder on cancer. To evaluate the prognosis of the identified biomarkers genes, we widely employed Cox Proportional Hazard Model to predict the most significant biomarkers genes associated with GBM progression and also associated with the survival of GBM cancer patients. Both univariate and multivariate regression analysis and the log-rank test was performed to determine the significant difference between alter and unaltered groups for each gene by estimating the survival function based on the PL estimation procedure. We found 15 DEGs between AD and GBM, 3 between ED and GBM, 10 between HD and GBM, 10 between MS and GBM and 16 between PD and GBM as biomarker genes that affect the survival of GBM which is tabulated in Table 3, Table 4, Table 5, Table 6 and Table 7, respectively, and survival curve for these biomarker genes are shown in Figure 5. A key reason for the use of the Cox PH model is that it relies on fewer assumptions compared to parametric models. The fundamental assumption in this model is the proportionality of the hazard function. Note that the Cox PH model assumes that the hazard ratio of two people is independent of time. A total of 54 prognostic genes were identified that affect the survival of the GBM patients by using the Cox PH model.



(a) Survival curve for the most significant DEGs common between CNS disorders and GBM

FIG. 5. The survival curve comparing altered and unaltered groups for each significant gene that are common between NDs (AD, ED, HD, MS and PD) and the GBM is shown in fig (a), (b) and (c). The gene numbers in the figure are divided into five groups where AD and GBM shares gene number 1–15, ED and GBM shares 16–18, HD and GBM share 19–28, MS and GBM share 29–38 and PD and GBM shares 39–54. Here, the cyan-colored lines in the survival curve indicate data from individuals with altered gene expression and the red indicates data from individuals with normal (i.e. unaffected) gene expression.

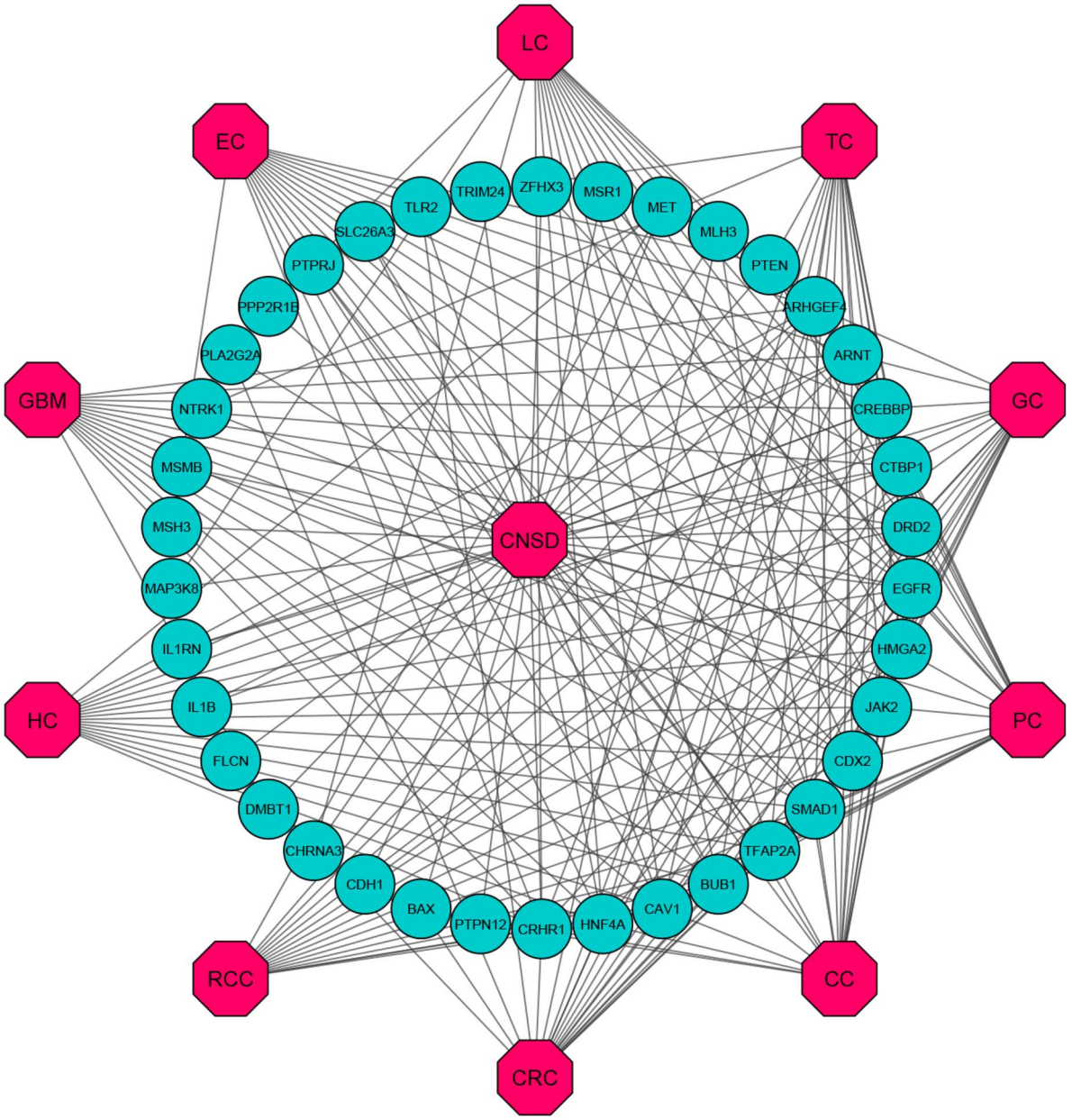


FIG. 6. Potential targets validating results using gold-standard databases. The octagon-shaped red color nodes indicate disease node and blue color nodes indicate genes. The red-colored center node is the acronym of central nervous system disorders which includes AD, ED, HD, MS and PD. The acronym CC for colon cancer, GC for gastric cancer, EC for endometrial cancer, GBM for glioblastoma, LC for lung cancer, CRC for colorectal cancer, HC for hepatocellular carcinoma, RCC for renal cell carcinoma and PC for prostate cancer.

