

A Predictive Framework Integrating Network Toxicology and Machine Learning Elucidates 6PPD-Quinone-Induced Glioblastoma Pathogenesis

Yuan Xu

The Affiliated Hospital of Qingdao University

Yu Yao

The Affiliated Hospital of Qingdao University

Junlei Huang

Peking University People's Hospital

Haojun Zhang

The Affiliated Hospital of Qingdao University

Guiming Zhang

zhangguiming9@126.com

The Affiliated Hospital of Qingdao University

Research Article

Keywords: 6PPD-Quinone, bioinformatics, glioblastoma, machine learning, molecular docking, environmental carcinogenesis, neural signaling pathways

Posted Date: October 14th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11193435/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Objective

This article aims to systematically explore the molecular mechanism of glioblastoma induced by 6PPD-Q.

Methods

Differential expression analysis and weighted gene co-expression network analysis were conducted on five independent GBM transcriptome datasets in the NCBI GEO database. It is used to identify the function-related gene modules related to the pathogenesis of GBM. This paper employs three complementary computational methods, to predict the possible protein targets of 6PPD-Q. This paper adopts machine learning algorithms to identify the core diagnostic genes.

Results

Firstly, we integrated differential expression analysis with WGCNA and identified 1,885 highly reliable genes related to GBM. Based on target prediction, we identified 2,623 potential 6PPD-Q protein targets. Among them, 225 genes have a very prominent overlap with GBM-related genes. Through machine learning analysis, we identified five core genes with special diagnostic performance, namely SCN2B, VIPR1, PAK1, MAP2K1 and SYNJ1. The AUC value of this integrated model in the validation queue reached 0.957, and its prediction accuracy is relatively high. From the analysis of SHAP interpretability, it can be found that MAP2K1 is the most influential predictor, and there are complex nonlinear relationships among core genes. Molecular docking simulations have provided structural evidence that can support the direct interaction between 6 PPD-Q and the target protein. The vina score ranges from -5.9 to -8.2, and they have a strong specific binding affinity.

Conclusion

This study provides computational evidence that 6PPD-Q may also promote the onset of glioblastoma by targeting specific genes and disrupting key signaling pathways involved in neurodevelopment and carcinogenic transformation. The study identified five core regulatory genes with special diagnostic functions. It also combines the molecular docking verification of direct binding interactions, which provides brand-new insights into the potential molecular mechanisms of the neurotoxicity and carcinogenicity of 6PPD-Q.

Introduction

Glioblastoma is the most common primary malignant brain tumor in adults, accounting for approximately 45% to 50% of all primary malignant brain tumors. Its incidence rate begins to increase after the age of 40 and peaks between the ages of 75 and 84 ¹. According to the classification of brain tumors by the World Health Organization (WHO), GBM is classified as grade 4, which is the most malignant type of astrocytoma. It is highly invasive, grows at an extremely fast rate, and often invades the surrounding brain tissue, making it one of the most lethal brain tumors ². Currently, the standard treatment methods for GBM include surgical tumor removal, radiotherapy combined with temozolomide for chemotherapy³. Although in the past ten years, there have been certain advancements in surgery, radiotherapy and chemotherapy, but the overall survival probability of GBM patients is still very low, with an average survival time of only 12 to 15 months¹. The causes of GBM involve complex interactions among genetic alterations, epigenetic modifications, and environmental exposure. These factors jointly promote the generation and development of tumors ⁴. Due to its heterogeneity, highly invasive disease progression, and resistance to existing treatment methods, GBM has always been a major clinical challenge. Although risk factors such as ionizing radiation, certain genetic syndromes and family history have now been identified, the impact of widespread environmental pollutants on the causes of GBM has not been fully understood. This is currently an area where active research is being conducted⁵.

Tire wear particles enter the aquatic environment via stormwater runoff and can release highly toxic pollutants with teratogenic and mutagenic effects. One chemical released from tire wear particles, N-(1,3-dimethylbutyl)-N'-phenyl-p-phenylenediamine-quinone (6PPD-quinone, 6PPD-Q), has been reported as the second most acutely toxic chemical to aquatic species ever identified⁶. Its precursor, N-(1,3-dimethylbutyl)-N'-phenyl-p-phenylenediamine (6PPD), is an antioxidant widely used in rubber tires and readily transforms into 6PPD-Q in the presence of ozone⁷. 6PPD-Q has recently gained significant attention as an emerging environmental contaminant. It was initially identified as the causative agent behind acute mortality in coho salmon exposed to urban stormwater runoff and has since been detected in various environmental media, including airborne particulate matter, soil, and water systems^{8,9}. The widespread use of tires and the consequent leaching of 6PPD and its transformation products into the environment suggest potential broad human exposure pathways, particularly in urban settings. The chemical stability and lipophilicity of 6PPD-Q raise concerns about its potential for bioaccumulation and tissue-specific toxicity.

6PPD-Q can rapidly penetrate the blood-brain barrier within such a short period of 0.5 hours. Its structure is quite similar to that of other known neurotoxic quinones. Its molecular weight is relatively low, and it has moderate lipophilicity. All these circumstances indicate that 6PPD-Q can indeed enter neural tissue^{10,11}. Previous studies have shown that various quinone compounds can induce oxidative stress, cause mitochondrial dysfunction, and damage the DNA of nerve cells. These mechanisms are also related to carcinogenesis. However, although 6PPD-Q has these worrying characteristics, However, the possible role that 6ppD-Q may play in the development of brain tumors and in glioblastoma has not been studied at all up to now.

Previous studies have shown that various quinone compounds can induce oxidative stress, cause mitochondrial dysfunction, and damage the DNA of nerve cells. These mechanisms are actually related to carcinogenesis^{12,13}. However, although 6PPD-Q has these worrying characteristics, it may play a role in the development of brain tumors. Its role in the development of glioblastoma has not been studied at all up to now.

The emergence of network toxicology and computational systems biology methods has provided a very powerful framework for uncovering the complex interactions between toxins and hosts. These methods, by integrating multi-omics datasets, dynamic network modeling, and computational structure-activity analysis, can systematically identify molecular targets and pathways that are disturbed by environmental exposure¹⁴. The integration of machine learning algorithms also enables us to prioritize key hub molecules and develop predictive models for toxicity assessment.

To fill the key knowledge gap regarding the possible glioma effects of 6PPD-Q, our study this time adopted a comprehensive triangular system toxicology approach, which consists of three main parts. The first part is to integrate multi-omics data and identify the possible targets of 6PPD-Q. To reconstruct the network of the pathogenesis of glioblastoma. The second part is to determine the priority of key hub molecules by conducting topological and functional enrichment analysis and using machine learning. The third part is to combine the molecules, verify the target binding at the atomic level, and describe the thermodynamic characteristics of the interaction by clarifying the possible multi-scale carcinogenic network of 6PPD-Q. This article aims to identify the nodes where drugs can be used for chemophobic and precise treatment of glioblastoma, thereby addressing a key unsolved need in environmental neuro-oncology.

Materials and Methods

Acquisition and Processing of GBM Transcriptomic Data

This article systematically sorted out five independent GBM transcriptome datasets from the NCBI Comprehensive Gene expression database (GSE12657, GSE42656, GSE50161, GSE67089 and GSE86574). The selection of these datasets was based on certain specific criteria, which included sample size, requiring each dataset to have at least 20 samples, and platform consistency. It should be based on the Affymetrix or Illumina platform. Additionally, the availability of the raw data and whether the dataset contains tumor and normal control samples must also be taken into consideration. These five datasets contain a total of 342 GBM samples and 44 normal brain tissue controls. Among them, GSE12657 and GSE42656 were designated as discovery cohorts with 98 samples, while the remaining three datasets (GSE50161, GSE67089, GSE86574) were independent validation cohorts with 288 samples.

