Taylor & Francis
Taylor & Francis Group

RESEARCH PAPER

OPEN ACCESS · Check for updates

Identification of a five-gene signature deriving from the vacuolar ATPase (V-ATPase) sub-classifies gliomas and decides prognoses and immune microenvironment alterations

Chunxiao Qi[a,b], Lei Lei[c], Jinqu Hu[a], Gang Wang[a], Jiyuan Liu[a], and Shaowu Ou[a]

[a]Department of Neurosurgery, The First Hospital of China Medical University, Shenyang, Liaoning, China; [b]Department of Neurosurgery, The Second Hospital of Dalian Medical University, Dalian, Liaoning, China; [c]Department of Rheumatology and Immunology, Dalian Municipal Central Hospital Affiliated of Dalian Medical University, Dalian, Liaoning, China

ABSTRACT

Aberrant expression of coding genes of the V-ATPase subunits has been reported in glioma patients that can activate oncogenic pathways and result in worse prognosis. However, the predictive effect of a single gene is not specific or sensitive enough. In this study, by using a series of bioinformatics analyses, we identified five coding genes (ATP6V1C2, ATP6V1G2, TCIRG1, ATP6AP1 and ATP6AP2) of the V-ATPase that were related to glioma patient prognosis. Based on the expression of these genes, glioma patients were sub-classified into different prognosis clusters, of which C1 cluster performed better prognosis; however, C2 cluster showed more malignant phenotypes with oncogenic and immune-related pathway activation. The single-cell RNA-seq data revealed that ATP6AP1, ATP6AP2, ATP6V1G2 and TCIRG1 might be cell-type potential markers. Copy number variation and DNA promoter methylation potentially regulate these five gene expressions. A risk score model consisted of these five genes effectively predicted glioma prognosis and was fully validated by six independent datasets. The risk scores also showed a positive correlation with immune checkpoint expression. Importantly, glioma patients with high-risk scores presented resistance to traditional treatment. We also revealed that more inhibitory immune cells infiltration and higher rates of "non-response" to immune checkpoint blockade (ICB) treatment in the high-risk score group. In conclusion, our study identified a five-gene signature from the V-ATPase that could sub-classify gliomas into different phenotypes and their abnormal expression was regulated by distinct mechanisms and accompanied with immune microenvironment alterations potentially act as a biomarker for ICB treatment.

Introduction

Glioma is a kind of incurable disease with highly heterogeneous and infiltrative features, which accounts for more than 70% of the intracranial malignant tumors [1]. Maximum resection in a safe degree with postoperative chemotherapy and radiotherapy have been considered as standard treatment strategies for glioma patients [2]; however, despite decades of exploration in surgical techniques and chemotherapy medicine, there is still limited progress in the choice of treatment regimens and the prognosis after standard treatment still remains clear distinctions [3]. The molecular events in glioma have been discovered gradually, which brought us with novel insights into treating this dismal disease. Therefore, it is still urgent to investigate in detail the molecular mechanisms of glioma and to find out new potential targets for therapy [3].

Concurrently, mounting evidence indicates that the tumor microenvironment (TME) plays important roles in glioma prognosis and therapeutic resistance [4,5]. TME mainly consists of cellular components and an extracellular matrix, which constitutes the heterogeneous features of cancers accounting for the immune resistance [6,7]. The interplay between TME and its regulatory factors has caught more attention over the years. The vacuolar ATPase (V-ATPase) is a macromolecular complex that is highly conserved and usually overexpressed in cancer cells [8,9]. It consists of two functional domains, the V0 domain, embedded

---

in the cellular membranes and responsible for proton translocation, and the V1 domain, cytosolic and responsible for the ATP hydrolysis. The V-ATPase is responsible for acidifying and maintaining the homeostasis of pH in intracellular organelles, including the Golgi apparatus, endosome, lysosome and secretory vesicles, and excreting intracellular protons into the extracellular space. Abnormal presence of the V-ATPase at the plasma membrane of cancer cells contributes to the acidity of TME and leads to cancer invasion and progression [10]. The effectiveness of V-ATPase inhibitors in cancers *in vitro* or *in vivo* proved that targeting the V-ATPase is promising for cancer treatment, but there is still a long way to go. The aberrant expression of V-ATPase subunits in malignant entities and cancer cell lines has been reported by previous studies. The overexpression of the V-ATPase at the plasma membrane of invasive cancer cells possibly facilitates the activation of proteinases under low pH conditions and further modifies components of the extracellular matrix, such as MMPs [10]. Overexpression of V0a2-4, V0C and V1C1 subunits has been reported in breast, ovarian, musculoskeletal, prostate, and head and neck cancers [9]. Our previous study identified that overexpression of TCIRG1 (V0a3 subunit) was found in glioma patients and correlated with immune cell infiltration in GBM [11]. Higher expression of TCIRG1 predicted a mesenchymal subtype of GBM and worse prognosis. Multi-cancer V-ATPase molecular signatures among different cancer types were identified [12,13], which were either used to identify cancer subtypes with distinct molecular events, or to function as cancer biomarkers as well as basis for drug development.

Since the V-ATPase plays essential roles in a variety of cellular processes in cancer, especially for tumor invasion, metastasis and activation of oncogenic pathways [10]. As a kind of highly heterogeneous tumor, glioma has specific TME and strong adaptability to treatment modalities. The prognostic value of multiple subunits of the V-ATPase and their roles in glioma TME alterations and treatment resistance remain unclear. We suppose that the existence of multiple subunits of the V-ATPase might reflect specific compositions for specific cell types and pathophysiological conditions in glioma patients and the comprehensive exploration of their correlation with TME alterations as well as potentially activated pathways might provide novel targets for glioma treatment. In the present study, we examined the coding genes of the V-ATPase subunits and chaperones, focusing on providing further insights into biological basis of the V-ATPase for glioma patients.

Materials and methods

Transcriptome data collection and processing

A coding gene signature of subunits and chaperones of the V-ATPase was acquired from a previous study [14], including 14 genes of the V1 subunit (ATP6V1A, ATP6V1B1, ATP6V1B2, ATP6V1C1, ATP6V1C2, ATP6V1C3, ATP6V1D, ATP6V1E1, ATP6V1E2, ATP6V1F, ATP6V1G1, ATP6V1G2, ATP6V1G3 and ATP6V1H), 13 genes of the V0 subunit (ATP6V0A1, ATP6V0A2, TCIRG1, ATP6V0A4, ATP6V0C, ATP6V0B, ATP6V0D1, ATP6V0D2, ATP6V0E1, ATP6V0E2, RNASEK, ATP6AP1 and ATP6AP2) and three chaperone molecules (TMEM199, VMA21 and CCDC115). Transcriptome data of glioma patients were acquired from the Cancer Genome Atlas (TCGA) using the UCSC Xena browser (TCGA-GBMLGG dataset (n = 702), https://xenabrowser.net/datapages/) and the Chinese Glioma Genome Atlas (CGGA) database (CGGA-mRNA693 (n = 693), CGGA-mRNA325 (n = 325) and CGGA-mRNA301 (n = 301), http://www.cgga.org.cn/) and Gliovis platform (Rembrandt (n = 472) and Gravendeel (n = 284), http://gliovis.bioinfo.cnio.es) [15–25]. The histological diagnosis is done at the tissue collecting institution following the 2007 WHO central nervous system tumor classification system [26]. The main clinical information in this study contains overall survival (OS) time, survival status, isocitrate dehydrogenase-1 (IDH-1) status, 1p19q co-deletion status, gender and age, and so on. We removed the genes with 50% of samples having an mRNA relative expression value equal to 0, and 28 genes were identified for further study. Cases without relevant survival information or with an OS of less than 30 days were deleted. Detailed clinical data of diffuse gliomas for this analysis were listed in Supplementary Table 1. The TCGA-PanCancer

dataset was acquired from the UCSC Xena browser, 33 cancer types were contained, including acute myeloid leukemia (LAML), adrenocortical carcinoma (ACC), cholangiocarcinoma (CHOL), bladder urothelial carcinoma (BLCA), breast invasive carcinoma (BRCA), cervical squamous cell carcinoma and endocervical adenocarcinoma (CESC), colon adenocarcinoma (COAD), uterine corpus endometrial carcinoma (UCEC), esophageal carcinoma (ESCA), glioblastoma (GBM), head and neck squamous carcinoma (HNSC), kidney chromophobe (KICH), kidney clear cell carcinoma (KIRC), kidney papillary cell carcinoma (KIRP), lymphoid neoplasm diffuse large B-cell lymphoma (DLBC), liver hepatocellular carcinoma (LIHC), lower grade glioma (LGG), lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), skin cutaneous melanoma (SKCM), mesothelioma (MESO), uveal melanoma (UVM), ovarian serous cystadenocarcinoma (OV), pancreatic adenocarcinoma (PAAD), pheochromocytoma and paraganglioma (PCPG), prostate adenocarcinoma (PRAD), rectum adenocarcinoma (READ), sarcoma (SARC), stomach adenocarcinoma (STAD), testicular germ cell tumor (TGCT), thymoma (THYM), thyroid carcinoma (THCA), uterine carcinoma (UCS), which were used for viewing transcriptional landscape of the V-ATPase in multi-cancers and acquiring the immune phenotypes of 33 type TCGA cancers, including C1 (wound healing), C2 (IFN-gamma dominant), C3 (inflammatory), C4 (lymphocyte depleted), C5 (immunologically quiet) and C6 (TGF-beta dominant) [27]. Otherwise, several online platforms were used and the corresponding information was included in the subsequent description.

Prognostic genes selection and a risk score model construction

We used the TCGA-GBMLGG dataset to identify prognosis-related genes. First, univariate Cox regression analysis was used, and then the selected prognosis-related genes were included for more functional investigation and developed possible risk score model using the LASSO Cox regression and multivariate Cox regression analyses. Tuning parameter (lambda) selection was applied in the LASSO model by using 10-fold cross-validation via the minimum and 1-standard error (SE) criteria. The 1-SE criterion was then used to define a list of genes that potentially composed a risk score model; finally, through using multivariate Cox regression analysis, we identified five genes and their coefficients to comprise a risk score model. As the following formula, the risk score = Exp_1*Coe_1+ Exp_2*Coe_2+ … … + Exp_n*Coe_n, Exp stands for the mRNA expression value, and Coe indicates the coefficients calculated by multivariate Cox regression analysis. Receiver operator characteristic curves (ROC) and the value of area under the curve (AUC) were used to explore the accuracy of risk model in evaluating OS of glioma patients. A nomogram was used to facilitate clinicians to predict the 1-, 3- and 5-year OS of glioma patients. "rms", glmnet", "survival", and "timeROC" R packages were used for the data analysis.