In general, the results of our analyses show that the NDs among the CNS disorders share dysregulated genes and molecular mechanisms, and an individual with CNS disorders (AD, ED, HD, MS and PD) is at increased risk of GBM development. Our results are useful for both biomedical research and health care utilization. Concerning biomedical research, our findings revealed why individuals with NDs are more vulnerable to GBM development. Our findings help us to acquire a better understanding of pathogenesis, which could lead to better prevention, diagnosis and treatment. The identification of potential biomarkers for GBM also opens new ways to think about treatment strategies for this disease comorbidity and the prognostic. We observe that our identified altered genes associated with GBM involve several molecular and ontological pathways that are related to cell proliferation, survival, migration and angiogenesis as complex pathogenesis. Our methods are thus able to identify the role of GBM comorbidity with NDs and lead to better means of predicting the survival of the GBM patients more precisely and helpful for the development of the health care system. We verified these potential biomarker genes with the gold standard benchmark databases and literature review as shown in Figure 6 and we confirm that our identified biomarker genes are strongly associated with the GBM progression and associated with the survival of GBM patients.

Table 3. The regression results of the Cox PH model for the DEGs that are common between AD and GBM. The parameter legends of the table are β for estimated coefficients, HR for hazard ratios.

Gene Symbol	Univariate			Multivariate								
	Overexpressed			Underexpressed			Overexpressed			Underexpressed		
	β	HR	P-value	β	HR	P-value	β	HR	P-value	β	HR	P-value
NABP1	-8.26E-02	9.21E-01	8.47E-01	1.06E+00	2.88E+00	1.45E-02	-5.92E-01	5.53E-01	4.88E-01	2.72E+00	1.52E+01	3.05E-03
GDPD2	1.17E+00	3.22E+00	4.16E-03	5.83E-01	1.79E+00	8.27E-02	5.28E-01	1.70E+00	4.66E-01	4.26E-01	1.53E+00	4.63E-01
PTBP1	6.93E-01	2.00E+00	3.85E-02	-3.95E-01	6.73E-01	5.08E-01	7.10E-01	2.03E+00	2.45E-01	-2.67E-01	7.66E-01	7.57E-01
HNF4A	2.53E-01	1.29E+00	7.26E-01	8.06E-01	2.24E+00	2.59E-02	8.48E-01	2.34E+00	3.74E-01	4.44E-01	1.56E+00	4.75E-01
BTN3A1	-6.52E-02	9.37E-01	9.00E-01	-1.03E+00	3.58E-01	1.53E-01	-8.76E-01	4.16E-01	2.79E-01	-1.98E+00	1.38E-01	4.82E-02
CAMK4	6.98E-01	2.01E+00	8.33E-02	7.58E-01	2.13E+00	4.74E-02	8.64E-01	2.37E+00	1.39E-01	2.79E-01	1.32E+00	6.16E-01
NXPH3	1.01E+00	2.75E+00	2.58E-03	1.53E+00	4.60E+00	3.53E-02	-1.95E-01	8.23E-01	7.46E-01	2.30E+00	9.94E+00	8.69E-03
PRKCB	-5.44E-01	5.80E-01	2.94E-01	9.57E-01	2.60E+00	2.65E-02	-1.59E+00	2.04E-01	3.45E-02	5.66E-01	1.76E+00	5.14E-01
PLAUR	-9.21E-01	3.98E-01	2.00E-01	1.15E+00	3.15E+00	5.58E-04	-1.40E+00	2.48E-01	2.47E-01	1.61E+00	5.02E+00	6.04E-02
CD44	8.96E-01	2.45E+00	5.63E-02	1.73E-01	1.19E+00	6.17E-01	1.76E+00	5.79E+00	3.59E-02	2.54E-01	1.29E+00	7.25E-01
EFNB1	-1.17E+00	3.10E-01	1.03E-01	2.02E-01	1.22E+00	6.68E-01	-2.27E+00	1.03E-01	1.73E-02	9.37E-01	2.55E+00	1.76E-01
RBL1	-2.32E-01	7.93E-01	6.54E-01	1.29E+00	3.62E+00	1.45E-03	4.21E-01	1.52E+00	5.43E-01	2.10E+00	8.15E+00	1.11E-03
MVB12A	-5.64E-01	5.69E-01	1.89E-01	1.02E+00	2.76E+00	2.98E-02	-4.78E-01	6.20E-01	4.01E-01	1.23E+00	3.43E+00	8.22E-02
PDLIM4	-7.44E-02	9.28E-01	9.41E-01	2.95E-01	1.34E+00	3.33E-01	6.01E-01	1.83E+00	6.90E-01	1.28E+00	3.60E+00	1.08E-02
APOL6	-1.27E+00	2.81E-01	2.09E-01	3.44E-01	1.41E+00	2.82E-01	-5.72E-01	5.65E-01	6.29E-01	7.98E-01	2.22E+00	1.18E-01

Table 4. The regression results of the Cox PH model for the DEGs that are common between ED and GBM. The parameter legends of the table are β for estimated coefficients, HR for hazard ratios.