Considering that the gene expression data in public databases usually contain some non-biological confounding factors, such as unknown batch effects or heterogeneity among individuals, which may blur

the true disease-related signals, we first adopt the method of alternative variable analysis to reduce such interference situations. We wanted to visually assess the corrective impact of SVA on the data and explore the natural clustering of the samples based on the overall gene expression pattern. Therefore, we conducted principal component analysis before and after SVA adjustment. With the combined strategy of SVA and PCA, we effectively controlled the data quality. It was also confirmed that the adjusted data could robustly reflect the transcriptomic characteristics related to the disease status.

Differential Gene Expression Analysis

Based on this purified dataset, this article uses limma software to conduct differential expression analysis on GBM tissues and normal brain controls, that is, to see what differences there are in gene expression between them. When identifying differentially expressed genes, a relatively strict threshold was adopted, requiring the FDR correction p value to be less than 0.05. Moreover, the log₂FC should be greater than 0.585, where 0.585 corresponds to a 1.5 times change. Subsequently, this paper uses the ggplot2 software package to visualize the analysis results. For instance, a volcano plot is used to display the distribution and statistical magnitude of expression changes, allowing us to more intuitively observe the distribution and magnitude of gene expression changes. Heat maps are also employed to illustrate the expression patterns of different groups, thus enabling a clear understanding of the characteristics of gene expression patterns in different groups.

Weighted Gene Co-expression Network Analysis (WGCNA)

A scale-free co-expression network was constructed using the WGCNA package in R to identify functionally related gene modules associated with GBM pathogenesis. When using the WGCNA package to construct gene co-expression networks, quality control is first performed through sample hierarchical clustering to remove outlier samples. Subsequently, this paper relies on the hierarchical clustering of topological overlap matrices to identify co-expressed modules. Here, the parameters are set to a minimum module size of 60, and the threshold for module merging is set at 0.25. Then, this paper will analyze the correlation between modules and traits. The screening criterion is that the Pearson correlation coefficient between the characteristic genes and phenotypes of the modules should be greater than 0.5. The threshold for greater magnitude is that the p value should be less than 0.05. In this article, hub genes will be screened based on in-mode connectivity, and genes with a kME greater than 0.8 will be retained.

Acquisition of Chemical Components and Targets of 6PPD-Q

6PPD-Q was systematically characterized through multisource database integration to ensure comprehensive target prediction. The canonical SMILES notation [CC(C)CC(C)NC1=CC(=O)C(=CC1=O)NC2=CC=CC=C2] and 2D molecular structure were extracted from the PubChem database.

Target prediction employed a tripartite strategy to maximize coverage and reliability:

ChEMBL Database: We queried the ChEMBL database (<https://www.ebi.ac.uk/chembl/>) for compounds structurally similar to 6PPD-Q. Targets associated with these similar compounds were extracted and considered as potential targets for 6PPD-Q.

SwissTargetPrediction: The canonical SMILES of 6PPD-Q was submitted to the SwissTargetPrediction web server (<https://swisstargetprediction.ch/>), which employs a combination of 2D and 3D similarity measures to predict protein targets based on known ligand-protein interactions.

PharmMapper Server: The 3D structure of 6PPD-Q was energy-minimized using the MMFF94 force field and submitted to the PharmMapper server (<https://www.lilab-ecust.cn/pharmmapper/index.html>) for reverse pharmacophore mapping against the human protein database. The selection criterion for potential targets is Norm Fit > 0.5.

All predicted targets from the three approaches were restricted to the Homo sapiens proteome, and redundant entries were removed to generate a final set of unique potential targets for 6PPD-Q.

Identification of 6PPD-Q-Associated Disease Targets

Intersection analysis between the union set of 6PPD-Q-predicted targets and the GBM-related genes (from DEGs and WGCNA hub genes) was performed to identify core targets potentially involved in 6PPD-Q-induced gliomagenesis. Venn diagrams were constructed to visualize the overlap using the VennDiagram package in R. The resulting overlapping genes were considered high-priority candidates for subsequent functional characterization and machine learning analysis.

Functional Enrichment Analysis

To elucidate the biological functions and pathways associated with the overlapping genes, we performed comprehensive Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses using the clusterProfiler package in R. Results were visualized using bar plots, dot plots, and circos plots to illustrate the relationships between genes and enriched terms.

Machine Learning-Based Core Gene Screening

This paper aims to systematically screen the diagnostic biomarkers related to 6ppd-q in GBM. Therefore, a stratified sampling method is adopted to divide the dataset into training subsets and internal validation subsets. This is done to keep the category distribution stable. This paper constructs a comprehensive multi-algorithm ensemble machine learning prediction framework. This framework integrates ten classic machine learning algorithms, including Lasso regression, Radial Basis Function kernel Support vector machine (SVM), Random Forest (RF), Gradient boosting machine (GBM), XGBoost, linear discriminant

analysis (LDA), Naive Bayes, and ensemble methods. For each algorithm, this paper has developed several models with different hyperparameter Settings. Eventually, a total of 113 prediction models were obtained. This paper relies on 5x cross-validation to optimize the hyperparameters and uses stratified sampling to separate training and validation. The performance of the models is evaluated using the area under the curve, or AUC, as a metric. Subsequently, the stacked ensemble learning method was adopted to integrate the predictions from the best-performing individual models. Those high-confidence models, that is, those achieving an AUC greater than 0.85 during both training and validation, were selected for this ensemble. Then, the features were ranked and analyzed based on their frequency of occurrence in these high-performance models. The genes with relatively high rankings have been identified as the core genes that may be involved in the formation of glioma induced by 6ppd-q.

Model Interpretation Using SHAP Analysis

This study aims to clarify the decision-making mechanism of the optimal prediction model and identify key features. Therefore, a SHAP framework based on R language was used to conduct an interpretability analysis of the model. We utilized the SHAP package to calculate the SHAP value of each feature in every sample from the trained optimal classification model. This value can quantitatively describe the magnitude and direction of the contribution of features to the model output.

Molecular Docking Analysis

To validate the interaction between the key target and 6PPD-Q at the atomic level, we performed molecular docking simulations. To validate the interaction between the key target and 6PPD-Q at the atomic level, we performed molecular docking simulations. The three-dimensional crystal structure of the core target protein was obtained from the Protein Data Bank (PDB) and downloaded in PDB format. Molecular docking was conducted using the online platform CB-Dock2, which can automatically identify potential binding pockets of the protein and perform rapid docking. The docking results were evaluated based on the binding free energy predicted by the Vina scoring function, and the optimal conformation with the highest score was selected for subsequent analysis.

Results

Identification of GBM-Related Target Genes

Multiple transcriptome datasets were successfully integrated, laying a foundation for the accurate identification of GBM-related genes. The comprehensive batch effect correction strategy we adopted effectively addressed the technical differences among different datasets, as evidenced by the significant improvement in data distribution consistency (Fig. 1A-B). Principal component analysis revealed that before correction, The two datasets, GSE12657 and GSE42656, are clearly separated from each other, and there are a large number of batch effects (Fig. 1C). After the coordinated pipeline was achieved, the samples of these two datasets presented an overlapping distribution in the dimensionality reduction

space, which confirmed the success of the integration (Fig. 1D). Such strict preprocessing of the data laid a reliable foundation for the subsequent differential expression and co-expression analysis.