Multi-omics validation of target gene in gliomas

The differential expression of hub genes between glioma samples and normal brain tissues was evaluated by the Genotype-Tissue Expression (GTEx) and TCGA datasets, and transcriptome data were acquired from the UCSC Xena browser (https://xenabrowser.net/datapages/) [15]. The transcription and protein levels validation of hub genes in glioma cell lines was performed based on Cancer Cell Line Encyclopedia (CCLE, https://sites.broadinstitute.org/ccle). The protein level exploration was performed using the Human Protein Atlas (HPA) database (https://www.proteinatlas.org) [28], which provides immunohistochemistry (IHC) staining of glioma samples and immunofluorescence of glioma U251 cell line. The Clinical Proteomic Tumor Analysis Consortium (CTPAC) database through the UALCAN online platform (ualcan.path.uab.edu/index.html) [29] was used to explore the protein level expression of hub genes between normal brain tissues and glioblastoma samples (Glioblastoma multiforme dataset, including normal brain tissues (n = 10) and GBM (n = 99), ATP6V1C2 was not contained in this database). Z-values represent standard deviations from the median across samples for the given cancer type. Log2 Spectral count ratio values from CPTAC were first normalized within each sample

profile and then normalized across samples. The Tumor Immune Single Cell Hub (TISCH, http://tisch.comp-genomics.org) [30], an interactive and online database that integrates single-cell transcriptome profiles of millions of cells from multiple high-quality cancer dataset across multiple cancer types. Glioma_GSE131928_10X dataset was used for single-cell analysis among 9 glioma patients consisted of 27 cell clusters and 3 cell types [31], including major-lineage (Astrocyte (AC)-like malignant cells, exhausted CD8 T cells (CD8Tex), mesenchymal (MES)-like malignant, malignant, monocytes or macrophages (Mono/Macro), Neural-progenitor-like malignant cells (NPC-like malignant), oligodendrocyte-precursor-cell-like malignant cells (OPC-like malignant) and Oligodendrocyte), minor-lineage (AC-like malignant, CD8Tex, M1 macrophage, MES-like malignant, malignant, monocyte, NPC-like malignant, OPC-like malignant and oligodendrocyte) and malignancy (immune cells, malignant cells and others).

Tumor microenvironment investigation and immune checkpoints blockade (ICB) treatment evaluation

Estimation of Stromal and Immune cells in Malignant Tumor tissues using Expression data (ESTIMATE) is a commonly used algorithm that can be utilized to calculate immune scores and stromal scores in tumor tissues based on expression profiles [32]. Higher immune scores represent higher immune cell infiltration in cancer microenvironment. ImmuCellAI is an online tool (http://bioinfo.life.hust.edu.cn/ImmuCellAI) that can be applied to estimate ICB effects in cancers based on gene expression profiles [33]. The single sample gene set enrichment analysis (ssGSEA) was used to define a enrichment score to represent the degree of entire enrichment of a gene set in each sample within a given dataset using R package "GSVA". Twenty-eight types of immune cell gene set signatures were obtained from the TISIDB database (cis.hku.hk/TISIDB/index.php) [34], which also provided online correlation analysis between target gene expression and tumor-infiltrating lymphocytes (TILs).

Functional enrichment analysis

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses were performed to investigate the potential function of hub genes and their similar genes by the Database for Annotation, Visualization and Integrated Discovery (DAVID) online platform (version 6.8, https://david.ncifcrf.gov) [35,36]. The gene set enrichment analysis (GSEA) was performed using the Omicshare online tool (https://www.omicshare.com/tools/), a free online platform for data analysis, of which the H (hallmark) and C2 (curated) KEGG gene sets were selected for analysis. Only the enrichment pathways with a P value less than 0.05 and a false discovery rate (FDR) less than 0.25 were considered statistically significant.

Subgroups classification and principal component analysis (PCA)

Consensus clustering is a class discovery technique for the detection of unknown possible clusters consisting of items with similar intrinsic features. Based on a comprehensive expression of the hub genes, we identified distinct subgroups of 654 tumor samples from the TCGA dataset with R "ConsensusClusterPlus" package (agglomerative km clustering with 1-pearson correlation distances and resampling 80% of the samples for 50 repetitions). The optimal number of clusters was determined using the empirical cumulative distribution function (CDF) plot. We then use PCA to verify the results of the grouping, which was performed through the ClustVis platform (https://biit.cs.ut.ee/clustvis/) [37], a web tool for visualizing clustering of multivariate data.

GSCALite, GeneMANIA, and ChIP-X enrichment analysis 3 (ChEA3) online platforms analyses

GSCALite platform (http://bioinfo.life.hust.edu.cn/web/GSCALite/) is an online algorithm for integrating genomic and immune data of 33 cancer types from the TCGA, drug responses from the Genomics of Drug Sensitivity in Cancer (GDSC) and the Cancer Therapeutic Response Portal (CTRP) and normal tissue data from the GTEx

[38]. The single nucleotide variation (SNV) and copy number variation (CNV) data (processed using GISTIC 2.0 method) were collected across 33 cancer types from the TCGA database. GeneMANIA (http://www.genemania.org) is a flexible, user-friendly web interface for generating hypotheses about gene function, analyzing gene lists and prioritizing genes for functional assays [39]. Association data include protein and genetic interactions, pathways, co-expression, co-localization and protein domain similarity. We used the GeneMANIA to find functionally similar genes with our hub genes using available genomics and proteomics data. ChEA3 (https://amp.pharm.mssm.edu/ChEA3) is a web-serve to provide transcription factor (TF) enrichment analysis tool that ranks TFs associated with user-submitted gene sets [40]. The ChEA3 background database contains a collection of gene set libraries generated from multiple sources including TF–gene co-expression from RNA-seq studies, TF–target associations from ChIP-seq experiments, and TF–gene co-occurrence computed from crowd-submitted gene lists. Enrichment results from these distinct sources are integrated to generate a composite rank that improves the prediction of the correct upstream TF compared to ranks produced by individual libraries.

Statistical analysis

All graphic and statistical work was completed using GraphPad Prism software (version no. 7; GraphPad Software, Inc.), R language (version no. 3.6.3) and aforementioned online tools. Kaplan–Meier survival analysis was performed to compare OS between two cohorts using the log-rank test. We used unpaired t-tests to compare two groups and ANOVA to compare three groups. The median value in each step was used to define the high (≥median value) and low (<median value) groups. Correlation analysis was performed by Pearson's or Spearman's correlation coefficient analysis. A receiver operating characteristics (ROC) curve was utilized to assess the sensitivity and specificity, and the area under the curve (AUC) was used to present the efficacy. Univariate and multivariate Cox regression analyses were also performed to evaluate the

prognostic value and to calculate hazard ratios. A $P < 0.05$ was considered to indicate a statistically significant difference in all steps.

Results

Prognostic hub genes selection

We first used the TCGA Pan-Cancer dataset to view transcriptional landscape of the V-ATPase in 33 type cancers. The result indicated that the coding genes of the V-ATPase subunits presented variable patterns among different cancers (Figure 1(a)). To determine the prognostic value of subunits of the V-ATPase in glioma patients, we first used univariate Cox regression analysis to search for prognosis-related genes and found 22 genes were correlated with the prognosis in glioma patients (Figure 1(b)). Then, LASSO regression predictive formula (Lambda.1-SE criteria) was used to identify the accurate prognosis-related genes, of which six genes were selected from the TCGA-GBMLGG cohort (Figure 1(c,d)). Moreover, multivariate Cox regression analysis was applied to further validate the results and obtained the associated coefficients. Finally, five genes (APT6V1C2, ATP6V1G2, TCIRG1, ATP6AP1 and ATP6AP2) were identified to play prognostic roles independently in glioma patients and were defined as hub genes, of which ATP6V1G2 showed protective effects and others showed oncogenes with hazard ratio >1 (Figure 1(e)). By Pearson's coefficient analysis, we revealed that the expression of ATP6V1G2 was negatively correlated with others in transcriptional level (Figure 1(f)).