Gene Symbol	Univariate			Multivariate								
	Overexpressed			Underexpressed			Overexpressed			Underexpressed		
	β	HR	P-value	β	HR	P-value	β	HR	P-value	β	HR	P-value
KCNJ16	1.50E+00	4.47E+00	1.11E-04	-4.46E-01	6.40E-01	3.60E-01	1.26E+00	3.53E+00	4.69E-03	-3.40E-01	7.12E-01	6.92E-01
ASCL1	3.32E-01	1.39E+00	7.42E-01	-1.32E+00	2.67E-01	1.07E-02	5.44E-01	1.72E+00	6.09E-01	-1.20E+00	3.03E-01	3.79E-02
SALL3	1.23E+00	3.43E+00	2.29E-03	-9.34E-01	3.93E-01	4.94E-02	8.80E-01	2.41E+00	5.31E-02	-6.25E-01	5.35E-01	3.17E-01

Table 5. The regression results of the Cox PH model for the DEGs that are common between HD and GBM. The parameter legends of the table are β for estimated coefficients, HR for hazard ratios.

Gene Symbol	Univariate			Multivariate								
	Overexpressed			Underexpressed			Overexpressed			Underexpressed		
	β	HR	P-value	β	HR	P-value	β	HR	P-value	β	HR	P-value
GYPC	-9.89E-01	3.72E-01	1.69E-01	1.45E+00	4.26E+00	6.73E-06	-3.07E+00	4.65E-02	1.80E-02	1.28E+00	3.61E+00	4.60E-02
TFEC	-1.15E-01	8.92E-01	7.74E-01	-1.71E+01	3.92E-08	9.96E-01	-2.05E+00	1.29E-01	1.14E-02	-1.72E+01	3.44E-08	9.96E-01
PITX1	1.04E-02	1.01E+00	9.86E-01	1.14E+00	3.11E+00	2.76E-03	6.25E-01	1.87E+00	4.47E-01	1.42E+00	4.12E+00	1.96E-02
CASP4	-5.99E-01	5.50E-01	4.05E-01	-2.43E-01	7.84E-01	5.75E-01	-3.20E+00	4.09E-02	1.85E-02	-9.14E-01	4.01E-01	3.08E-01
HOXB9	-3.66E-01	6.94E-01	4.80E-01	1.18E+00	3.27E+00	6.01E-03	-3.97E-01	6.72E-01	5.64E-01	1.59E+00	4.89E+00	3.30E-02
FOXD2	-7.68E-01	4.64E-01	2.86E-01	1.12E+00	3.06E+00	3.21E-02	-4.90E-01	6.13E-01	5.56E-01	1.91E-01	1.21E+00	8.50E-01
OTP	6.64E-02	1.07E+00	9.11E-01	-5.04E-02	9.51E-01	9.01E-01	-2.67E+00	6.91E-02	3.38E-03	1.10E+00	3.01E+00	1.07E-01
F13A1	1.16E+00	3.18E+00	2.45E-03	-6.92E-01	5.01E-01	2.44E-01	2.29E+00	9.83E+00	2.24E-02	-2.72E-01	7.62E-01	7.79E-01
MEOX2	8.50E-01	2.34E+00	2.00E-02	-1.40E+01	8.27E-07	9.96E-01	2.29E+00	9.83E+00	5.91E-05	-2.19E+01	3.16E-10	9.99E-01
ID3	7.36E-01	2.09E+00	1.56E-01	-3.66E-01	6.93E-01	4.80E-01	-5.24E-01	5.92E-01	7.27E-01	-2.16E+00	1.15E-01	1.68E-02

Table 6. The regression results of the Cox PH model for the DEGs that are common between MS and GBM. The parameter legends of the table are β for estimated coefficients, HR for hazard ratios.

Gene Symbol	Univariate			Multivariate								
	Overexpressed			Underexpressed			Overexpressed			Underexpressed		
	β	HR	P-value	β	HR	P-value	β	HR	P-value	β	HR	P-value
MAPT	9.61E-01	2.62E+00	7.73E-03	-7.18E-01	4.88E-01	7.30E-02	2.15E+00	8.56E+00	2.39E-02	-4.23E-01	6.55E-01	6.40E-01
RIMKLA	8.35E-02	1.09E+00	8.58E-01	4.67E-01	1.60E+00	1.99E-01	2.64E+00	1.40E+01	1.70E-02	-4.47E-01	6.40E-01	6.51E-01
TEF	5.50E-01	1.73E+00	1.73E-01	-7.55E-01	4.70E-01	2.88E-02	-2.42E-01	7.85E-01	8.19E-01	-1.46E+00	2.33E-01	3.78E-02
PYGO1	1.11E+00	3.05E+00	1.08E-02	-3.98E-02	9.61E-01	9.21E-01	6.43E-01	1.90E+00	5.98E-01	1.15E+00	3.15E+00	2.34E-01
KALRN	8.27E-01	2.29E+00	2.91E-02	2.50E-01	1.28E+00	4.09E-01	-9.22E-01	3.98E-01	3.27E-01	2.68E+00	1.47E+01	9.11E-03
GLIS3	1.10E+00	2.99E+00	4.14E-03	-3.69E-01	6.91E-01	2.42E-01	-2.40E-01	7.86E-01	8.01E-01	3.11E-01	1.37E+00	6.85E-01
RAB27B	6.77E-01	1.97E+00	6.02E-02	4.06E-01	1.50E+00	2.21E-01	1.63E+00	5.11E+00	8.79E-02	2.22E+00	9.23E+00	1.03E-02
PTPRT	6.63E-01	1.94E+00	5.41E-02	1.83E-01	1.20E+00	5.57E-01	-7.64E-01	4.66E-01	3.57E-01	2.03E+00	7.60E+00	2.21E-02
FOCAD	-1.69E-01	8.45E-01	8.67E-01	5.16E-01	1.68E+00	4.70E-02	-4.76E+00	8.56E-03	1.13E-01	1.02E+00	2.77E+00	8.26E-02
WDR49	1.03E+00	2.80E+00	1.56E-03	-2.31E-02	9.77E-01	9.47E-01	3.13E+00	2.29E+01	8.18E-04	-1.81E+00	1.64E-01	9.78E-02

Table 7. The regression results of the Cox PH model for the DEGs that are common between PD and GBM. The parameter legends of the table are β for estimated coefficients, HR for hazard ratios.