Differential expression analysis was conducted on GBM tissues and the normal brain control group. A total of 1,660 significantly differentially expressed genes that met our strict criteria were identified, with fdr correction $P < 0.05$ and $\log_2\text{FC} > 0.585$. We visualized the analysis results using a volcano map (Fig. 2A). The heat map shows the different expression patterns of these DEGs. It can be seen that there is a significant separation between GBM and the control group, and it also highlights the molecular heterogeneity of glioblastoma (Fig. 2B).

Based on the results of WGCNA, we identified the key gene modules significantly associated with 6PPD-Q exposure. The module-trait relationship analysis showed that the positive correlation between the turquoise module and the black module and the treated trait was the strongest. Their correlation coefficients were 0.73 and 0.71 respectively, and the p values were both less than $1e-16$. Under the exposure of 6PPD-Q, the co-expression changes of genes in these two modules were the most prominent (Fig. 2C-D). An internal analysis of the modules revealed that among these key modules, there was a high degree of consistency between the module membership degree of the treated traits and the gene magnitude. The cor of the turquoise module was 0.73, with a p value of $2.5e-165$, and the cor of the black module was 0.76, with a p value of $4e-22$. This means that in the turquoise and black modules, those genes located at the center of the network, that is, with stronger connectivity, have a closer association between their expression levels and 6PPD-Q exposure. These findings verify that these modules are biologically reliable and can be given priority when exploring the toxicological mechanisms of the network in the future.

Integration of the 1,660 DEGs from conventional differential expression analysis and the 989 turquoise module genes from WGCNA yielded a final, high-confidence set of 1,885 unique GBM-related genes after removing duplicates (Fig. 2E). This integrated approach enhanced the reliability of GBM-associated gene identification by combining both differential expression and functional co-expression criteria.

Identification of 6PPD-Q Target Proteins and Associated Disease Targets

The canonical chemical structure of 6PPD-Q was obtained from the PubChem database (Fig. 3A). Subsequently, target prediction was performed using the ChEMBL, PharmMapper, and SwissTargetPrediction platforms. After data integration and removal of redundant entries, a total of 2,623 unique potential targets for 6PPD-Q were identified (Fig. 3B). The distribution of these targets can be visualized using a Venn diagram. This relatively comprehensive target prediction result shows that 6PPD-Q has the potential to interact with multiple proteins, which is consistent with its structural complexity. After conducting a key cross-analysis of the 2,623 predicted targets of 6PPD-Q and 1,885 GBM-related genes, 225 overlapping genes were found (Fig. 3C). These 225 overlapping genes form a set of core potential molecular targets. 6PPD-Q May participate in the formation process of glioma

through these targets, so they were selected for subsequent functional characterization and machine learning analysis.

Functional Enrichment Analysis of 6PPD-Q-Associated Targets in GBM

To systematically elucidate the biological functions of the potential core targets of 6PPD-Q, we performed Gene Ontology (GO) enrichment analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis. The integrated results clearly demonstrate that the toxic mechanisms of 6PPD-Q are highly concentrated in nervous system functions and cancer-related pathways.

At the biological function level, GO analysis revealed that the core targets were significantly enriched in key neurological processes such as "regulation of membrane potential", "learning or memory", "gliogenesis", "cognition", "regulation of synaptic plasticity" and "potassium ion transport". In terms of cellular components, the targets were mainly located at key neuronal structures, such as the "synaptic membrane", "ion channel complexes", and "postsynaptic specialization" (Fig. 4A). These findings together suggest that 6PPD-Q May impair the electrophysiological signals and synaptic functions of neurons, thereby affecting higher-order neural activities.

At the level of signaling pathways, KEGG analysis provided more detailed insights into the mechanism. In typical neural signaling pathways such as "glutamatergic synapses", "cholinergic synapses", and "retrograde endogenous cannabinoid signaling", there was a very prominent enrichment of the targets, which strongly proved the synaptic functional terms discovered in the graphene oxide analysis. A coherent chain of evidence has been formed, ranging from molecular functions to signaling pathways. The analysis also revealed that multiple pathways closely related to tumorigenesis were highly enriched, including the "PI3K-Akt signaling pathway", "focal adhesion", "Ras signaling pathway", and various solid cancer pathways (Fig. 4B). This finding links the neurotoxic effects of 6PPD-Q with potential carcinogenic risks. Its toxicity in the nervous system may be related to the activation or disruption of key oncogenic signaling networks.

Based on the comprehensive GO and KEGG enrichment analysis, the neurotoxicity of 6PPD-Q is derived from the interaction among multiple processes and pathways, mainly involving the disruption of neuronal synaptic transmission and plasticity, as well as the possible abnormal activation of cancer signaling pathways related to the pathogenesis of glioblastoma. This provides a clear direction and a crucial theoretical basis for studying the specific toxicological mechanism of 6PPD-Q.

Machine Learning-Based Identification of Core Genes in 6PPD-Q-Induced Gliomagenesis

This paper aims to establish a reliable diagnostic model for glioblastoma induced by 6ppd-q and identify the most critical molecular targets. Therefore, a comprehensive machine learning framework was implemented, which integrates multiple algorithms and integration strategies. Based on the expression

profiles of 225 overlapping genes, this paper employed 10 classic machine learning algorithms. A total of 113 prediction models were established.

Model Performance Evaluation demonstrated exceptional predictive capability across both training and validation cohorts (Fig. 5A). The top-performing models included: RF+NaiveBayes (AUC = 0.997 training, 0.995 validation, combined score = 0.957), RF+Ridge (AUC = 0.999 training, 0.995 validation, combined score = 0.953), XGBoost (AUC = 1.000 training, 0.986 validation, combined score = 0.946), Ridge (AUC = 0.998 training, 0.997 validation, combined score = 0.942) and RF+Enet[alpha = 0.4] (AUC = 0.999 training, 0.993 validation, combined score = 0.942). The integrated model performed particularly well in the independent validation cohort (GSE50161, GSE67089 and GSE86574), with an average AUC of 0.934, an accuracy rate of 0.892, and an f1 score of 0.885. This confirms that the diagnostic framework is relatively robust.

Through core gene identification and feature criticality analysis, it can be known that there are five key genes with the highest diagnostic value. They are SCN2B, VIPR1, PAK1, MAP2K1 and SYNJ1 respectively (Fig. 5B). The volcano plot clearly shows that there are significant differences in the expression of these genes between the treatment group and the control group.

To assess the diagnostic potential of these five key genes, we conducted a receiver operating characteristic curve analysis. The analysis results indicated that all genes demonstrated strong discriminative power, with the area under the curve values all exceeding 0.94. It should be mentioned here that the prediction accuracy demonstrated by MAP2K1 and PAK1 is particularly outstanding (Fig. 5C). This indicates that the screened key genes can very reliably distinguish 6PPD-Q exposed samples from control samples, demonstrating strong potential as biomarkers.

Model Interpretation Using SHAP Analysis

Based on SHAP model interpretability analysis, we further revealed the contribution levels and functional patterns of five key genes in predicting 6PPD-Q exposure. SHAP value analysis clearly quantified the importance of each feature, with MAP2K1 contributing most significantly to the model output (mean |SHAP value| = 0.075), followed by PAK1 (0.0575) and SYNJ1 (0.0442), while SCN2B and VIPR1 also showed substantial influence (Fig. 6A).