Multiple platforms validation of hub genes in glioma samples and cell lines

In the aforementioned study, we preliminarily clarified the important roles of these five genes (ATP6V1C2, ATP6V1G2, TCIRG1, ATP6AP1 and ATP6AP2) in glioma patients. The differential transcription levels between glioma patients and normal brain cortex tissues were performed by comparing the GTEx and TCGA datasets. We found that except for ATP6V1G2, other genes showed higher expression in GBM compared
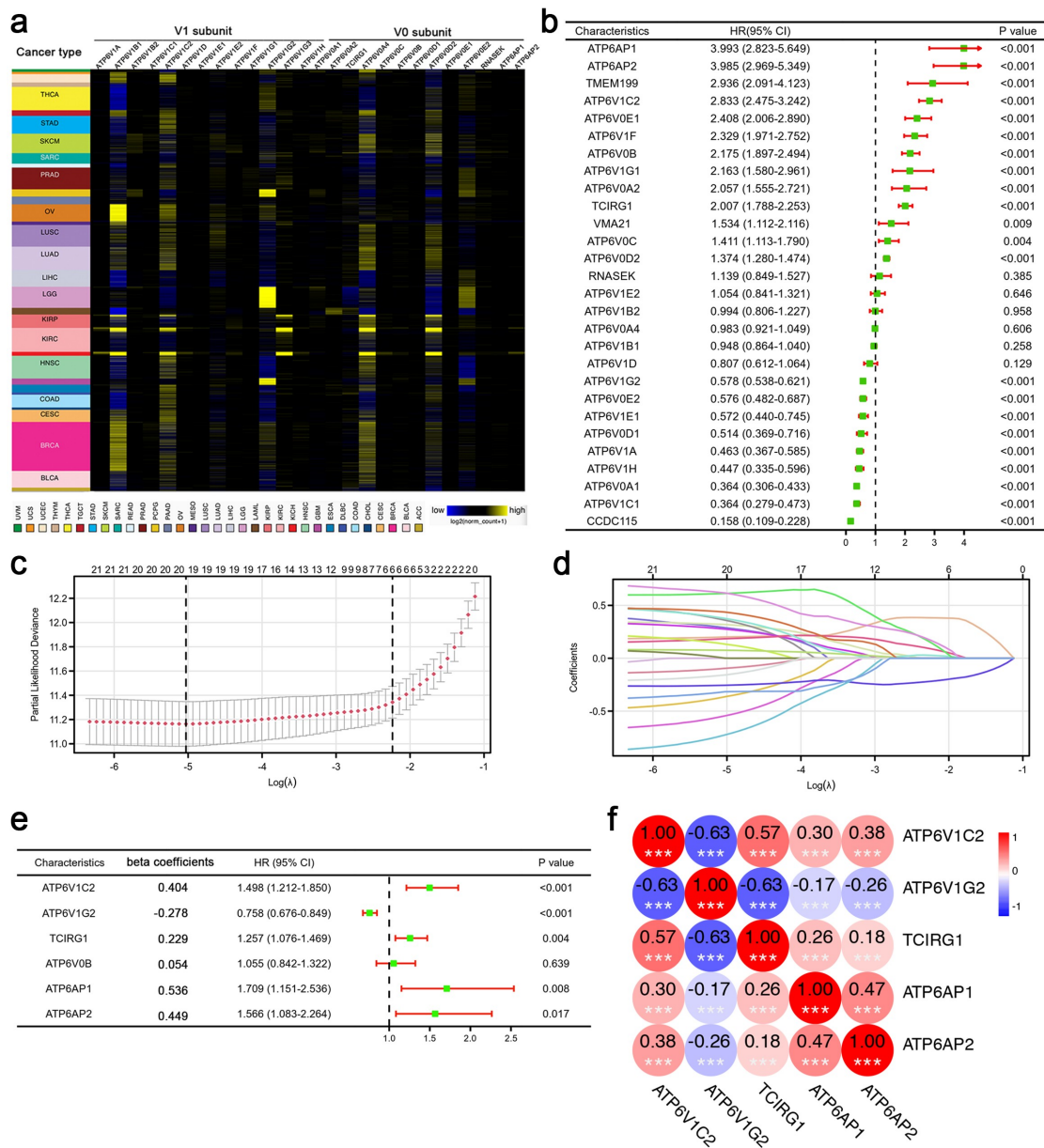
Figure 1. Coding genes expression of the V-ATPase subunits in the TCGA Pan-Cancer dataset and their prognostic value for glioma patients. (a) Coding genes expression profile of V1 and V2 subunits of the V-ATPase in the TCGA cancers. (b) Univariate Cox regression analysis for V-ATPase subunits coding genes and three chaperones in glioma patients based on the TCGA-GBMLGG dataset. (c,d) LASSO regression analysis for prognosis-related genes in the TCGA-GBMLGG dataset and 1-SE criterion was used to select candidate genes. (e) Multivariate Cox regression analysis showed that ATP6V1C2, ATP6V1G2, TCIRG1, ATP6AP1 and ATP6AP2 played independent prognostic roles in glioma patients. (f) Correlation analysis showed that the expression of ATP6V1G2 was negatively correlated with other genes. LAML: acute myeloid leukemia; ACC: adrenocortical carcinoma; CHOL: cholangiocarcinoma; BLCA: bladder urothelial carcinoma; BRCA: breast invasive carcinoma; CESC: cervical squamous cell carcinoma and endocervical adenocarcinoma; COAD: colon adenocarcinoma; UCEC: uterine corpus endometrial carcinoma; ESCA: esophageal carcinoma; GBM: glioblastoma; HNSC: head and neck squamous carcinoma; KICH: kidney chromophobe; KIRC: kidney renal clear cell carcinoma; KIRP: kidney renal papillary cell carcinoma; DLBC: lymphoid neoplasm diffuse large B-cell lymphoma; LIHC: liver hepatocellular carcinoma; LGG: lower grade glioma; LUAD: lung adenocarcinoma; LUSC: lung squamous cell carcinoma; SKCM: skin cutaneous melanoma; MESO: mesothelioma; UVM: uveal melanoma; OV: ovarian serous cystadenocarcinoma; PAAD: pancreatic adenocarcinoma; PCPG: pheochromocytoma and paraganglioma; PRAD: prostate adenocarcinoma; READ: rectum adenocarcinoma; SARC: sarcoma; STAD: stomach adenocarcinoma; TGCT: testicular germ cell tumor; THYM: thymoma; THCA: thyroid carcinoma; UCS: uterine carcinoma. ***: $P < 0.001$.
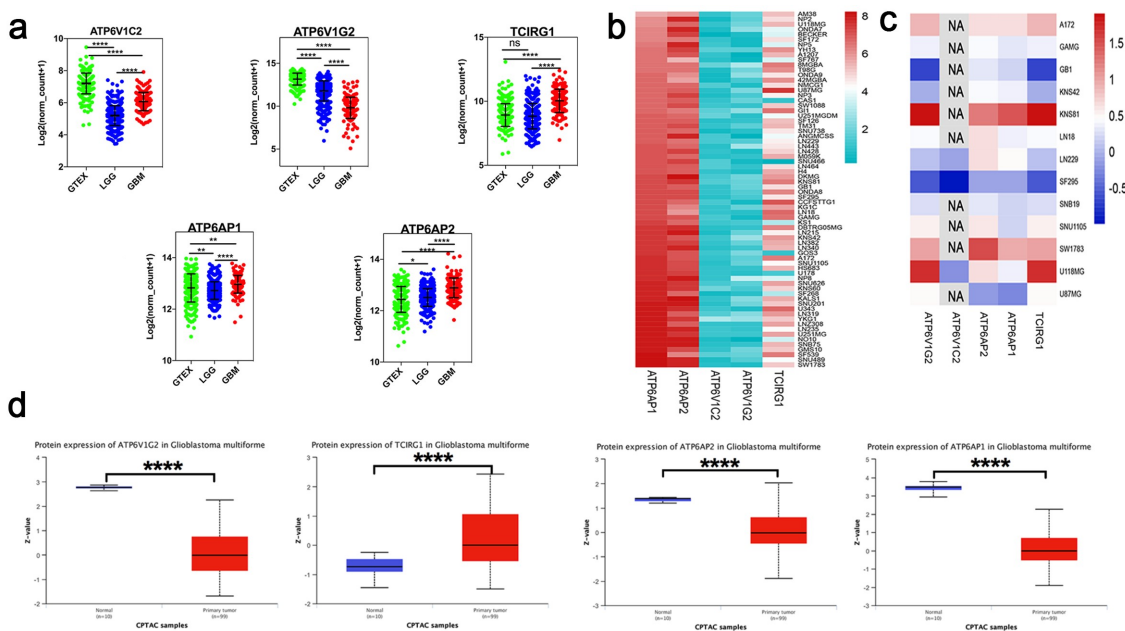
Figure 2. Multi-omics analysis of hub genes expression in glioma patients and glioma cell lines. (a) The differential expression of five hub genes between the GTEx (brain cortex) and TCGA-GBMLGG. (b) Five hub genes expression in transcription levels among 70 glioma cell lines based on the CCLE database. (c) Protein levels validation of five hub genes in glioma cell lines based on the CCLE database. (d) Different protein levels expression of ATP6V1G2, ATP6AP1, TCIRG1 and ATP6AP2 between GBM and normal brain tissues based on CPTAC database. *:P < 0.05; **:P < 0.01; ***:P < 0.001;****:P < 0.0001.

with LGG, but for ATP6V1C2 and ATP6V1G2, higher expression was also found in normal brain tissues compared with glioma samples (Figure 2(a)). Otherwise, we searched for the CCLE database, the transcription levels of five hub genes in 70 glioma cell lines were identified (Figure 2(b)), and the protein levels of 5 hub genes in 13 glioma cell lines were investigated (Figure 2(c)), which will unquestionably facilitate the *in vitro* and *in vivo* experiments in the future. Through data mining in the HPA database, positive staining intensity by IHC of hub genes was found in glioma samples (Figure Supplementary 1 (a,b)). Mining data in the CPTAC database showed that high protein levels expression of ATP6V1G2, ATP6AP1 and ATP6AP2 was found in normal brain tissues, TCIRG1 showed high expression in GBM samples (Figure 2(d)). To characterize the intracellular localization of these five genes in glioma cell lines, we assessed the distribution of TCIRG1, ATP6V1C2, ATP6AP1 and ATP6AP2 (ATP6V1G2 was not acquired) within the microtubules and nucleus in glioma U251 cell line. Immunofluorescence results

showed that ATP6AP2 co-localized with the nuclear marker, TCIRG1 was mainly localized in mitochondria, ATP6AP1 was mainly localized in plasma membrane and cytoplasm, and ATP6V1C2 was mainly localized in mitochondria (Figure supplementary 1(c)). The different localization in glioma cell line further identified the diverse functions of these five hub genes.

Analysis of the hub genes expression at single-cell level and correlations with tumor-infiltrating lymphocytes (TILs) in glioma patients

Glioma is a kind of highly heterogeneous cancer that presents little benefit from immunotherapy and other treatment strategies [41]. The rapid development of scRNA-seq technology has been widely adopted to feature the TME of cancers. Hence, we explored the transcription levels of hub genes at a single-cell resolution based on the TISCH database [30]. According to different clusters and cell types, we found that immune
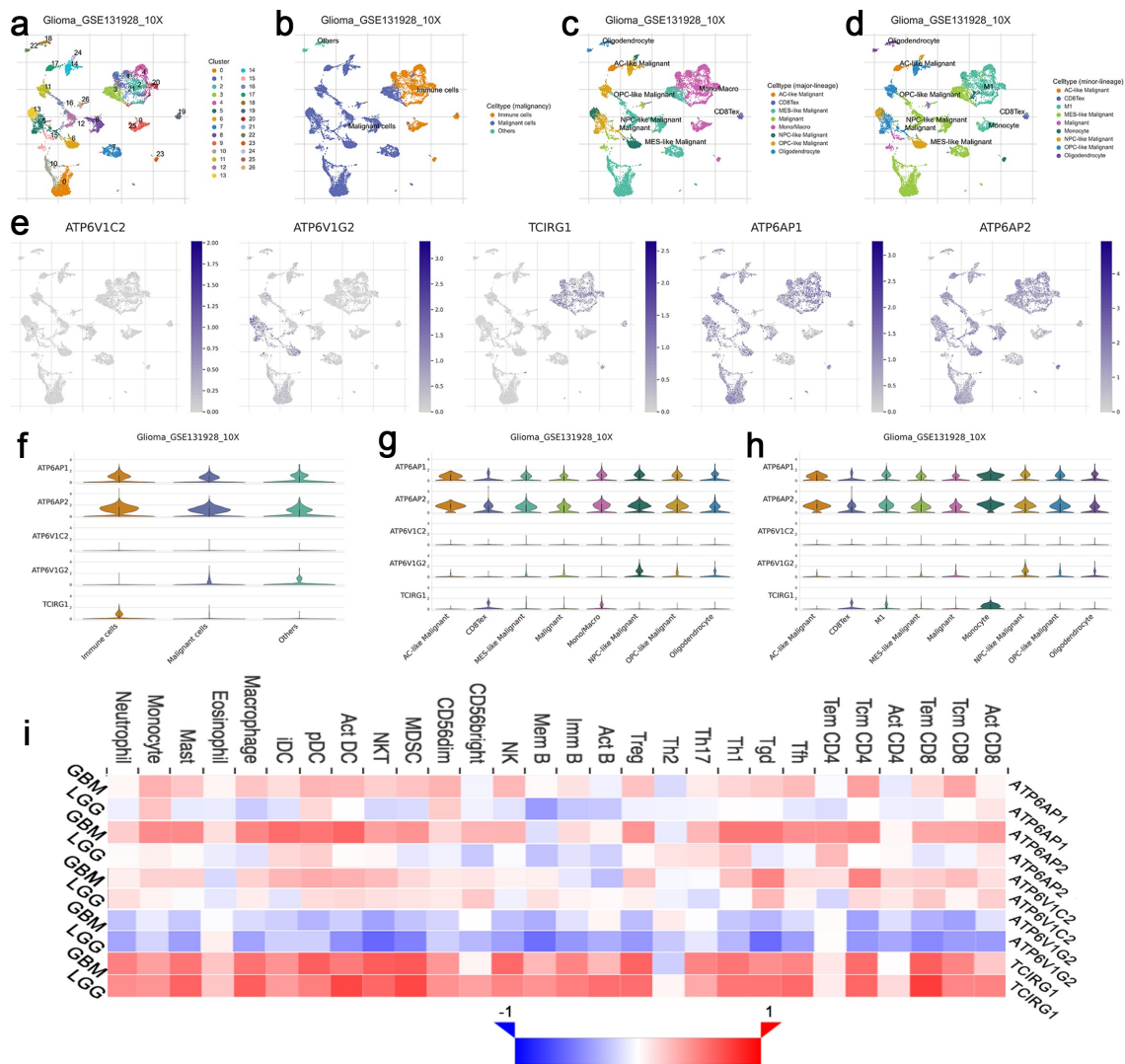
Figure 3. Hub genes expression at a single-cell resolution and their correlations with tumor-infiltrating lymphocytes (TILs) in glioma patients. (a) UMAP visualization of dataset glioma_GSE131928_10X. Colors represent the major-lineage cluster ID. (b-d) UMAP visualization of dataset glioma_GSE131928_10X. Colors represent the malignancy (b), major-lineage (c) and minor-lineage (d) cell-types. (e) Comparison of five hub genes expression at a single-cell resolution in glioma. (f-h) The grid violin plot reflects the expression distribution of five hub genes in different cell types of glioma. (i) The Spearman correlation analysis between five hub genes expression and abundance of TILs in LGG and GBM was shown in a heatmap. AC: astrocyte; CD8Tex: exhausted CD8 T cells; MES: mesenchymal; Mono/Macro: monocytes or macrophages; NPC: neural-progenitor-like malignant cells, OPC: oligodendrocyte-precursor-cell; Act: activated; Tcm: central memory T cell; Tem: effector memory T cell; Tfh: T follicular helper cell; Tgd: gamma delta T cell; Th: T helper cell; Treg: regulatory T cell; Mem B: memory B cell; Imm B: immature B cell; NK: natural killer cell; MDSC: myeloid derived suppressor cell; pDC: plasmacytoid dendritic cell; iDC: immature dendritic cell.