Gene Symbol	Univariate						Multivariate					
	Overexpressed			Underexpressed			Overexpressed			Underexpressed		
	β	HR	P-value	β	HR	P-value	β	HR	P-value	β	HR	P-value
TSKU	7.20E-01	2.06E+00	9.52E-02	7.06E-01	2.03E+00	7.85E-02	9.00E-01	2.46E+00	2.74E-01	1.76E+00	5.84E+00	2.31E-02
CD58	-1.27E+00	2.79E-01	2.06E-01	1.55E-01	1.17E+00	7.03E-01	-3.37E+00	3.44E-02	4.11E-02	-1.18E+00	3.09E-01	2.46E-01
EDA2R	4.21E-01	1.52E+00	4.18E-01	1.04E+00	2.82E+00	6.48E-03	1.83E+00	6.21E+00	4.71E-02	9.07E-01	2.48E+00	2.47E-01
ITGB8	1.31E+00	3.72E+00	6.61E-05	-1.06E-01	8.99E-01	9.17E-01	1.53E+00	4.60E+00	2.25E-02	4.44E+00	8.49E+01	1.79E-02
E2F8	-1.16E+00	3.14E-01	2.51E-01	7.43E-01	2.10E+00	5.23E-02	-1.53E-01	8.58E-01	9.25E-01	2.55E+00	1.28E+01	1.94E-02
SLC18A1	-3.50E-01	7.05E-01	6.27E-01	1 1.066420e+00	2.90E+00	8.22E-03	1.84E+00	1.59E-01	1.95E-01	-3.11E-01	7.33E-01	7.20E-01
NBL1	-5.07E-02	9.51E-01	9.22E-01	1.12E+00	3.06E+00	5.39E-03	-9.96E-01	3.70E-01	2.46E-01	3.87E-01	1.47E+00	6.39E-01
POU4F1	3.10E-01	1.36E+00	3.19E-01	1.21E+00	3.34E+00	4.38E-02	7.03E-01	2.02E+00	1.57E-01	3.81E+00	4.51E+01	1.57E-03
SLC4A7	2.89E-01	1.33E+00	5.01E-01	4.71E-01	1.60E+00	3.13E-01	1.99E+00	7.29E+00	4.41E-02	1.82E-01	1.20E+00	8.65E-01
GABBR1	7.65E-01	2.15E+00	2.71E-02	1.49E+00	4.41E+00	1.62E-03	9.59E-01	2.61E+00	1.18E-01	3.43E+00	3.10E+01	3.37E-03
GLI2	7.40E-01	2.09E+00	8.78E-02	5.16E-01	1.68E+00	1.74E-01	2.05E+00	7.75E+00	1.96E-02	2.67E+00	1.45E+01	2.85E-03
HOXB3	1.18E+00	3.26E+00	6.04E-03	9.69E-01	2.64E+00	5.49E-03	2.56E+00	1.29E+01	7.16E-03	3.99E-01	1.49E+00	5.93E-01
SIT1	-3.93E-01	6.75E-01	5.07E-01	1.27E+00	3.54E+00	3.07E-05	2.31E+00	1.01E+01	4.54E-02	8.16E-01	2.26E+00	3.03E-01
ITGA4	7.14E-01	2.04E+00	4.97E-02	-8.06E-01	4.47E-01	2.67E-01	3.47E-01	1.41E+00	6.29E-01	-1.40E+00	2.48E-01	5.24E-01
FOXM1	-4.62E-01	6.30E-01	4.35E-01	1.06E+00	2.88E+00	2.24E-03	-1.31E+00	2.70E-01	2.63E-01	9.41E-01	2.56E+00	3.26E-01
SLC16A4	3.24E-01	1.38E+00	5.33E-01	4 6.155572e-04	1.00E+00	9.99E-01	3.47E-01	7.07E-01	7.62E-01	-3.00E+00	5.00E-02	1.54E-02

CONCLUSION

In conclusion, the co-occurrence of direct and inverse GBM progression in patients with NDs/CNS disorders led us to identify biological and nonbiological relations. Our findings provide an in-depth understanding of the pathobiology of GBM cancer and why individuals with NDs are relatively vulnerable to GBM development. In our study, we identified several gene mutations associated with GBM progression but all of them were not significantly related to the survival of GBM patients. For this reason, we expanded our study to identify the joint effect of clinical factors and genes to identify biomarker genes that play roles for the survival of the GBM patients and we identified 54 genes that influence survival. From the clinical science perspective, we suggest that the potential markers we identified in this work could be useful for disease diagnostic and may help in finding useful and novel therapeutic targets. At the same time, our findings will assist in a variety of clinical and diagnostic activities. Medical practitioners have recently begun to consider comorbidity interactions as an important part of treatment modalities, and started to investigate the possibilities of using multi-omics data by bioinformatics approach. Our proposed methodologies can be used for this approach and may detect many disease features, especially those at the earliest time points before the symptoms appear, and help to gain a better understanding of the complex pathogenesis of disease risk phenotypes and the heterogeneity of disease comorbidity. Thus, it could be applied to improve individualized medicine and clinical bioinformatics. Thus, our study may lead to improve health outcomes and reduce diagnostic costs. In this work, we investigated how gene expression, multi-omics, clinical and molecular data can be integrated and analyzed to identify diseasome and disease comorbidity interactions using bioinformatics and machine learning models. In general, we used datasets of low sample size and different cell types for cross disease or comorbidity analysis, so it is possible to miss the genes associated with diseases. Our future direction is to attain greater reproducibility of biological findings in disease comorbidity identification, therapeutic targets/biomarker discovery that will further advance the quality of individualized medicine for GBM patients that also suffer from ND comorbidities.