A bee swarm plot detailed the distribution of SHAP values for each gene and its relationship with gene expression levels. It should be noted that if MAP2K1 and PAK1 have relatively high expression, it is easier to increase the prediction score of the model. This indicates that they are upregulated in 6PPD-Q exposed samples and play a crucial positive role in promoting the classification results. In other words, their high expression has a very crucial positive effect on the classification results. In contrast to this situation, SYNJ1 and SCN2B exhibit more complex patterns. Their low expression states and high expression states respectively have positive and negative impacts on model predictions, which implies that there may be more complex regulatory mechanisms, as shown in Fig. 6B.

In addition, the SHAP-dependent plot reveals possible biological interactions among key genes. The expression levels of MAP2K1 and PAK1 tend to change in a coordinated way. When both genes show high expression, their positive combined effect on model prediction is the strongest. Similarly, there is also a very prominent covariant relationship between MAP2K1 and SYNJ1. These genes may be jointly involved in and regulate the same signaling pathways or cellular processes in the neurotoxic mechanism of 6PPD-Q, as can be seen from Fig. 6C.

Overall, SHAP analysis confirmed from the perspective of machine learning that MAP2K1, PAK1, SYNJ1, SCN2B and VIPR1 are the core genes that respond to 6PPD-Q exposure. It also explained their respective directional effects and potential functional synergies. Such analysis results provide clear priority goals for the subsequent experimental verification and mechanism research, and also offer theoretical basis.

Molecular Docking Validation of 6PPD-Q-Core Gene Interactions

In this paper, to verify the possible direct interaction between 6PPD-Q and the identified core genes, molecular docking simulation was carried out. The results showed that 6PPD-Q had relatively strong binding affinity with the five target proteins, and the binding energy was always lower than -5.0 kcal/mol (Table 1). The molecular interaction was stable and the specificity was also very strong.

Table 1. Binding energies of 6PPD-Q

Ligand	Receptor	Vina score
6PPD-Q	MAP2K1	-7.2
6PPD-Q	PAK1	-7.8
6PPD-Q	SCN2B	-5.9
6PPD-Q	SYNJ1	-8.2
6PPD-Q	VIPR1	-7.7

According to the standards already established in molecular docking research, if the binding energy is less than 0 kilocalories per mole, it indicates that the substance has the ability to bind spontaneously. If the binding energy is less than -5.0 kilocalories per mole, it indicates that the binding affinity is relatively good. If the binding energy is less than -7.0 kilocalories per mole, it indicates that the binding affinity is very strong. As shown in Fig. 7, molecular docking revealed the binding interactions between 6PPD-Q and the receptors. The binding energies of these five complexes all exceeded the threshold that could achieve strong binding affinity. Among them, MAP2K1 demonstrated the most favorable interaction.

Discussion

This research is pioneering. It systematically interprets the possible molecular mechanisms that link the emerging environmental pollutant 6PPD-Q with the pathogenesis of glioblastoma⁷. By adopting an integrated computational strategy, we have coordinated multi-omics data, advanced machine learning, and structural biology simulations. A very notable interaction network has been discovered. Our research results show that 6PPD-Q may promote glioma formation by targeting a specific group of genes, which are particularly crucial for neural function and oncogenic signal transduction. The most important aspect of our research is identifying 225 highly reliable targets that can link 6PPD-Q exposure to GBM. Then, relying on a powerful machine learning framework, a detailed analysis was conducted on the five core diagnostic genes. SHAP interpretability analysis was used to explain their functional roles and mutual interactions. Additionally, molecular docking was employed to structurally verify the binding ability of 6PPD-Q to these targets^{15,16}.

Our work has a key and novel discovery: at the unique intersection of neuro-specific processes and mature oncogenic pathways, there are clearly many potential 6PPD-Q targets^{17,18}. After a comprehensive analysis of GO and KEGG, it is unanimously indicated that there is a dual attack mechanism: 6PPD-Q seems to disrupt fundamental neurophysiological processes, including regulating membrane potential, synaptic plasticity, glial formation and ion channel activity, while also interfering with key cancer driver pathways such as PI3K-Akt, Ras and Focal Adhesion¹⁶. This situation is crucial. It speculates that the neurotoxicity of 6PPD-Q is not only a parallel effect but may also be intrinsically linked to its carcinogenic potential. After the core neuron signals and homeostasis are disrupted, it may create a loose cellular environment or cause initial damage. When combined with the activation of proliferation and survival pathways, this will prompt the transformation of glial cells¹⁴. This provides a computationally derived mechanism hypothesis for the sequence "environmental exposure - neurodysfunction - glioma formation", which goes beyond the known links between quinine drugs and general oxidative stress as well as DNA damage by accurately identifying specific downstream pathways¹⁹.

By comprehensively applying the machine learning framework, which includes 113 different models, it has been proven that this approach can eliminate the complexity of transcriptomic data, identify powerful biomarkers, and the five core genes verified in independent cohorts have excellent diagnostic performance, with all AUCs greater than 0.94²⁰. This highlights their possible role as biomarkers for 6PPD-Q exposure or early detection of related pathological changes. What is even more worth mentioning is that the SHAP interpretability analysis has transformed these biomarkers from mere statistical associations into functional suspects with contextual implications. After quantifying the criticality of the features, it was determined that MAP2K1 and PAK1 are the main driving factors for the model's prediction^{20,21}. This discovery is quite astonishing from a biological perspective. MAP2K1 is the central kinase of the RAS/RAF/MEK/ERK pathway, which is a key basis for cell proliferation and survival that are often dysregulated in cancer, while PAK1 is a key node that integrates Rho GTPases signaling to regulate cytoskeletal dynamics, cell movement and transcription²². Just as the SHAP dependency diagram shows, the synergistic relationship among them indicates that 6PPD-Q exposure may affect

these interrelated growth-promoting and invasion-promoting signaling modules²³. After adding SYNJ1, SCN2B and VIPR1, a more comprehensive picture is depicted. This suggests that there are dysfunctions in synaptic communication, ionic homeostasis and glial signaling in the toxic mechanism^{24,25}.

Molecular docking simulation provides a crucial structural basis for our hypothesis. Predictions show that 6PPD-Q has always had a strong binding affinity with the protein products of all five core genes, which provides a reasonable physical basis for the observed computational association²⁶. The binding with MAP2K1 is particularly favorable, and it may be the main target. It may cause abnormal activation of the entire MAPK signal cascade. Although the results obtained from these computer simulations are highly enlightening, they cannot serve as the ultimate evidence of interactions or functional regulation in biological systems^{27,28}. They are the key starting point for future experimental work.

This study also has some shortcomings. It is mainly based on calculation, and the causal relationship between 6PPD-Q exposure and glioma formation needs to be verified through rigorous experiments using both in vitro and in vivo models²⁹. It remains to be determined what exact functional outcomes will occur when 6PPD-Q binds to targets such as MAP2K1 or pak1, whether it activates their functions or inhibits them. Additionally, although the model organisms used in source transcriptomics studies have certain correlations, they may not fully reflect the complexity of human glioma formation^{30,31}.

Overall, our comprehensive analysis has constructed a very convincing new theory regarding the potential gliogenicity risk of 6PPD-Q. We have mapped out a coherent molecular network, prioritizing those highly credible core genes with specific diagnostic capabilities and pointing out the synergistic relationships among them. It also provides structural evidence for direct targeting^{31,32}. This work has transformed 6PPD-Q from a common poison into a molecule that may cause specific and potential damage to key carcinogenic and neurodevelopmental pathways. The core genes identified, along with the synergistic genes for MAP2K1 and PAK1, are promising research objects for future mechanotoxicological studies. It may also be a biomarker for environmental monitoring and risk assessment of the population exposed to this ubiquitous environmental pollutant.