and malignant cells occupied the vast majority (Figure 3(a-d)). We then explored the expression of five hub genes in different cell types. The results indicated that ATP6AP1, ATP6AP2, TCIRG1 and ATP6V1G2 might be cell-type markers (Figure 3(e-h) and Supplementary Table 2), but no significance was found for ATP6V1C2. Otherwise, we used a heatmap to present the relations between hub genes expression and abundance of 28 TILs (Figure 3(i)).

The results preliminarily elucidated the correlation between the five-hub genes expression and different cell types in glioma patients.

Five hub genes could be an applicable criterion to stratify glioma patients into different subgroups

Given the important role of five hub genes, we tried to figure out whether they could be utilized to identify the molecular subtypes and potentially activated pathways in glioma patients. Based on the expression of five hub genes, glioma patients in the TCGA-GBMLGG cohort can be stratified into two subgroups (Figure 4(a-c)). Moreover, the PCA showed that two subgroups could be divided into different directions (Figure 4(d)), which further

confirmed the classification was effective. We performed survival analysis between the two groups and the results revealed that glioma patients in two groups had different OS and progression free survival (PFS). Patients in the C1 cluster had longer OS and PFS compared with those in the C2 cluster (Figure 4(e,f)). Furthermore, we used ssGSEA to calculate 28 immune cell infiltration scores in glioma patients. Most of the infiltration scores in the C2 cluster were higher than that in the C1 cluster (Figure 4(g)). Glioma patients in the C2 cluster had more malignant phenotypes (Figure 4(h)). By GSEA, several tumor-related pathways, such as angiogenesis and epithelial mesenchymal transition (EMT), and immune-related pathways might be activated in the C2 cluster (Figure supplementary 2(a,b)).
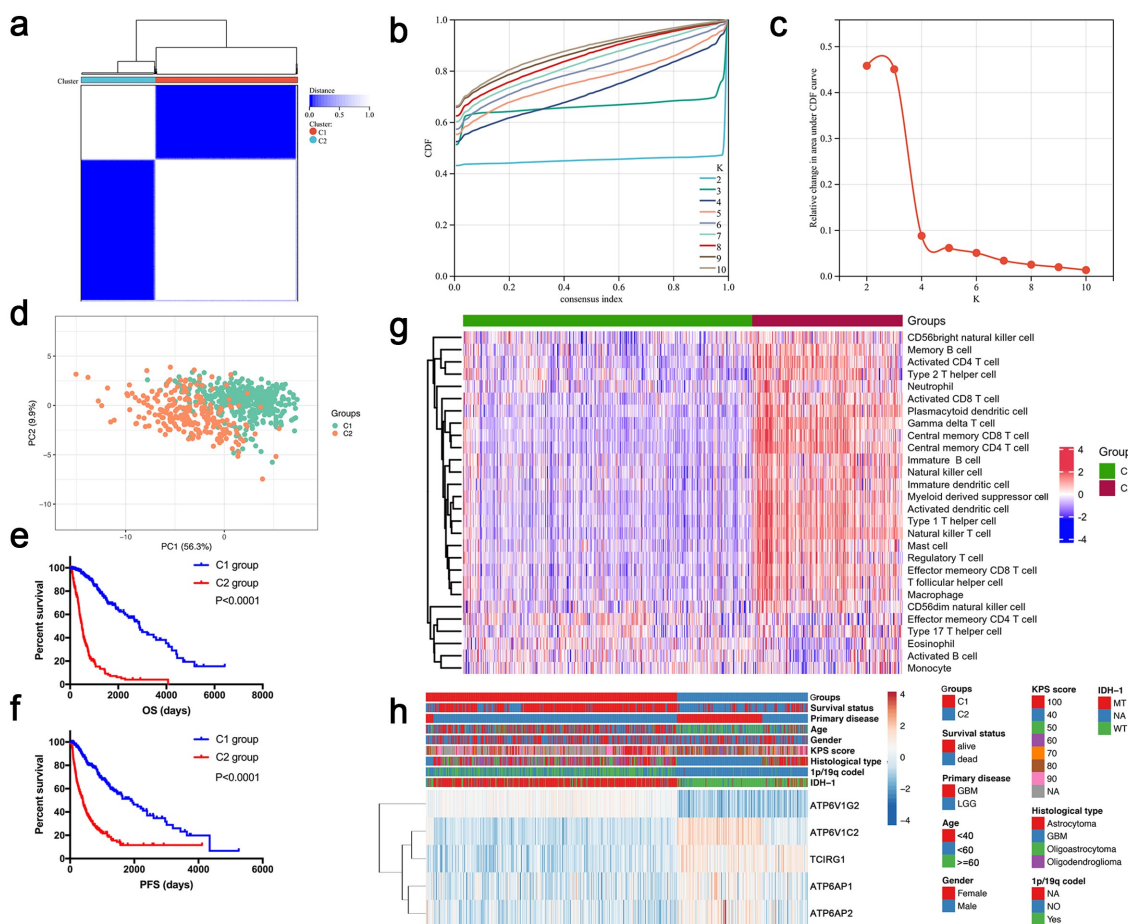

Figure 4. Consensus cluster analysis and principal component analysis (PCA).(a-c) Cumulative distribution function (CDF) and relative change in the area under the CDF curve of the consensus cluster for k = 2–10 in TCGA-GBMLGG; k = 2 was presented and selected for further analysis. (d) PCA analysis of the RNA-seq profile of glioma patients in the TCGA-GBMLGG. (e,f) Kaplan–Meier survival curves for glioma patients in the TCGA-GBMLGG, patients in C1 cluster had longer OS and progression free survival (PFS). (g) Patients in C2 cluster had more immune cells infiltration than that in C1 cluster. (h) Heatmap and clinicopathological traits of the two clusters (cluster 1 and cluster 2) defined by the 5-gene consensus expression based on the TCGA-GBMLGG.
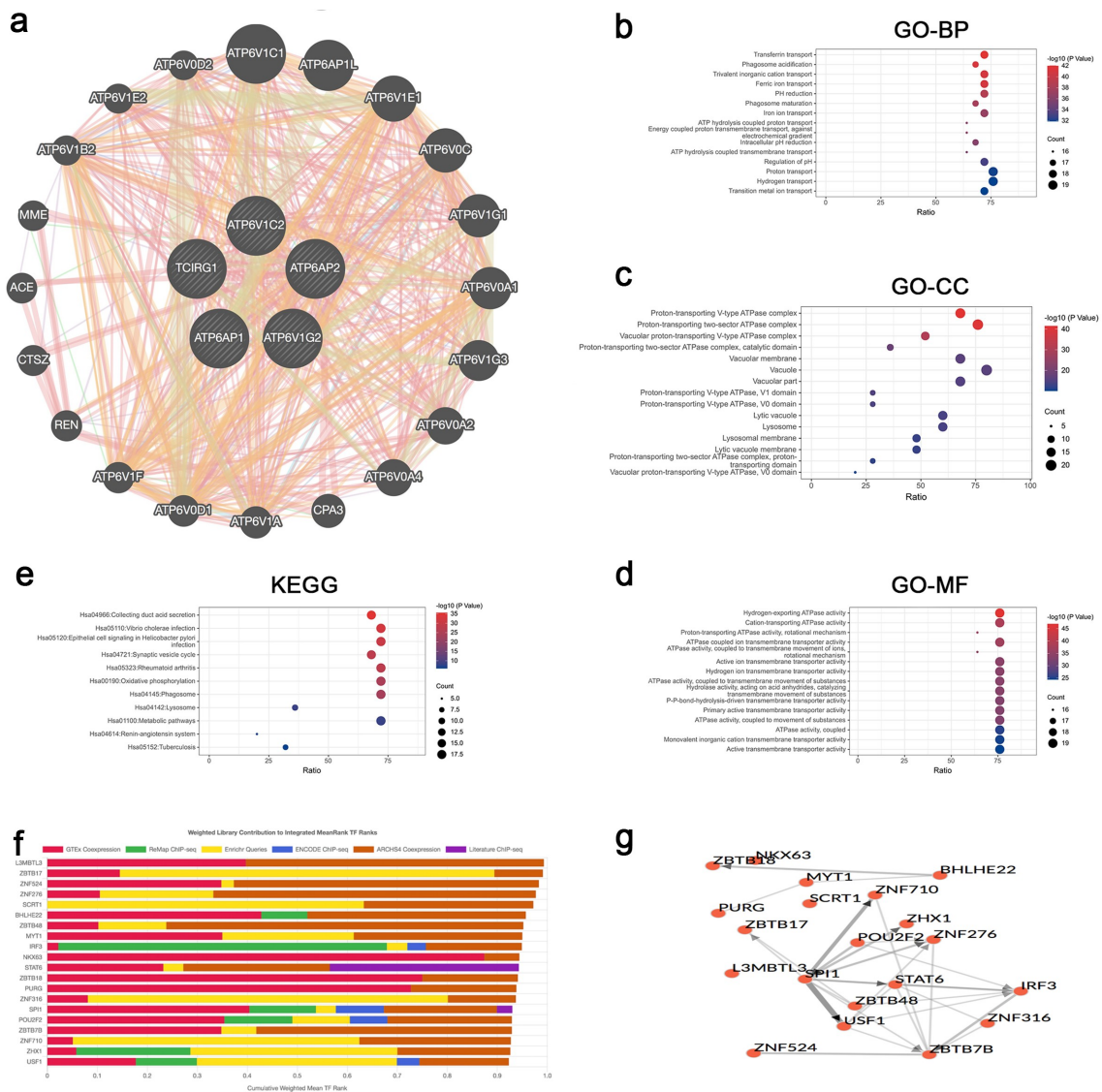
Figure 5. A co-expression network construction and transcription factors (TFs) regulation of five hub genes in glioma patients. (a) A network of five hub genes and their 20 related genes was analyzed by GeneMANIA. (b) Biological process (BP); (c) Cellular component (CC); (d) Molecular function (MF); (e) KEGG pathway analysis. (f,g) Top 20 TFs with strong regulatory relationships with five hub genes and their interaction network by ChEA3 analysis.