Key Point

- Epidemiological and clinical studies link the central nervous system (CNS) disorders and suggest that their transcriptomic profiles could have a number of molecular mechanisms in common. Our main objectives are to develop an integrated framework based on the bioinformatics and machine learning model to unravel shared differentially expressed genes(DEGs) and cell signaling pathways that can link CNS disorders and glioblastoma (GBM);
- After analyzing transcriptomic data of selected CNS disorders and GBM, and identifying DEGs employing our framework, disease-gene association network, signaling pathway, gene ontology (GO) analysis as well as hub proteins identifications were performed to predict the function of these DEGs;
- We expanded our study to determine the significant genes that play a significant role in GBM progression and affect the survival of the GBM patients by exploiting clinical and genetic factors using the Cox Proportional Hazard Model and the Kaplan–Meier estimator. Our biomarker genes that influence patient survival can be targets for therapeutic drug development;
- Our findings provide an in-depth understanding of the pathobiology of GBM cancer and why individuals with NDs are relatively vulnerable to GBM development. The identified hub proteins are potential biomarkers which may lead to new GBM therapeutic targets and may play significant roles in signal transduction during the progression of GBM;
- Our proposed approach is generalized and can be used for identifying key genetic and clinical factors of other types of cancers.

Acknowledgments

We thank CAS-TWAS Presidents Fellowship for providing doctoral support to the first author (M.H.R.) by Fellowship No. 2016CTF014.

Author Contributions

M.H.R., S.P. and M.A.M. conceived and designed the study; M.H.R. performed the computational analyses and wrote the draft manuscript; H.K.R. helped in the preparation of the tables and figures. J.M.W.Q., X.H. and C.C. were involved in the preparation of the important intellectual content and critical revision; S.P. supervised the whole study. All authors approved the final version for submission.

References

1. Valderas JM, Starfield B, Sibbald B, *et al*. Defining comorbidity: implications for understanding health and health services. *Ann Fam Med* 2009 Jul 1; **7**(4): 357–63.
2. García-Azorín D, Martínez-Pías E, Trigo J, *et al*. Neurological comorbidity is a predictor of death in Covid-19 disease: a cohort study on 576 patients. *Front Neurol* 2020 Jul 7; **11**:781.
3. Catalá-López F, Suárez-Pinilla M, Suárez-Pinilla P, *et al*. Inverse and direct cancer comorbidity in people with central nervous system disorders: a meta-analysis of cancer incidence in 577,013 participants of 50 observational studies. *Psychother Psychosom* 2014; **83**(2): 89–105.
4. Ibáñez K, Boullosa C, Tabarés-Seisdedos R, *et al*. Molecular evidence for the inverse comorbidity between central nervous system disorders and cancers detected by transcriptomic meta-analyses. *PLoS Genet* 2014; **10**(2): p.e1004173.
5. Plun-Favreau H, Lewis PA, Hardy J, *et al*. Cancer and neurodegeneration: between the devil and the deep blue sea. *PLoS Genet* 2010; **6**(12): p.e1001257.
6. Feng DD, Cai W, Chen X. The associations between Parkinson's disease and cancer: the plot thickens. *Translational neurodegeneration* 2015; **4**(1): p.20.4a.
7. Ong EL, Goldacre R, Goldacre M. Differential risks of cancer types in people with Parkinson's disease: a national record-linkage study. *Eur J Cancer* 2014 Sep 1; **50**(14): 2456–62.
8. Sevenich L. Brain-resident microglia and blood-borne macrophages orchestrate central nervous system inflammation in neurodegenerative disorders and brain cancer. *Front Immunol* 2018 Apr 6; **9**:697.