Although this study has provided new insights into the possible molecular mechanisms of glioblastoma induced by 6PPD-Q, there are some limitations that need to be noted. Firstly, most of these research results are obtained through computational analysis and publicly available transcriptome data, lacking direct experiments to prove them. Although the results of molecular docking can indicate the possibility of binding between 6PPD-Q and the core target, these predictions based on calculation still need to be confirmed through in vitro and in vivo experiments to determine the causal relationship between them. The transcriptome datasets used in this study have significant heterogeneity and were obtained based on different platforms. Although strict methods were adopted to correct batch effects, there might still be some unprocessed confounding factors. Although the machine learning models demonstrated relatively high prediction accuracy, they were trained on existing glioblastoma data. Perhaps it is impossible to fully capture the dynamic changes and heterogeneity of glioblastoma induced by 6PPD-Q. Because the data was mainly obtained from human tissue samples, this study did not take into account

the possible differences in toxicological responses among different species, which limited the generalization of the research results in the in vivo model system. In future research, it is necessary to integrate multi-omics analyses in cell and animal models exposed to 6PPD-Q. This is crucial for verifying these computational predictions and clarifying the functional pathways in the context of biology.

Conclusion

Based on comprehensive computational analysis, this article determined that 6PPD-quinone might be involved in the pathogenesis of glioblastoma by disrupting key molecular networks. We identified 225 potential targets related to 6PPD-Q exposure and GBM, and extracted five core diagnostic genes using machine learning methods. The five core diagnostic genes are MAP2K1, PAK1, SYNJ1, SCN2B and VIPR1 respectively. These genes exhibit rather unique diagnostic properties and are involved in neural functions and carcinogenic pathways. The identified genes can be regarded as promising candidates for future experimental verification and potential environmental risk assessment biomarkers. This research work highlights the value of computational methods in the field of environmental toxicology. It also emphasized the necessity of conducting experimental verification on these prediction results.

Declarations

Acknowledgements

Not applicable.

Authors' contributions

G.Z.: Conceptualization, Data curation, Formal analysis, Investigation, Validation, Supervision; Y.X.: Investigation, Data Curation, Writing—Original Draft; Y.Y., J.L. and H.J.: Writing—Original Draft, Writing—Review and Editing, Supervision.

Funding

This study was partly funded by the BEIJING DADI MEDICAL CHARITY FOUNDATION (DDYL-A-KT-20250619-0089).

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Conflicts of Interest

The authors declare no competing interests.

Clinical trial number

Not applicable.

References

1. Schaff, L. R. & Mellinghoff, I. K. Glioblastoma and Other Primary Brain Malignancies in Adults. *Jama* **329** (2023). <https://doi.org:10.1001/jama.2023.0023>
2. Pace, A. et al. European Association for Neuro-Oncology (EANO) guidelines for palliative care in adults with glioma. *The Lancet Oncology* **18**, e330-e340 (2017). [https://doi.org:10.1016/s1470-2045\(17\)30345-5](https://doi.org:10.1016/s1470-2045(17)30345-5)
3. Lan, Z., Li, X. & Zhang, X. Glioblastoma: An Update in Pathology, Molecular Mechanisms and Biomarkers. *International Journal of Molecular Sciences* **25** (2024). <https://doi.org:10.3390/ijms25053040>
4. Esemien, Y. et al. Molecular Pathogenesis of Glioblastoma in Adults and Future Perspectives: A Systematic Review. *International Journal of Molecular Sciences* **23** (2022). <https://doi.org:10.3390/ijms23052607>
5. Pouyan, A. et al. Glioblastoma multiforme: insights into pathogenesis, key signaling pathways, and therapeutic strategies. *Molecular Cancer* **24** (2025). <https://doi.org:10.1186/s12943-025-02267-0>
6. Chen, L. et al. Solar-light-activated periodate for degradation and detoxification of highly toxic 6PPD-quinone at environmental levels. *Nature Water* **2**, 453-463 (2024). <https://doi.org:10.1038/s44221-024-00236-3>
7. Li, G. et al. Respiratory toxicity mechanism of 6PPD and 6PPD-quinone: An integrated study based on network toxicology and molecular docking. *Ecotoxicology and Environmental Safety* **301** (2025). <https://doi.org:10.1016/j.ecoenv.2025.118494>
8. Zhang, S.-Y. et al. 6PPD and its metabolite 6PPDQ induce different developmental toxicities and phenotypes in embryonic zebrafish. *Journal of Hazardous Materials* **455** (2023). <https://doi.org:10.1016/j.jhazmat.2023.131601>
9. Fang, J. et al. 6PPD-quinone exposure induces neuronal mitochondrial dysfunction to exacerbate Lewy neurites formation induced by α -synuclein preformed fibrils seeding. *Journal of Hazardous Materials* **465** (2024). <https://doi.org:10.1016/j.jhazmat.2023.133312>
10. Li, Y. et al. A Review of N-(1,3-Dimethylbutyl)-N'-phenyl-p-Phenylenediamine (6PPD) and Its Derivative 6PPD-Quinone in the Environment. *Toxics* **12** (2024). <https://doi.org:10.3390/toxics12060394>

11. Huang, Z. *et al.* Protective role of ghrelin against 6PPD-quinone-induced neurotoxicity in zebrafish larvae (*Danio rerio*) via the GHSR pathway. *Ecotoxicology and Environmental Safety* **285** (2024). <https://doi.org/10.1016/j.ecoenv.2024.117031>
12. Greer, J. B., Dalsky, E. M., Lane, R. F. & Hansen, J. D. Tire-Derived Transformation Product 6PPD-Quinone Induces Mortality and Transcriptionally Disrupts Vascular Permeability Pathways in Developing Coho Salmon. *Environmental Science & Technology* **57**, 10940-10950 (2023). <https://doi.org/10.1021/acs.est.3c01040>
13. Jana, S. *et al.* Mitochondrial dysfunction mediated by quinone oxidation products of dopamine: Implications in dopamine cytotoxicity and pathogenesis of Parkinson's disease. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* **1812**, 663-673 (2011). <https://doi.org/10.1016/j.bbadis.2011.02.013>
14. Cecarini, V. *et al.* Crosstalk between the ubiquitin–proteasome system and autophagy in a human cellular model of Alzheimer's disease. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* **1822**, 1741-1751 (2012). <https://doi.org/10.1016/j.bbadis.2012.07.015>
15. Gao, J. *et al.* Integrating machine learning and molecular docking to decipher the molecular network of aflatoxin B1-induced hepatocellular carcinoma. *International Journal of Surgery* **111**, 4539-4549 (2025). <https://doi.org/10.1097/js9.0000000000002455>
16. Liao, X.-L., Zhou, J.-M., Wang, Y., Chen, Z.-F. & Cai, Z. Network pharmacology and transcriptomics reveal androgen receptor as a potential protein target for 6PPD-quinone. *Science of The Total Environment* **957** (2024). <https://doi.org/10.1016/j.scitotenv.2024.177678>
17. Wang, Y. *et al.* Mechanisms underlying the therapeutic effects of Qingfei Yin in treating acute lung injury based on GEO datasets, network pharmacology and molecular docking. *Computers in Biology and Medicine* **145** (2022). <https://doi.org/10.1016/j.combiomed.2022.105454>
18. Yu, H. *et al.* Exposure to 6PPD-Q induces dysfunctions of ovarian granulosa cells: Its potential role in PCOS. *Journal of Hazardous Materials* **486** (2025). <https://doi.org/10.1016/j.jhazmat.2024.137037>
19. Guo, B., Jiang, X., Zhu, L. & He, X. Exploring the Diagnostic Potential of Core Targets of 6PPD and Its Metabolite 6PPD-Q in Cardiovascular Diseases: An Integrated Analysis Based on Network Toxicology, Molecular Docking, and In Vitro Validation. *Journal of Applied Toxicology* (2025). <https://doi.org/10.1002/jat.4892>
20. Crampon, K., Giorkallos, A., Deldossi, M., Baud, S. & Steffanel, L. A. Machine-learning methods for ligand–protein molecular docking. *Drug Discovery Today* **27**, 151-164 (2022). <https://doi.org/10.1016/j.drudis.2021.09.007>
21. Kuang, L. *et al.* Explore the toxicological mechanism of 6PPD-Q on human health through a novel perspective: The interaction between lactate dehydrogenase and 6PPD-Q. *International Journal of Biological Macromolecules* **293** (2025). <https://doi.org/10.1016/j.ijbiomac.2024.139266>
22. Wang, X. *et al.* Inhibition of the long non-coding RNA MALAT1 downregulates MAP2K1, suppressing the progression of hypopharyngeal squamous cell carcinoma. *Biomolecules and Biomedicine* **25**, 1023-1037 (2024). <https://doi.org/10.17305/bb.2024.11151>