Functional enrichment analysis of hub genes

We created a network of hub genes and their 20 related genes by the Gene MANIA (Figure 5(a)). Proteins that interact with hub genes include ATP6V1C1, ATP6AP1L, ATP6V1E1, ATP6V0C, ATP6V1G1, ATP6V0A1, ATP6V1G3, ATP6V0A2, ATP6V0A4, CPA3, ATP6V1A, ATP6V0D1, ATP6V1F, REN, CTSZ, ACE, MME, ATP6V1B2, ATP6V1E2 and ATP6V0D2. GO and KEGG pathways analyses of hub genes and their related genes were performed through the DAVID online database. The biological processes, such as GO:0033572 (transferrin transport),

GO:0090383 (phagosome acidification), GO:0072512 (trivalent inorganic cation transport), GO:0015682 (ferric iron transport), GO:0045851 (pH reduction), GO:0090382 (phagosome maturation), GO:0006826 (iron ion transport), GO:0015991 (ATP hydrolysis coupled proton transport), GO:0015988 (energy coupled proton transmembrane transport, against electrochemical gradient), GO:0051452 (intracellular pH reduction), GO:0090662 (ATP hydrolysis coupled transmembrane transport), GO:0006885 (regulation of pH), GO:0015992 (proton transport), GO:0006818 (hydrogen transport) and GO:0000041 (transition

metal ion transport) were remarkably regulated by the five hub genes and regulated genes (Figure 5(b)). For cellular components, GO:0033176 (proton-transporting V-type ATPase complex), GO:0016469 (proton-transporting two-sector ATPase complex), GO:0016471 (vacuolar proton-transporting V-type ATPase complex), GO:0033178 (proton-transporting two-sector ATPase complex, catalytic domain), GO:0005774 (vacuolar membrane), GO:0005773 (vacuole), GO:0044437 (vacuolar part), GO:0033180 (proton-transporting V-type ATPase, V1 domain), GO:0033179 (proton-transporting V-type ATPase, V0 domain), GO:0000323 (lytic vacuole), GO:0005764 (lysosome), GO:0005765 (lysosomal membrane), GO:0098852 (lytic vacuole membrane), GO:0033177 (proton-transporting two-sector ATPase complex, proton-transporting domain) and GO:0000220 (vacuolar proton-transporting V-type ATPase, V0 domain) were associated with these genes (Figure 5(c)). Besides, five hub genes and related genes potentially influence the molecular functions, such as GO:0036442 (hydrogen-exporting ATPase activity), GO:0019829 (cation-transporting ATPase activity), GO:0046961 (proton-transporting ATPase activity, rotational mechanism), GO:0042625 (ATPase coupled ion transmembrane transporter activity) and GO:0044769 (ATPase activity, coupled to transmembrane movement of ions, rotational mechanism) (Figure 5(d)).

For KEGG pathway analyses, these pathways including hsa04966 (Collecting duct acid secretion), hsa05110 (Vibrio cholerae infection), hsa05120 (Epithelial cell signaling in Helicobacter pylori infection), hsa04721 (Synaptic vesicle cycle), hsa05323 (Rheumatoid arthritis), hsa00190 (Oxidative phosphorylation) were correlated with the functions of five hub genes and related genes (Figure 5(e)).

In addition, we identified the top 20 TFs with strong regulatory relationships with five hub genes and their interaction networks by ChEA3 analysis (Figure 5(f,g)). The GO-BP analysis showed that these TFs mainly enriched in gene expression regulation processes, such as regulation of transcription, DNA-templation, and regulation of RNA biosynthetic processes (Figure supplementary 2(c)).

Potential regulation mechanisms for hub genes and drugs sensitivity analysis

To identify the potential mechanisms that regulate hub gene expression in cancers, we investigate the copy number variation (CNV) and methylation data from the TCGA database. The CNV alteration showed that the main CNV types in glioma and other cancers were heterozygous amplification and deletion (Figure 6(a) and Figure supplementary 3(a)). TCIRG1 and ATP6V1C2 had the higher rate of heterozygous amplification (10.33% and 6.07%) in LGG and GBM, respectively (Figure 6(b)). For the heterozygous deletion, ATP6AP2 (18.91%) and ATP6AP1 (19.41%) showed higher rate in LGG and GBM (Figure 6(b)). However, the rates of homozygous amplification and deletion were lower (less than 2%) in glioma patients (Figure 6(c)).

Correlation analysis indicated that mRNA expression of ATP6AP2, ATP6V1G2 and ATP6V1C2 was positively correlated with CNV in LGG, but TCIRG1 showed negatively correlated with CNV in LGG. For GBM, only ATP6AP2 showed significant correlation with CNV (Figure 6(d)).

Aiming at a comprehensive understanding of the mechanism regulating the expression of these hub genes, we also investigated the pan-cancer dataset. CNV percentage analysis showed that heterozygous amplification of ATP6V1C2 in ACC, ESCA, KIRP, OV, LUAD and HNSC; ATP6AP2 in ACC, UCS, KIRP and SARC; ATP6V1C2 in LUSC, UCS, TGCT, OV, ESCA, BLCA, LUAD and CESC; ATP6V1G2 in UCS, SKCM, UVM, OV, LIHC, LUAD, READ and BLCA; and TCIRG1 in LUAD, UCS, OV, ESCA, HNSC, LUSC and KICH were greater than 25%. Heterozygous deletion of ATP6AP1 in KICH, CHOL and SARC; ATP6AP2 in KICH, OV and CHOL; ATP6V1C2 in KICH; ATP6V1G2 in KICH; and TCIRG1 in TGCT were greater than 40% (Figure supplementary 3(b) and Supplementary Table 3). Homozygous analysis showed that the percentage of amplified and deleted genes of TCGA cancers was less than
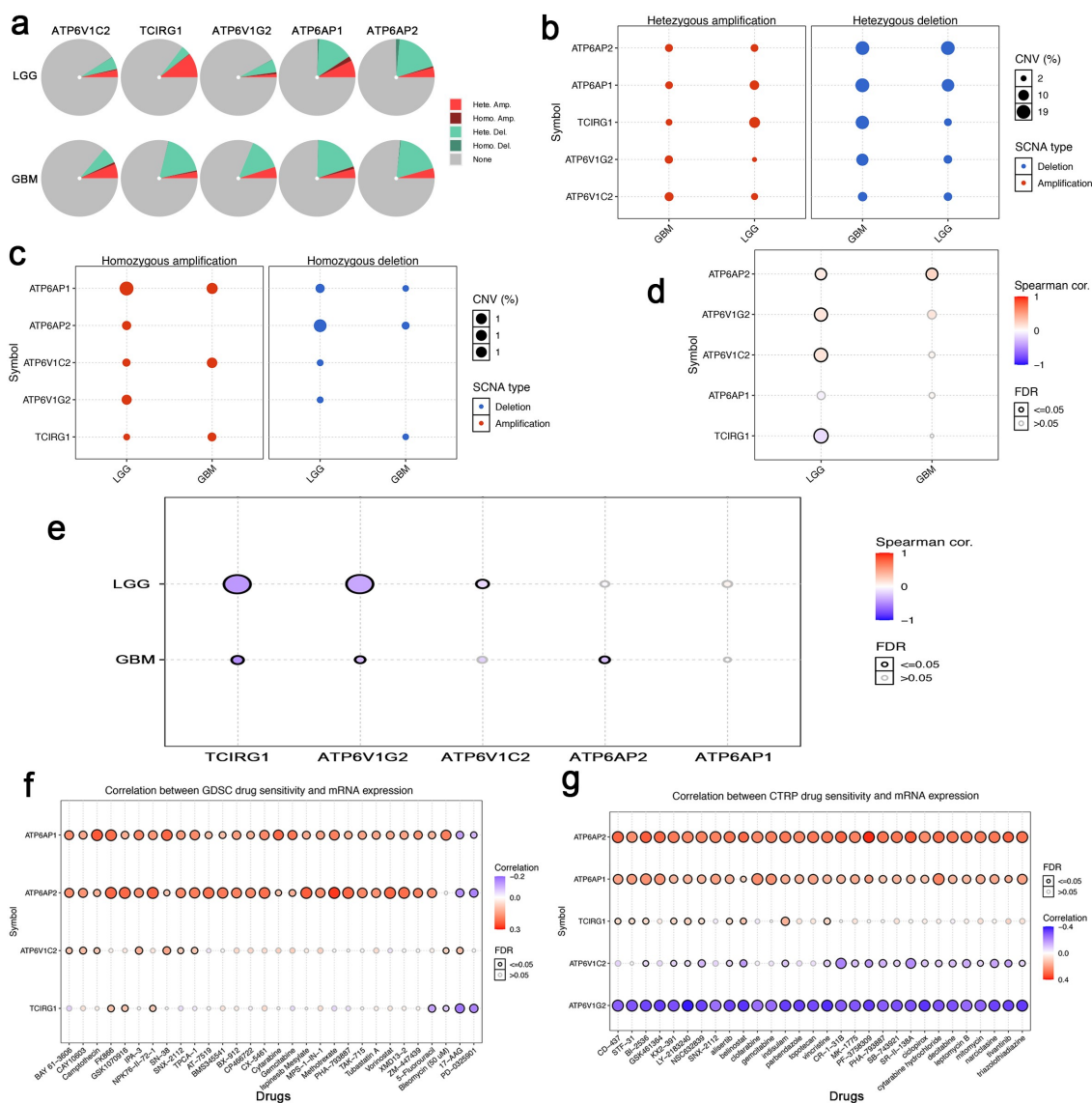
Figure 6. Potential regulation mechanisms of hub genes and drugs sensitivity analysis. (a) Copy number variation (CNV) distribution in LGG and GBM. CNV pie chart showing the combined heterozygous/homozygous CNV of each gene in each cancer. A pie chart representing the proportion of different types of CNV of one cancer, and different colors represent different types of CNV. (b,c) Heterozygous and homozygous CNV profile showing the percentage for each gene in LGG and GBM. (d) The correlation between CNV percentage and paired gene expression in glioma by Spearman correlation analysis. The size of the point represents the statistical significance, the bigger the dot size, the higher the statistical significance. (e) The correlation between DNA methylation and hub genes expression by Spearman correlation analysis in LGG and GBM. Blue points represent a negative correlation and red points represent a positive correlation. (f,g) The correlation between hub genes expression and drug IC50 in the GDSC and CTRP databases by Pearson correlation analysis. The top 30 drugs ranked by the Pearson correlation coefficients were listed. LGG: lower grade glioma; GBM: glioblastoma; GDSC: genomics of drug sensitivity in cancers; CTRP: cancer therapeutics response portal.