9. Majd S, Power J, Majd Z. Alzheimer's disease and cancer: when two monsters cannot be together. *Front Neurosci* 2019 Mar 1; **13**:155.
10. Ma LL, Yu JT, Wang HF, Meng XF, Tan CC, Wang C, Tan L (2014) Association between cancer and Alzheimer's disease: systematic review and meta-analysis. *J Alzheimers Dis* **42**(2): 565–573. https://doi.org/10.3233/JAD-14016 8
11. Musicco M, Adorni F, Di Santo S, *et al*. Inverse occurrence of cancer and Alzheimer disease: a population-based incidence study. *Neurology* 2013 Jul 23; **81**(4): 322–8.
12. Shafi O. Inverse relationship between Alzheimer's disease and cancer, and other factors contributing to Alzheimer's disease: a systematic review. *BMC Neurol* 2016 Dec; **16**(1): 1–7.
13. Hammond TC, Xing X, Wang C, *et al*. β-amyloid and tau drive early Alzheimer's disease decline while glucose hypometabolism drives late decline. *Communications biology* 2020 Jul 6; **3**(1): 1–3.
14. Qian Z, Alaa AM, Bellot A, Rashbass J, van der Schaar M. Learning dynamic and personalized comorbidity networks from event data using deep diffusion processes. arXiv preprint arXiv:2001.02585. 2020 Jan 8.
15. Liu T, Ren D, Zhu X, *et al*. Transcriptional signaling pathways inversely regulated in Alzheimer's disease and glioblastoma multiform. *Sci Rep* 2013 Dec 10; **3**:3467.
16. Sánchez-Valle J, Tejero H, Ibáñez K, *et al*. A molecular hypothesis to explain direct and inverse co-morbidities between Alzheimer's disease. *Glioblastoma and Lung cancer Scientific reports* 2017 Jun 30; **7**(1): 1–2.
17. Candido S, Lupo G, Pennisi M, *et al*. The analysis of miRNA expression profiling datasets reveals inverse microRNA patterns in glioblastoma and Alzheimer's disease. *Oncol Rep* 2019 Sep 1; **42**(3): 911–22.
18. Cai S, Li J, Wu Y, *et al*. De novo mutations of TUBB2A cause infantile-onset epilepsy and developmental delay. *J Hum Genet* 2020 Mar; **16**:1–8.
19. Huberfeld G, Vecht CJ. Seizures and gliomas—towards a single therapeutic approach. *Nat Rev Neurol* 2016 Apr; **12**(4): 204.
20. Liang S, Zhang J, Zhang S, *et al*. Epilepsy in adults with supratentorial glioblastoma: incidence and influence factors and prophylaxis in 184 patients. *PloS one* 2016 Jul 20; **11**(7): e0158206.
21. Snijders TJ, Berendsen S, Seute T, *et al*. Glioma-associated epilepsy: toward mechanism-based treatment. *Transl Cancer Res* 2017 Mar 1; **6**:S337.
22. Berntsson SG, Malmer B, Bondy ML, *et al*. Tumor-associated epilepsy and glioma: are there common genetic pathways? *Acta Oncol* 2009 Jan 1; **48**(7): 955–63.
23. Dabrowska M, Ciolak A, Kozlowska E, *et al*. Generation of new isogenic models of Huntington's disease using CRISPR-Cas9 technology. *Int J Mol Sci* 2020 Jan; **21**(5): 1854.
24. Chandra S, Suarez-Cedeno G, Stimming E. Malignant gliomas in patients with Huntington's disease. In: *MOVEMENT DISORDERS 2018 Oct 1 (Vol. 33, pp. S379-S379). 111 RIVER ST, HOBOKEN 07030-5774*. NJ USA: WILEY.
25. Sørensen SA, Fenger K, Olsen JH. Significantly lower incidence of cancer among patients with Huntington disease: an apoptotic effect of an expanded polyglutamine tract? *Cancer: Interdisciplinary International Journal of the American Cancer Society* 1999 Oct 1; **86**(7): 1342–6.
26. Chandra S, Suarez-Cedeno G, Stimming E. Malignant gliomas in patients with Huntington's disease [abstract]. *Mov Disord* 2018; **33**(suppl 2). https://www.mdsabstracts.org/abstract/malignant-gliomas-in-patients-with-huntingtons-disease/ Accessed October 18, 2020.
27. Rana HK, Akhtar MR, Ahmed MB, *et al*. Genetic effects of welding fumes on the progression of neurodegenerative diseases. *Neurotoxicology* 2019 Mar 1; **71**:93–101.
28. Pulido-Valdeolivas I, Andorrà M, Gómez-Andrés D, *et al*. Retinal and brain damage during multiple sclerosis course: inflammatory activity is a key factor in the first 5 years. *Sci Rep* 2020 Aug 7; **10**(1): 1–1.
29. Plantone D, Renna R, Sbardella E, *et al*. Concurrence of multiple sclerosis and brain tumors. *Front Neurol* 2015 Mar 4; **6**:40.
30. Islam T, Rahman MR, Karim MR, *et al*. Detection of multiple sclerosis using blood and brain cells transcript profiles: insights from comprehensive bioinformatics approach. *Informatics in Medicine Unlocked* 2019; **16**:100201.
31. Tan EK, Chao YX, West A, *et al*. Parkinson disease and the immune system—associations, mechanisms and therapeutics. *Nat Rev Neurol* 2020 Apr; **24**:1–6.
32. Ye R, Shen T, Jiang Y, *et al*. The relationship between parkinson disease and brain tumor: a meta-analysis. *PloS one* 2016 Oct 20; **11**(10): e0164388.
33. Yamamoto T, Fukaya C, Obuchi T, *et al*. Glioblastoma multiforme developed during chronic deep brain stimulation for Parkinson disease. *Stereotact Funct Neurosurg* 2016; **94**(5): 320–5.
34. Mencke P, Hanss Z, Boussaad I, *et al*. Bidirectional relation between Parkinsons disease and glioblastoma multiforme. *Front Neurol* 2020; **11**:898.
35. Ye R, Shen T, Jiang Y, *et al*. The relationship between parkinson disease and brain tumor: a meta-analysis. *PloS one*. 2016 Oct 20; **11**(10): e0164388.
36. Savaskan NE, Seufert S, Hauke J, *et al*. Dissection of mitogenic and neurodegenerative actions of cystine and glutamate in malignant gliomas. *Oncogene* 2011; **30**(1): 43.
37. Louis DN, Perry A, Reifenberger G, *et al*. The 2016 World Health Organization classification of tumors of the central nervous system: a summary. *Acta Neuropathol* 2016; **131**(6): 803–20.
38. Furnari FB, Fenton T, Bachoo RM, *et al*. Malignant astrocytic glioma: genetics, biology, and paths to treatment. *Genes Dev* 2007; **21**(21): 2683–710.
39. Festuccia C, Biordi AL, Tombolini V, *et al*. Targeted molecular therapy in glioblastoma. *J Oncol* 2020 Jan; **14**:2020.
40. Rajaratnam V, Islam MM, Yang M, *et al*. Glioblastoma: pathogenesis and current status of chemotherapy and other novel treatments. *Cancer* 2020 Apr; **12**(4): 937.
41. Silantyev AS, Falzone L, Libra M, *et al*. Current and future trends on diagnosis and prognosis of glioblastoma: from molecular biology to proteomics. *Cell* 2019 Aug; **8**(8): 863.
42. Tuaeva NO, Falzone L, Porozov YB, *et al*. Translational application of circulating DNA in oncology: review of the last decades achievements. *Cell* 2019 Oct; **8**(10): 1251.
43. Liu J, Yu J, Shen H, *et al*. Mass spectrometric analysis of cerebrospinal fluid protein for glioma and its clinical application. *Contemporary Oncology* 2014; **18**(2): 100.
44. Meier F, Geyer PE, Virreira Winter S, *et al*. BoxCar acquisition method enables single-shot proteomics at a depth of 10,000 proteins in 100 minutes. *Nat Methods* 2018 Jun; **15**:440–8.
45. Miyauchi E, Furuta T, Ohtsuki S, *et al*. Identification of blood biomarkers in glioblastoma by SWATH mass spectrometry and quantitative targeted absolute proteomics. *PloS one*. 2018 Mar 7; **13**(3): e0193799.