23. Zhang, W. *et al.* EGFR and CYP signaling disruption underlies 6PPD-quinone hepatotoxicity: Insights from a network and machine learning approach. *Toxicology* **517** (2025).
<https://doi.org/10.1016/j.tox.2025.154235>
24. Song, D. Y. *et al.* α -Synuclein induces deficiency in clathrin-mediated endocytosis through inhibiting synaptojanin1 expression. *Journal of Neurochemistry* **167**, 461-484 (2023).
<https://doi.org/10.1111/jnc.15974>
25. Choudhry, H., Aggarwal, M. & Pan, P.-Y. Mini-review: Synaptojanin 1 and its implications in membrane trafficking. *Neuroscience Letters* **765** (2021). <https://doi.org/10.1016/j.neulet.2021.136288>
26. Song, Y., Weng, W. & Wu, S. Investigating the Potential Effects of 6PPDQ on Prostate Cancer Through Network Toxicology and Molecular Docking. *Toxics* **12** (2024).
<https://doi.org/10.3390/toxics12120891>
27. Lampron, M.-C., Paré, I., Al-Zharani, M., Semlali, A. & Loubaki, L. Cannabinoid Mixture Affects the Fate and Functions of B Cells through the Modulation of the Caspase and MAP Kinase Pathways. *Cells* **12** (2023). <https://doi.org/10.3390/cells12040588>
28. Dankner, M. *et al.* Clinical Activity of Mitogen-Activated Protein Kinase Inhibitors in Patients With MAP2K1 (MEK1)-Mutated Metastatic Cancers. *JCO Precision Oncology* (2025).
<https://doi.org/10.1200/po.24.00199>
29. Nair, P. *et al.* Structure and Toxicity Characterization of Alkyl Hydroxylated Metabolites of 6PPD-Q. *Environmental Science & Technology* **59**, 7474-7484 (2025). <https://doi.org/10.1021/acs.est.4c11823>
30. Luo, J. *et al.* Current understanding of gliomagenesis: from model to mechanism. *International Journal of Medical Sciences* **19**, 2071-2079 (2022). <https://doi.org/10.7150/ijms.77287>
31. Zhou, L. Generation and banking of patient-derived glioblastoma organoid and its application in cancer neuroscience. *American Journal of Cancer Research* **14**, 5000-5010 (2024).
<https://doi.org/10.62347/nsva5836>
32. Cao, Z. *et al.* Screening COPD-Related Biomarkers and Traditional Chinese Medicine Prediction Based on Bioinformatics and Machine Learning. *International Journal of Chronic Obstructive Pulmonary Disease* **Volume 19**, 2073-2095 (2024). <https://doi.org/10.2147/copd.S476808>

Figures

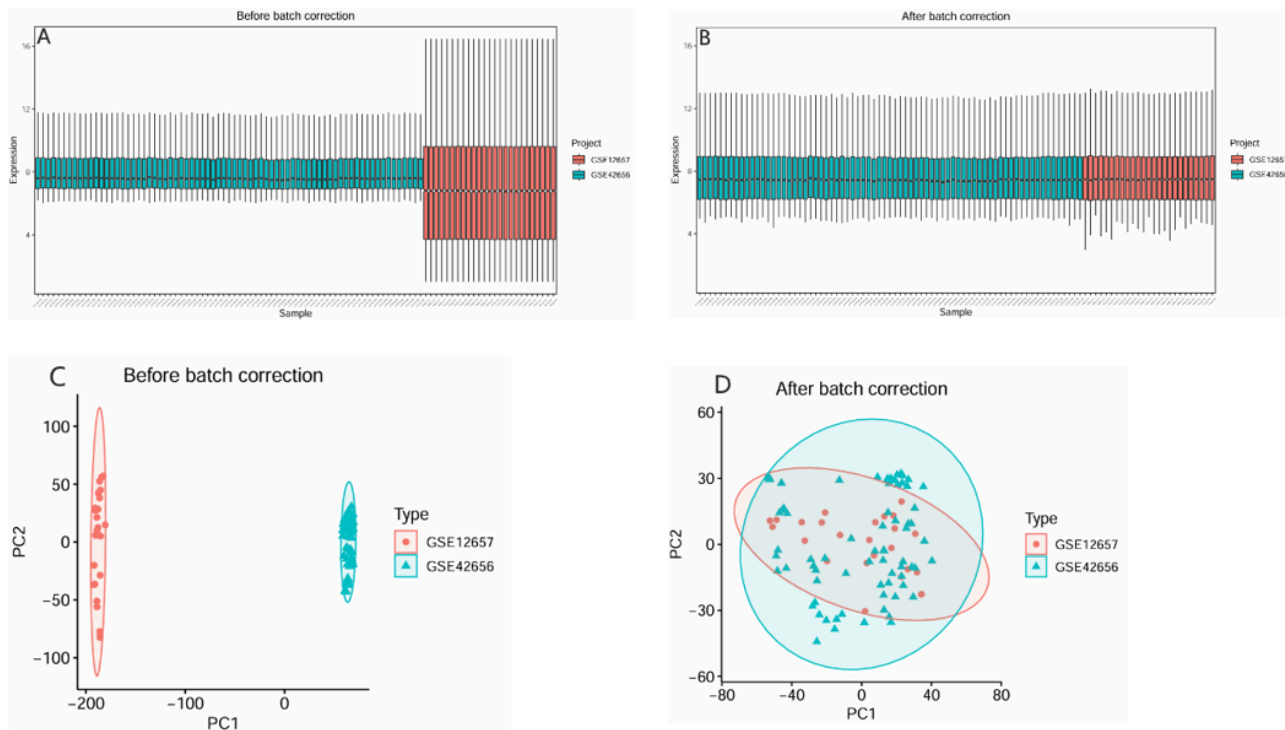


Figure 1

PCA and 2D visualizations before and after SVA correction: elimination of dataset-specific clustering and technical artifacts.