10% (Figure supplementary 3(c)). Correlation analysis indicated that mRNA expression was positively correlated with CNV in most cancer types; however, the negative correlation was found for ATP6AP2 in KIRP, KICH, KIRC and SKCM; APT6AP1 in PCPG and KIRP; and ATP6V1G2

in THYM (FDR<0.05) (Figure supplementary 3(d)).

These results revealed that the CNV in five hub genes might mediate their abnormal expression and exert an important role in cancer progression. In addition, DNA methylation and mRNA

expression correlation analysis revealed that most of the expression levels of hub genes were negatively correlated with their DNA methylation levels, only ATP6AP1 in THYM and TGCT showed a positive correlation between methylation and gene expression (FDR<0.05) (Figure 6(e) and Figure supplementary 3(e)).

CNV and DNA methylation in cancers are two important mechanisms that critically regulate gene expression. Genomic abnormals influence the effects of clinical response to chemotherapy and targeted therapy treatment [42]. We searched the GDSC and CTRP databases to investigate the roles of the five hub genes in chemotherapy and targeted therapy. Pearson's correlation analysis showed that drug sensitivity in GDSC database (ATP6V1G2 was not acquired in this database) toward BAY 61–3606, CAY10603, Camptothecin, FK886, GSK1070916, IPA-3, NPK76-II-72-1, SN-38, SNX-2112, TPCA-1, AT-7519, BMS345541, BX-912, CP466722, CX-5461, Cytarabine, Gemcitabine, Ispinesib Mesylate, MPS-1-IN-1, Methotrexate, PHA-793887, TAK-715, Tubastatin A, Vorinostat, XMD13-2, ZM-447439 and 5-Fluorouracil was correlated with the expression of ATP6AP1 and ATP6AP2 (positive correlation with IC50); however, 17-AAg and PD-0325901 showed negative correlation with the expression of ATP6AP1 and ATP6AP2. The expression of ATP6V1C2 was positively correlated with BAY 61–3606, CAY10603, Camptothecin, IPA-3, SN38, SNX-2112, TPCA-1, Bleomycin and 17AAG. The expression of TCIRG1 was positively correlated with FK886, GSK1070916 and NPK76-II-72-1, but negatively correlated with 5-Fluorouracil, Bleomycin, 17-AAG and PD-0325901 (Figure 6(f)). Through data mining in the CTRP database, we found that the expression of ATP6AP2 and ATP6AP1 was positively correlated with CD437, STF-31, BI-2536, GSK461364, KX2-391, LY-2183240, NSC632839, SNX-2112, alisertib, belinostat, clofarabine, gemcitabine, indisulam, parbendazole, topotecan, vincristine, CR-1-31B, MK-1775, PF-3758309, PHA-793887, SB-743921, SR-II-138A, ciclopirox, cytarabine hydrochloride, decitabine, leptomycin B, mitomycin, narciclasine, tivantinib and triazolothiadiazine. However, the expression of ATP6V1G2 was negatively correlated with these drugs. For ATP6V1C2, its expression was negatively correlated with BI-2536, KX2-391, LY-2183240, NSC632839, alisertib, belinostat, indisulam, vincristine, CR-1-31B, MK-1775, PF-3758309, PHA-793887, SB-743921, SR-II-138A, ciclopirox, cytarabine hydrochloride, decitabine, leptomycin B, mitomycin, narciclasine, tivantinib and triazolothiadiazine. For TCIRG1, its expression was positively correlated with CD437, STF-31, BI-2536, KX2-391, LY-2183240, NSC632839, alisertib, belinostat, indisulam and vincristine (Figure 6(g)). In summary, our results indicate that the aberrant expression of these hub genes potentially mediated resistance to chemotherapy and targeted drug therapy, and they are valuable indicators for cancer treatment.

Multiple types of mutations for hub genes

We analyzed SNV data to detect the variant frequency and types of hub genes in glioma patients and other-type cancers through GSCALite online platform [38]. Of the five hub genes, the top three mutated cancer types were UCEC, SKCM and STAD (Figure supplementary 4(a)). In glioma patients, SNV percentage analysis indicated that TCIRG1 had the most mutated frequency rate (38%), followed by ATP6AP1 (31%), ATP6V1C2 (23%), ATP6AP2 (15%) and ATP6V1G2 (15%) (Figure 7(a,b)). However, in the Pan-Cancer analysis, ATP6V1C2 had the most mutated frequency rate (31%), followed by TCIRG1 (28%), ATP6AP1 (28%), ATP6AP2 (24%) and ATP6V1G2 (7%) (Figure supplementary 4 (b)). Among the five hub genes, the most common variant type was missense mutation (Figure 7(b) and Figure supplementary 4 (b)). Mutation in genes can influence the immune phenotypes of cancers [43,44]. We compared the hub gene expression with Pan-Cancer immune phenotypes [27]. The results showed that the five hub genes had different expression distributions in six cancer immune phenotypes (One-way ANOVA test, P < 0.05) (Figure 7(c)). For all the glioma patients, ATP6AP1 and ATP6AP2 had similar patterns of expression distribution, of which high expression was found in C3 (inflammatory) and C4 type (lymphocyte depleted). The expression distribution patterns between TCIRG1 and ATP6V1C2
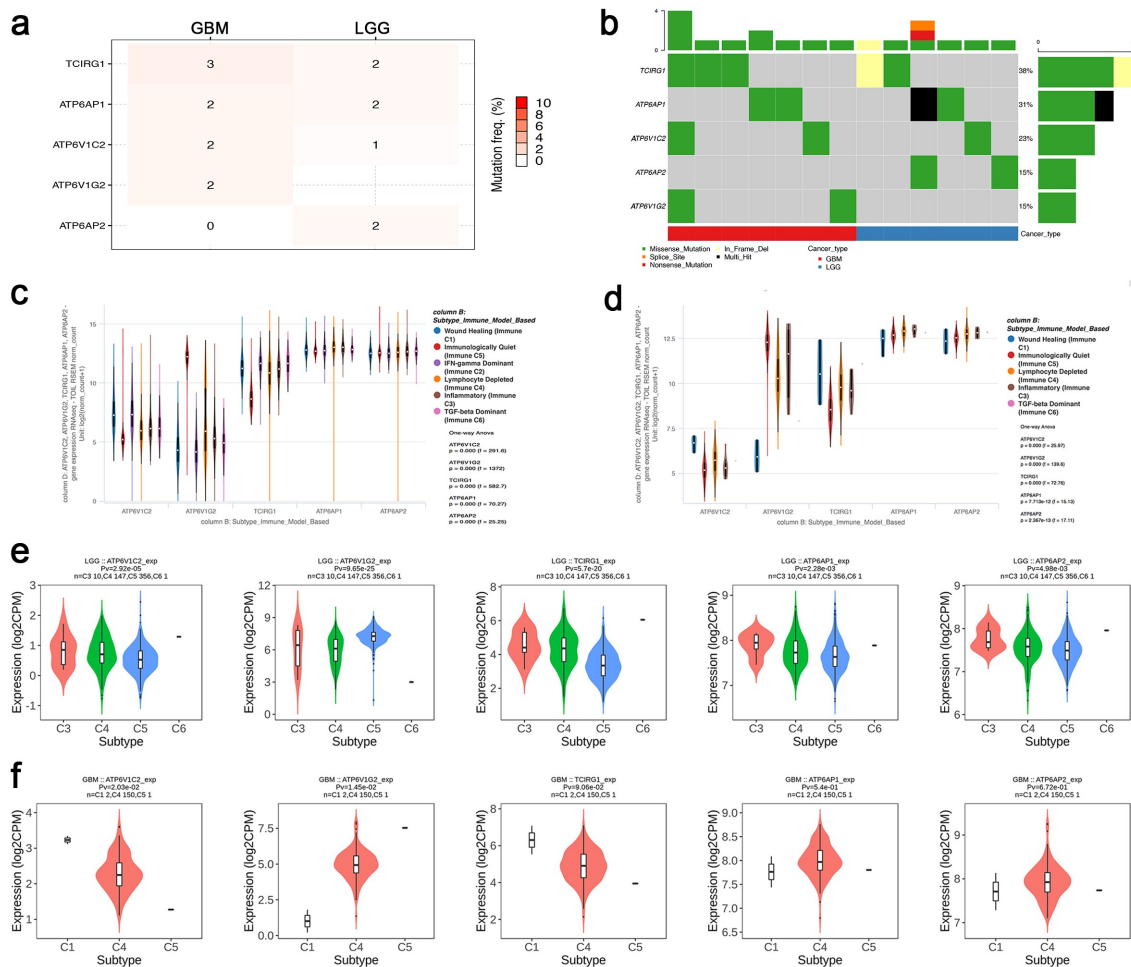
Figure 7. Single nucleotide variation (SNV) frequency and variant types of hub genes in glioma patients. (a) Mutation frequency of five hub genes in glioma patients. Numbers represent the number of samples that have the corresponding mutated gene for a given cancer. Zero indicates that there was no mutation in the gene coding region, and no number indicates there was no mutation in any region of the gene. (b) SNV oncoplot. An oncoplot showing the mutation distribution of five hub genes and a classification of SNV types in glioma patients. (c) Five hub genes showing different expression distribution in six immune phenotypes among the TCGA Pan-Cancer dataset (ANOVA test, P < 0.01). (d) Five hub genes expression distribution in immune phenotypes of glioma patients (ANOVA test, P < 0.01). (e) Five hub genes expression in different immune phenotypes of LGG based on TISIDB (ANOVA test). (f) Five hub genes expression in different immune phenotypes of GBM based on TISIDB (ANOVA test). LGG: lower grade glioma; GBM: glioblastoma; Pv: P value.

were similar that high expression was found in C1 type (wound healing); however, the expression distribution of ATP6V1G2 was different from others, which showed high expression in C5 type (immunologically quiet) (Figure 7(d)). For LGG, higher ATP6V1G2 expression was found in C5 type, but other hub genes showed a higher expression in the C3 type (Figure 7(e)). Only ATP6V1G2 and ATP6V1C2 showed significant distributions

among distinct GBM subtypes, whereas no significant difference was found in other hub genes (Figure 7(f)).