46. Palmirotta R, Lovero D, Cafforio P, *et al*. Liquid biopsy of cancer: a multimodal diagnostic tool in clinical oncology. *Therapeutic advances in medical oncology* 2018 Aug; **10**: 1758835918794630.
47. Ladomersky E, Scholtens DM, Kocherginsky M, *et al*. The coincidence between increasing age, immunosuppression, and the incidence of patients with glioblastoma. *Front Pharmacol* 2019 Mar 27; **10**:200.
48. Squillaro T, Schettino C, Sampaolo S, *et al*. Adult onset brain tumors and neurodegeneration: are polyphenols protective? *J Cell Physiol* 2018; **233**(5): 3955–67.
49. Sevenich L. Brain-resident microglia and blood-borne macrophages orchestrate central nervous system inflammation in neurodegenerative disorders and brain cancer. *Front Immunol* 2018 Apr 6; **9**:697.
50. Savaskan E, Nicolai ZF, Broggini T, *et al*. Neurodegeneration in the brain tumor microenvironment: glutamate in the limelight. *Curr Neuropharmacol* 2015; **13**(2): 258–65.
51. Meldrum BS. Glutamate as a neurotransmitter in the brain: review of physiology and pathology. *J Nutr* 2000; **130**(4): 1007S–15S.
52. Ahmed T, Begum M. Association between gene expression, clinical factors and survival in patients with breast cancer. *Journal of Biomedical Analytics* 2018 Feb 1; **1**(1): 1–4.
53. Yu Q, Hsu Y. Asymptotic normality of the product-limit-estimator. *Journal of Nonparametric Statistics* 2016 Oct 1; **28**(4): 802–12.
54. Barrett T, SE, P, *et al*. NCBI geo: archive for functional genomics data sets—update. *Nucleic Acids Res* 2012; **41**(D1): D991–5.
55. Brazma A, Parkinson H, Sarkans U, *et al*. ArrayExpress—a public repository for microarray gene expression data at the EBI. *Nucleic Acids Res* 2003; **31**(1): 68–71.
56. Blalock, Eric M and Buechel, Heather M and Popovic, Jelena and Geddes, James W and Landfield, Philip W. Microarray analyses of laser-captured hippocampus reveal distinct gray and white matter signatures associated with incipient Alzheimer's disease. *J Chem Neuroanat*; **42**(2): 118–26, Oct, 2011. PMID: 21756998.
57. Niesen CE, Xu J, Fan X, *et al*. Transcriptomic profiling of human peritumoral neocortex tissues revealed genes possibly involved in tumor-induced epilepsy. *PLoS One* 2013; **8**(2) e56077: PMID: 23418513.
58. Labadorf A, Hoss AG, Lagomarsino V, *et al*. RNA sequence analysis of human Huntington disease brain reveals an extensive increase in inflammatory and developmental gene expression. *PLoS One* 2015; **10**(12) e0143563: PMID: 26636579.
59. Lieury A, Chanal M, Androdias G, *et al*. Tissue remodeling in periplaque regions of multiple sclerosis spinal cord lesions. *Glia* 2014; **62**(10): 1645–58, Oct PMID: 24910450.
60. Lewandowski, Nicole M and Ju, Shulin and Verbitsky, Miguel and Ross, Barbara and Geddie, Melissa L and Rockenstein, Edward and Adame, Anthony and Muhammad, Alim and Vonsattel, Jean Paul and Ringe, Dagmar and others. Polyamine pathway contributes to the pathogenesis of Parkinson disease. *Proc Natl Acad Sci U S A*; **107**(39): 16970–5, Sep 28, 2010. PMID: 20837543.
61. Gill BJ, Pisapia DJ, Malone HR, *et al*. MRI-localized biopsies reveal subtype-specific differences in molecular and cellular composition at the margins of glioblastoma. *Proc Natl Acad Sci U S A* 2014 Aug 26; **111**(34): 12550–5 PMID: 25114226.
62. Rana HK, Akhtar MR, Islam MB, *et al*. Machine learning and bioinformatics models to identify pathways that mediate influences of welding fumes on cancer progression. *Sci Rep* 2020 Feb 17; **10**(1): 1–5.
63. Cerami E, Gao J, Dogrusoz U, *et al*. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. 2012;401–4.
64. Hoadley KA, Yau C, Hinoue T, *et al*. Cell-of-origin patterns dominate the molecular classification of 10,000 tumors from 33 types of cancer. *Cell* 2018; **173**(2): 291–304.
65. Rahman MH, Peng S, Hu X, *et al*. 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 Public Health* 2020 Jan; **17**(3): 1035.
66. Liang J, Lv X, Lu C, *et al*. Prognostic factors of patients with Gliomas-an analysis on 335 patients with glioblastoma and other forms of Gliomas. *BMC Cancer* 2020 Dec; **20**(1): 1–7.
67. Rahman MH, Peng S, Hu X, *et al*. Bioinformatics methodologies to identify interactions between type 2 diabetes and neurological comorbidities. *IEEE Access* 2019 Dec 16; **7**:183948–70.
68. Ritchie ME, Phipson B, Wu D, *et al*. Limma powers differential expression analyses for rna-sequencing and microarray studies. *Nucleic Acids Res* 2015; **43**(7): e47–7.
69. Anders, Simon, and Wolfgang Huber. " Differential expression of RNA-Seq data at the gene level-the DESeq package." Heidelberg, Germany: European Molecular Biology Laboratory (EMBL) (2012).
70. Moni MA, Lio' P. Genetic profiling and comorbidities of zika infection. *The J infectious diseases* 2017; **216**:703–12.
71. Rana HK, Akhtar MR, Islam MB, *et al*. Genetic effects of welding fumes on the development of respiratory system diseases. *Comput Biol Med* 2019 May 1; **108**:142–9.
72. Moni MA, Rana HK, Islam MB, *et al*. A computational approach to identify blood cell-expressed Parkinson's disease biomarkers that are coordinately expressed in brain tissue. *Comput Biol Med* 2019 Oct 1; **113**:103385.
73. Kuleshov MV, Jones MR, Rouillard AD, *et al*. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic acids research, 2016* **44**(W1): W90–7.
74. Kanehisa M, Goto S, Sato Y, *et al*. Kegg for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res* 2011; **40**(D1): D109–14.
75. Croft D, O'kelly G, Wu G, *et al*. Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Res* 2010; **39**:D691–7.
76. Slenter DN, Kutmon M, Hanspers K, *et al*. Wiki pathways: a multifaceted pathway database bridging metabolomics to other omics research. *Nucleic Acids Res* 2017; **46**(D1): D661–7.
77. Nishimura D. BioCarta. *Biotech Software & Internet Report* 2001; **2**(3): 117–20. doi: 10.1089/152791601750294344.
78. Gene Ontology Consortium. Gene ontology consortium: going forward. *Nucleic Acids Res* 2014; **43**(D1): D1049–56.
79. Szklarczyk D, Morris JH, Cook H, *et al*. The string database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res* p. gkw9372016.
80. Xia J, Gill EE, Hancock RE. Networkanalyst for statistical, visual and network-based meta-analysis of gene expression data. *Nat Protoc* 2015; **10**(6): 823.
81. Moni MA, Liò P. How to build personalized multi-omics comorbidity profiles. *Frontiers in cell and developmental biology* 2015 Jun 24; **3**:28.
82. Hossain MA, Islam SM, Quinn JM, *et al*. Machine learning and bioinformatics models to identify gene expression patterns