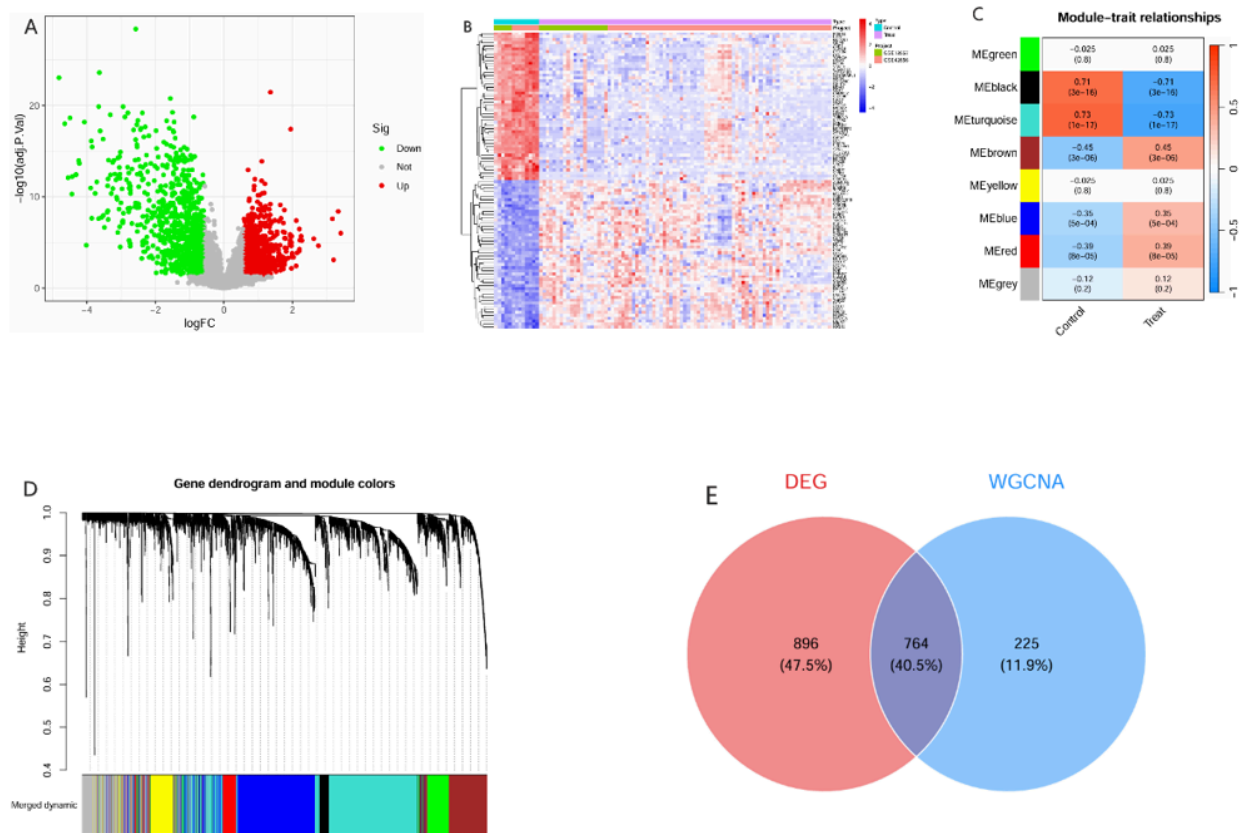


Figure 2

Analysis of differentially expressed genes and WGCNA module associations between treatment and control groups. (A) Volcano plot of DEGs between treatment and control groups ($\text{FDR} < 0.05$, $|\log_{2}(\text{FC})| > 0.585$); (B) Heatmap displaying normalized expression patterns of significant DEGs across samples; (C) Module-trait relationships from WGCNA. Turquoise and black modules show strongest correlation with treatment ($r = 0.73, 0.71$; $p < 1 \times 10^{-16}$); (D) Gene dendrogram with module assignments from WGCNA analysis; (E) Venn diagram identifying 225 overlapping genes between DEGs and WGCNA modules for subsequent analysis.

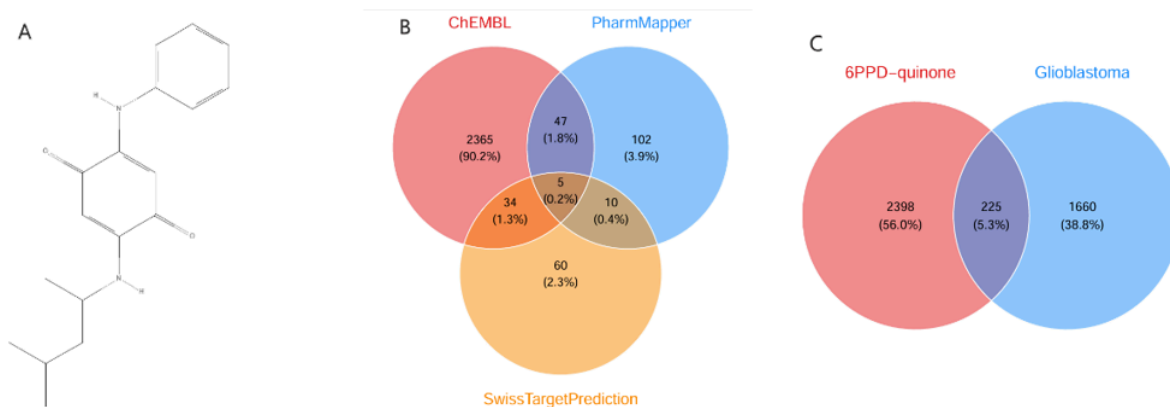


Figure 3

Chemical structure of AFB1, predicted 6PPD-Q targets, and identification of core candidate genes in glioblastoma. (A) Chemical structure of AFB1; (B) Overlap of 6PPD-Q targets predicted by three databases, totaling 2,623 unique targets; (C) Integration identifies 225 overlapping genes between 6PPD-Q targets and glioblastoma genes as core candidates.

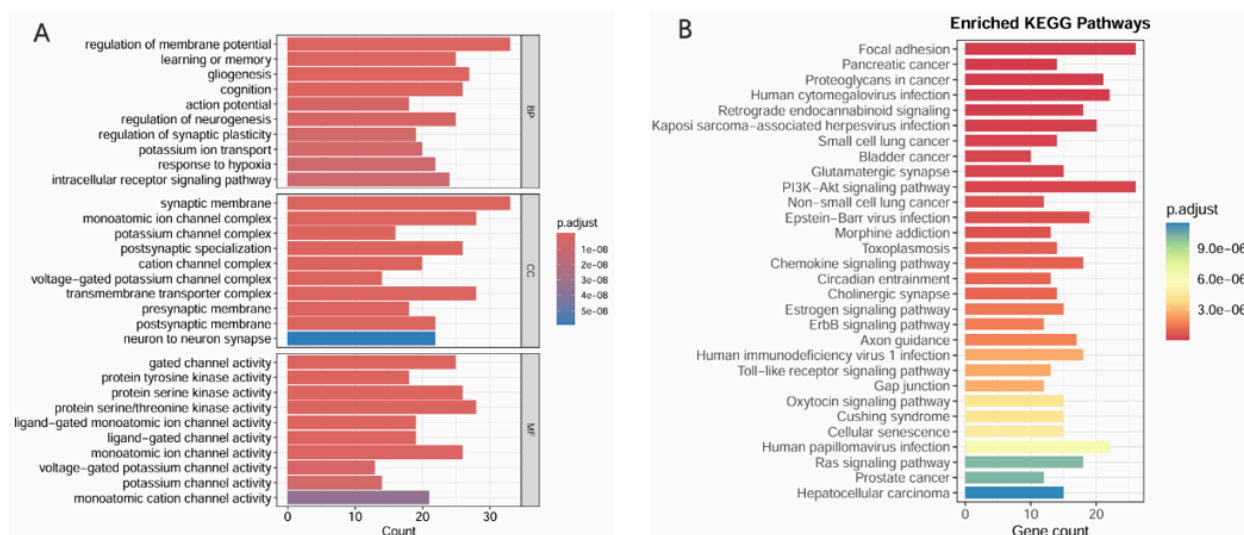


Figure 4

GO and KEGG enrichment analyses: neuronal and synaptic regulation, cancer-related pathways, and neuroactive ligand-receptor interactions. (A) GO enrichment analysis reveals significant terms in biological process, cellular component, and molecular function, highlighting neuronal and synaptic regulation; (B) KEGG pathway analysis identifies key cancer-related pathways and neuroactive ligand-receptor interactions.