A risk score model construction and its application in clinical

We created a risk score model using the five hub genes as followed formula: risk score = 0.404*ATP6V1C2-
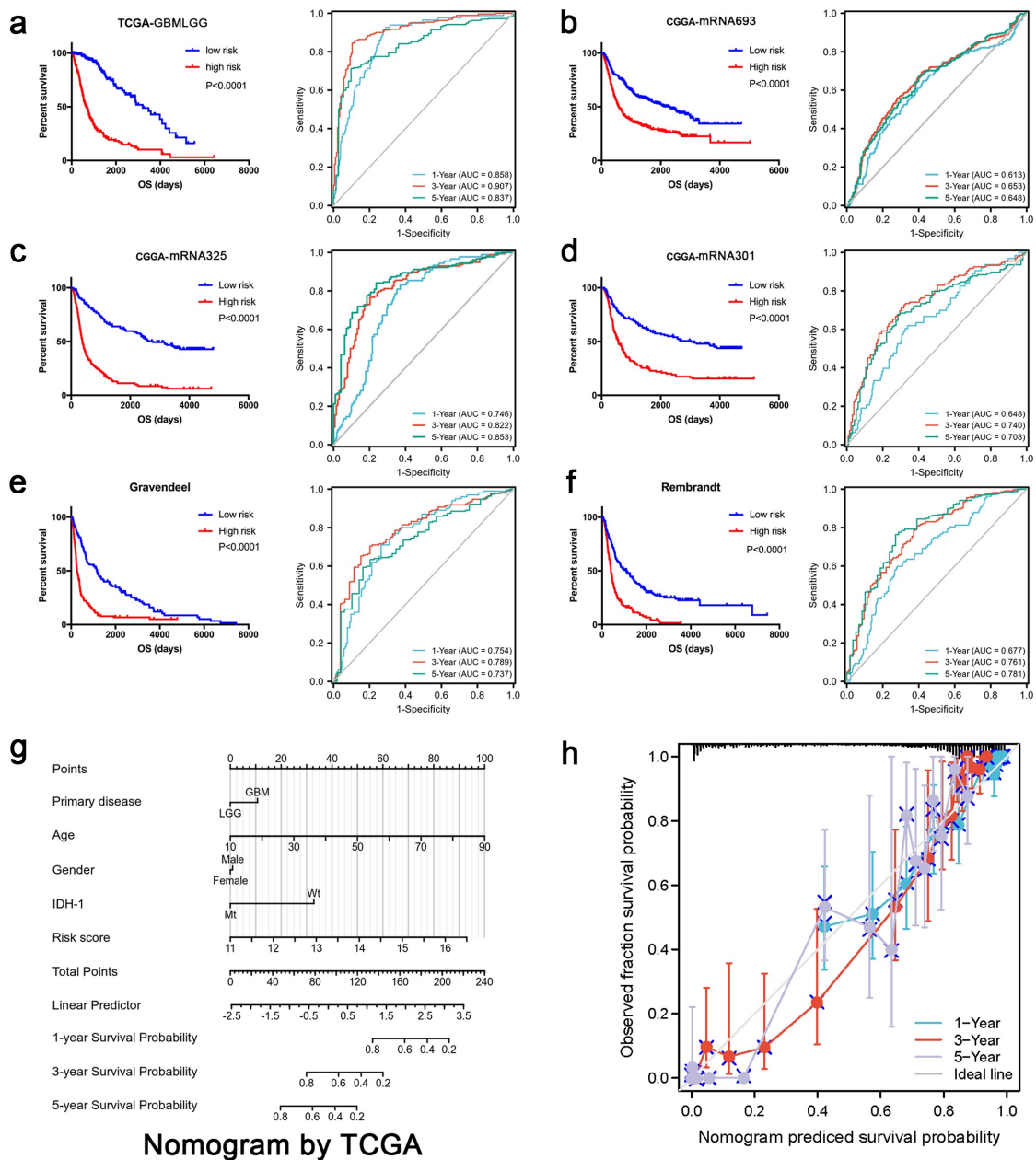
Figure 8. A risk score model consists of five hub genes can effectively predict glioma patients' prognosis. (a-f) Multiple independent datasets from the TCGA-GBMLGG, CGGA-mRNA693, CGGA-mRNA325, CGGA-mRNA301, Gravendeel and Rembrandt datasets proved that our risk score model could effectively predict OS of glioma patients and their corresponding ROC curves confirmed the conclusion. (g) Combined with other clinical parameters, the risk score model can be used to create a nomogram to help clinicians predict the 1-, 3- and 5-year OS of glioma patients. (h) The caliberation curve validated the accuracy of the nomogram.

0.278*ATP6V1G2 + 0.229*TCIRG1 + 0.536*ATP6AP1- + 0.449*ATP6AP2, of which the coefficients were calculated by the aforementioned multivariate Cox regression analysis. We confirm the prognostic value of the risk score model in glioma patients. We used six independent glioma datasets to investigate the OS between low- and high-risk score groups and the AUC value of ROC curves was used to examine the

accuracy of our risk score model. We found that glioma patients with high-risk scores had poor prognosis compared with low-risk score groups. The AUC value for 1-, 3- and 5-year OS in the TCGA dataset was 85.8%, 90.7% and 83.7% (Figure 8(a)). Similar results that the risk score model performed better in predicting OS of glioma patients were also found in other independent datasets (CGGA-mRNA693, CGGA-

mRNA325, CGGA-mRNA301, Gravendeel and Rembrandt) (Figure 8(b-f)). Given the prognostic value of our risk score model in glioma patients, we further used multivariate Cox regression analysis to explore the independent role of this risk score model in glioma patients. Using the TCGA dataset, after adjusting age, gender, IDH-1 status and 1p19q co-deletion status, the risk score still played an independent prognostic role in glioma patients (Supplementary Table 4). A consistent result was also found in the CGGA-325 dataset (Supplementary Table 5). For providing clinicians with a convenient method to predict OS of glioma patients, we made a nomogram that combined the risk scores with different clinical parameters. As shown in Figure 8(g,h), the nomogram could be used to evaluate the 1-, 3- and 5-year OS. All the results reveal that our risk score model can be efficiently applied to predict glioma prognosis.

High-risk scores predict malignant phenotypes of glioma patients and treatment resistance

To elucidate the predictive function of our risk score model on malignant phenotypes of glioma patients, we used the TCGA-GBMLGG, CGGA mRNA-325, CGGA mRNA-693 and CGGA mRNA-301 datasets to compare risk scores with glioma pathology, IDH-1 status and 1p19q co-deletion and survival status. In the TCGA-GBMLGG dataset, we found that glioma patients diagnosed as GBM, 1p19q non-codeletion status and IDH-1 wild type had higher-risk scores than their counterparts; concurrently, glioma patients who were in alive status had lower-risk scores than the death (Figure 9(a)). We further explored its effects on treatment results. For glioma patients who received chemotherapy, the higher risk scores predicted shorter OS (Figure 9(b)). In addition,
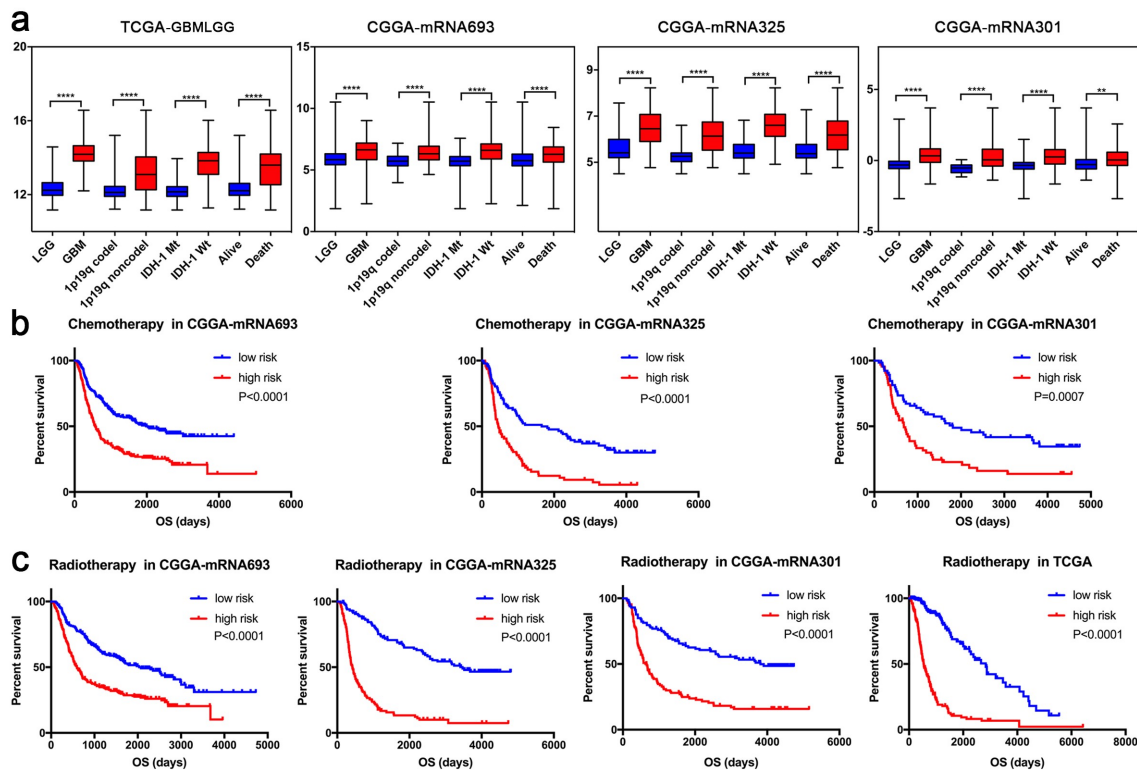