of ovarian cancer associated with disease progression and mortality. *J Biomed Inform* 2019 Dec 1; **100**:103313.
83. Peng G, Luo L, Siu H, *et al*. Gene and pathway-based analysis: second wave of genome-wide association studies. *Nature Precedings* 2008 Jul 14;1.
84. Jin L, *et al*. Pathway-based analysis tools for complex diseases: a review. *Genomics Proteomics Bioinformatics* 2014; **12**:210–20.
85. Beisswanger E, Lee V, Kim JJ, *et al*. Gene regulation ontology (GRO): design principles and use cases. *InMIE* 2008; Jan (pp. 9-14).
86. Shannon P, Markiel A, Ozier O, *et al*. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003 Nov 1; **13**(11): 2498–504.
87. Reifenberger G, Wirsching HG, Knobbe-Thomsen CB, *et al*. Advances in the molecular genetics of gliomas—implications for classification and therapy. *Nat Rev Clin Oncol* 2017 Jul; **14**(7): 434.
88. Backes C, Harz C, Fischer U, *et al*. New insights into the genetics of glioblastoma multiforme by familial exome sequencing. *Oncotarget* 2015 Mar 20; **6**(8): 5918.
89. Crespo I, Vital AL, Gonzalez-Tablas M, *et al*. Molecular and genomic alterations in glioblastoma multiforme. *Am J Pathol* 2015 Jul 1; **185**(7): 1820–33.
90. Appin CL, Brat DJ. Molecular genetics of gliomas. *The Cancer Journal* 2014 Jan 1; **20**(1): 66–72.
91. Cheng W, Ren X, Zhang C, *et al*. Bioinformatic profiling identifies an immune-related risk signature for glioblastoma. *Neurology* 2016 Jun 14; **86**(24): 2226–34.
92. Li QJ, Cai JQ, Liu CY. Evolving molecular genetics of glioblastoma. *Chin Med J (Engl)* 2016 Feb 20; **129**(4): 464.
93. Dunn GP, Rinne ML, Wykosky J, *et al*. Emerging insights into the molecular and cellular basis of glioblastoma. *Genes Dev* 2012 Apr 15; **26**(8): 756–84.
94. Bo L, Wei B, Li C, *et al*. Identification of potential key genes associated with glioblastoma based on the gene expression profile. *Oncol Lett* 2017 Aug 1; **14**(2): 2045–52.
95. Cheng F, Guo D. MET in glioma: signaling pathways and targeted therapies. *J Exp Clin Cancer Res* 2019 Dec 1; **38**(1): 270.