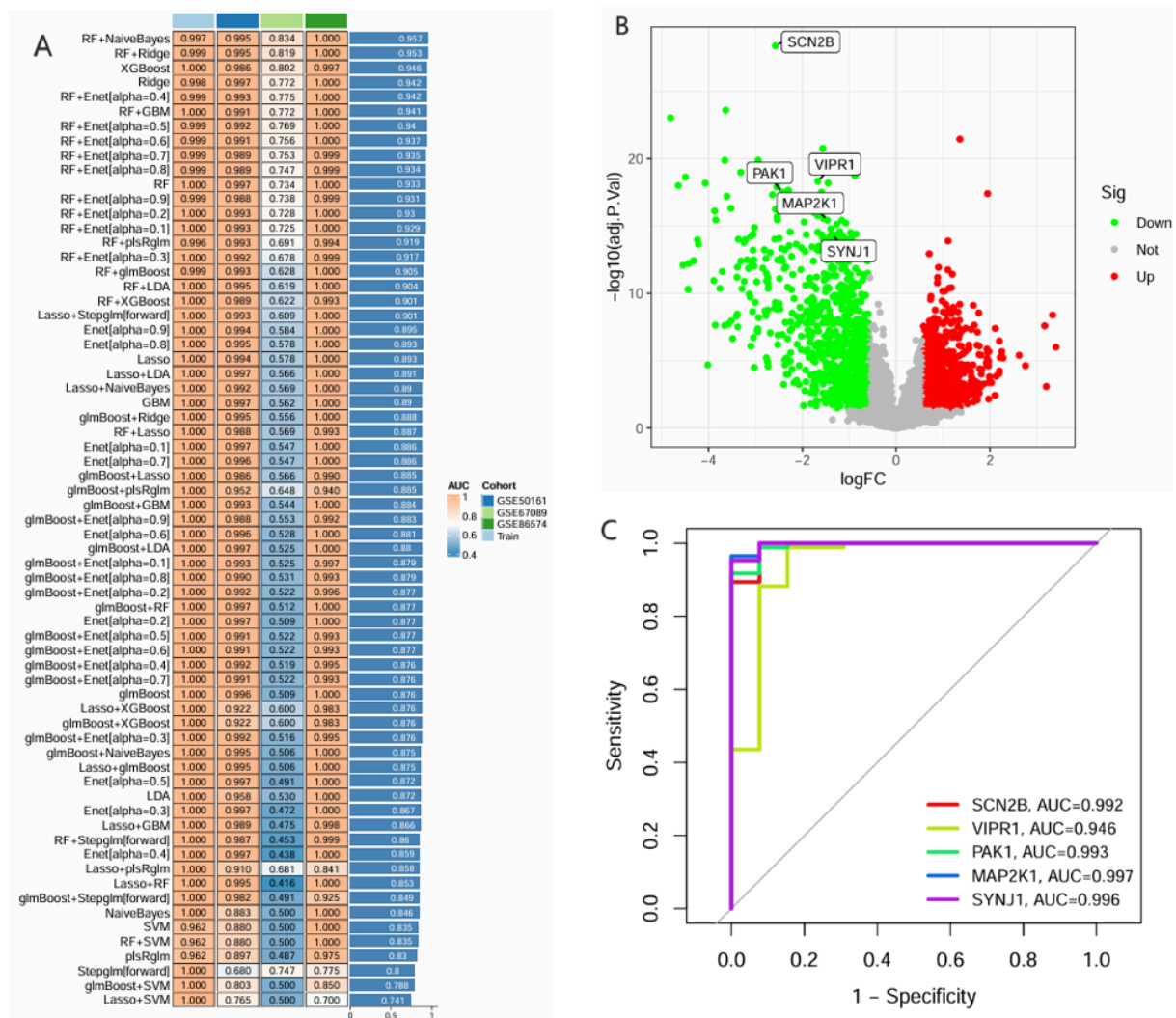


Figure 5

Performance ranking of 113 machine learning models and identification and validation of five key diagnostic genes. (A) Performance ranking of 113 machine learning models. Top models achieved AUC >0.99 in training and >0.95 in validation; (B) Volcano plot identifying five key diagnostic genes (MAP2K1, PAK1, SYNJ1, SCN2B, VIPR1) from DEG analysis; (C) ROC curves validating diagnostic performance of five key genes, with all AUCs >0.94 (MAP2K1: 0.997; PAK1: 0.993).



Figure 6

SHAP-based feature importance, directionality, and synergistic interactions of key genes. (A) Feature importance ranking. MAP2K1 contributes most (mean |SHAP value|=0.075), followed by PAK1 (0.0575); (B) Bee swarm plot showing directionality of feature impacts. High expression of MAP2K1/PAK1 increases prediction scores (positive SHAP values); (C) Dependence plots revealing synergistic interactions between key genes.

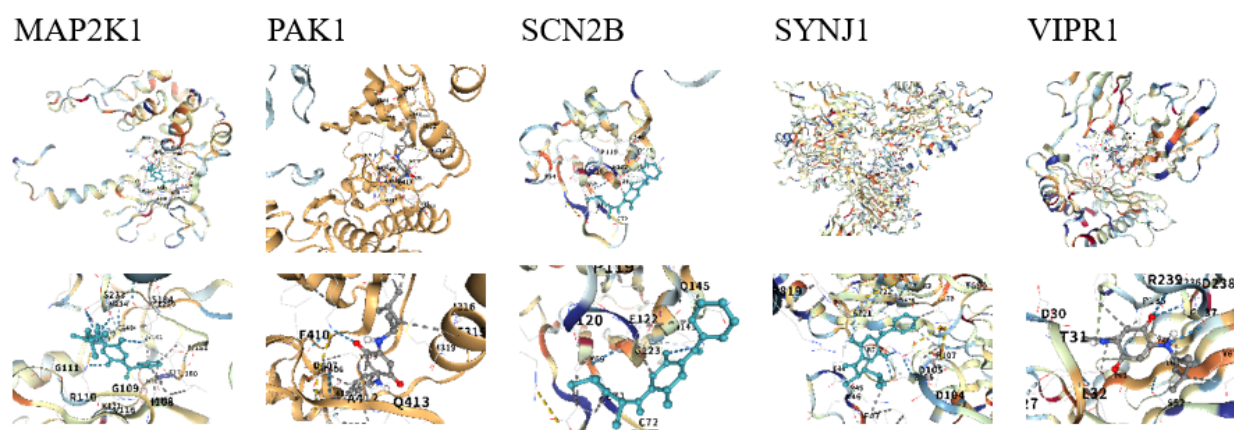


Figure 7

Docking results of receptors with 6PPD-Q.