Figure 9. High-risk scores indicate malignant phenotypes of glioma patients and treatment resistance. (a) Analyses in multiple datasets showed consistent results that glioma patients with high-risk scores presented GBM, 1p19q non-codeletion, IDH-1 wild type and high risk of death. (b) Glioma patients with high-risk scores had shorter OS in chemotherapy groups. (c) Glioma patients with high-risk scores had shorter OS in radiotherapy groups. LGG: lower-grade glioma; GBM: glioblastoma; IDH: isocitrate dehydrogenase. *:P < 0.05; **:P < 0.01; ***:P < 0.001;****:P < 0.0001.
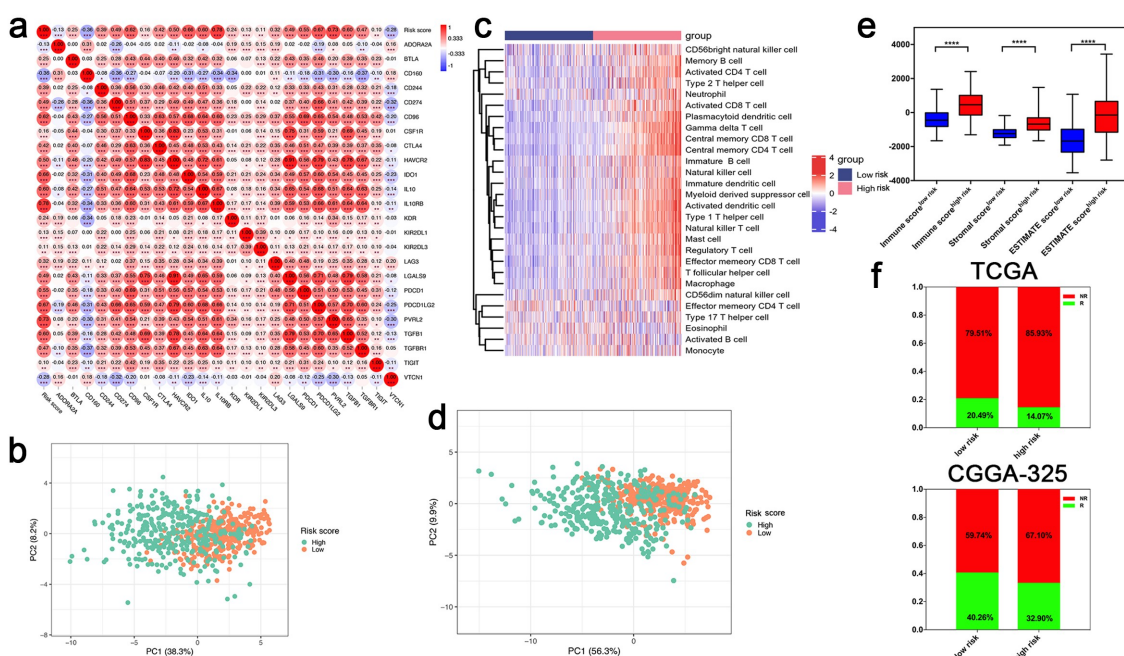
Figure 10. High-risk scores in glioma patients indicated high immune cells infiltration and immune checkpoints blockade (ICB) treatment resistance. (a,b) The correlation analysis showed that risk scores were positively correlated with immune checkpoints and immunoinhibitory molecules expression, PCA showed that immune checkpoints and immunoinhibitory molecules in high- and low-risk score groups had different expression profiles. (c,d) Glioma patients in high-risk score group had more TIL infiltration by ssGSEA, and PCA showed two groups had distinct immune cells infiltration levels. (e) Glioma patients with high-risk scores had higher immune, stromal and ESTIMATE scores. (f) Through ImmuCellAI method, patients in high-risk score group had lower "response" rate to ICB treatment compared with their counterparts in the TCGA-GBMLGG and CGGA-325 datasets. *:P < 0.05; **:P < 0.01; ***: P < 0.001;****:P < 0.0001. NR: non-response; R: response.

similar results were also found in radiation therapy groups (Figure 9(c)). These results indicate that our risk score model not only predicts malignant phenotypes of glioma patients but also promotes traditional therapy resistance.

High-risk scores indicated more immune cells infiltration and immunotherapy resistance

Immune checkpoints and inhibitory molecules play important roles in cancer immune suppression, and their inhibitors are promising strategies for cancer treatment [5,41]. We investigated the relationship between the risk score and the expression of critical immune checkpoints and inhibitory molecules in the TCGA dataset by Pearson correlation analysis. We found that the risk scores showed significant correlations with the expression of PDCD1, CD274, CTLA-4, LAG-3 and other immune inhibitory molecules; however, only the expression of ADORA2A (Pearson's r = −0.13, P < 0.001), CD160 (Pearson's r = −0.36, P < 0.001) and VTCN1 (Pearson's

r = −0.28, P < 0.001) was negatively correlated with risk scores (Figure 10(a)). Therefore, high-risk scores might present an immune resistance microenvironment in glioma patients. By the PCA, the expression of these immune inhibitory molecules could be separated into two clusters based on risk score groups (Figure 10(b)). We further used the ssGSEA to calculate 28 types of immune cell infiltration scores in the TCGA glioma samples, which showed that patients in the high-risk score group had higher immune cells infiltration (Figure 10(c)). The immune cell infiltration scores could be subdivided into different directions by PCA, which further confirmed that the risk scores in distinct groups presented different immune microenvironments (Figure 10(d)). The ESTIMATE method was then used to evaluate immune scores in glioma patients: the higher the risk scores, the higher the immune infiltration levels in glioma patients. Glioma patients in the high-risk score group had higher immune, stromal and ESTIMATE scores (Figure 10(e)).

In view of the positive correlation between risk scores and immunoinhibitory molecular

expression, we used the ImmuCellAI method to predict ICB treatment effects in the TCGA-GBMLGG and CGGA mRNA-325 datasets. The results revealed that glioma patients in the high-risk score groups had a lower rate of "response" to ICB treatment than those in the low-risk score group (Figure 10(f)). Taken together, the risk score model was not only associated with high levels of immune cell infiltration but may also become a valuable indicator for immunotherapy in the future.

Discussion

The identification of key molecules in gliomas has provided superior diagnostic and prognostic values, and as a result, our understanding of glioma behavior has rapidly evolved [3]. Mutations in isocitrate dehydrogenase 1 and 2 are present in the majority of adult patients with grade II and III gliomas when combined with 1p/19q codeletion status for classification; the prognostic distinction between grade II and III gliomas is diminished; meanwhile, the WHO grade II and III gliomas have been classified as "Lower-grade glioma" [3,45]. However, for heterogeneous cancers, the prognosis of GBM is quite different from other types of gliomas, and rapid recurrence after standard treatment is still inevitable. The complicated pathophysiological features and changeable TME of glioma have made multiple treatment strategies ineffective, including the current emerging immunotherapy [41]. Therefore, novel indicators are urgently needed to overcome these difficulties.

The V-ATPase is a multi-protein complex that catalyzes the ATP-dependent transport of protons across intracellular and plasma membranes. The resulting acidification of organelle lumens and the extracellular space will promote multiple cancer-related processes and signal pathway activation [10], as well immunosuppressive cells infiltration and treatment resistance will occur in cancers [9,10,46]. Inhibition of specific subunits of the V-ATPase has exhibited promising in suppression of migration and invasiveness in different cancers and reduction of cancer growth and metastasis [10,47,48]. To identify the key subunits of the V-ATPase in glioma patients, we analyzed the coding genes and chaperones of the V-ATPase in

glioma patients and used a series of bioinformatics analyses to identify and validate the hub genes among multi-omics datasets.

A growing number of evidence has indicated that the V-ATPase is related to invasiveness, metastasis, angiogenesis, proliferation, tumorigenesis, and drug resistance in tumors [10,12,47,48], therefore, is emerging as a potential drug target in cancer therapy. In highly metastatic and invasive cancers, the V-ATPase is at the plasma membrane of cancer cells, and preferential to be used to maintain their intra- and extracellular pH balance [48], which has been proposed to contribute alkalization of the tumor cytoplasm and to the acidification of the extra-cellular TME. Several specific subunits of the V-ATPase were overexpressed by cancer cells and their inhibitors might be novel targets for cancer treatment to overcome the adverse effects of nonselective inhibition of V-ATPase had brought [10,11,47,48]. In the present study, we identified five genes deriving from different subunits of the V-ATPase showing independent prognostic value in glioma patients and finally were regarded as hub genes. Molecular functions and clinical roles of these five genes in cancers have been reported. Silencing of ATP6AP1 and ATP6AP2 *in vitro* resulted in impaired vesicle acidification, redistribution of endosomal compartments, and accumulation of intra-cytoplasmic granules, recapitulating the cardinal phenotypic characteristics of granular cell tumors and providing a novel genotypic-phenotypic correlation [49]. High expression of TCIRG1 (V0a3 subunit) was found in breast cancer, hepatocellular carcinoma (HCC) and glioma patients [11,47,48]. TCIRG1 knockdown suppressed tumor cell growth and proliferation in HCC cell lines and also inhibited the metastatic potential of HCC cells by selectively regulating the epithelial-mesenchymal transition (EMT) [47]. For GBM patients who had higher expression of TCIRG1 presented a significant infiltration of immunosuppressive cells and also indicated worse prognosis [11]. Knockdown of V0a3 subunit significantly decreased migration of breast cancer MMB231 cell line [50]. High expression of ATP6V1C2 had been reported in colorectal and ovarian cancers [51], its high expression predicted shorter OS for renal clear carcinoma patients [51], and our findings confirmed that glioma patients

with high expression of ATP6V1C2 had worse prognosis. In addition, McConnell M and colleagues found that in breast cancers, overexpressed ATP6V1C2 was not significant and had no effects on OS [52]. However, its biological functions in cancers have not been fully elucidated. ATP6V1G2 was less reported in cancers, and the roles of ATP6V1G2 in cancers were probably diverse, which presented adverse effects on colorectal cancers [53] but showed protective effects on colon cancers [54]. Our results revealed its protective roles in glioma patients and highlighted its valuable application in cancer treatment.

Compared with most predictive models that were based on unexplored genes, which was difficult for clinicians to capture the specific function of members in the prognostic signature and might result in failing to detect the potential targets obtained from signature for treatment development. To address this issue, we developed a function-specific signature that may better explain the innate function of related genes and further confirm the potential therapeutic value in clinical. In the present study, we focused on identifying and validating a gene signature deriving from the V-ATPase subunits and primarily exploring the molecular functions of this gene signature. As reported by previous studies [10,46], the abnormal expression of V-ATPase accompanied with tumor immune microenvironment change, however, the specific subunits expression in different immune cells was unknown. Through the scRNA-seq technology, we found that mRNA levels of hub genes (ATP6AP1, ATP6AP2, TCIRG1 and ATP6V1G2) might be immune cell-type markers for gliomas, otherwise, the tight correlations between hub genes expression and TILs further confirmed this hypothesis. We deem that these genes might become novel targets for immunotherapy in the future.

Other reported gene signatures have been widely reported in clinical studies, these signatures can be used not only to identify sub-classification of cancers but also to create risk score models to predict prognosis and treatment resistance. Based on gene expression of our signature, glioma patients could be divided into different groups that showed distinct prognoses and immune cell infiltration levels; importantly, we also identified

many cancer-related pathway activation in the C2 group, which possibly provided us with more insights into internal interactions of glioma patients and novel targets for cancer treatment. By Terrasi's study [13], a specific V-ATPase signature from GBM and recurrent gliomas could resolve the heterogeneous class of IDH-wild type lower-grade gliomas. Another study from Couta-Viera and colleagues suggested that a differential expression pattern of V-ATPase was found in different tumors, particularly, a high V0c1 subunit and a low V0c2 subunit expression pattern accurately discriminated esophageal carcinoma from normal tissues [12]. However, their studies have not investigated the potential regulatory system of specific subunits expression and their roles in TME alterations remain unknown. In this study, we revealed that DNA promoter methylation and CNV were two possible methods that regulated hub gene expression. Abnormal hypomethylation of hub genes mediated their upregulation and was associated with poor survival in several cancers, suggesting that the hypomethylation might be a driver that mediated carcinogenesis [55]. Otherwise, TFs are a type of molecules that play important roles in regulation of gene expression. By using the ChEA3 online platform that ranks TFs associated with user-submitted gene sets [40], we primarily obtained the gene expression regulatory network and found the top 20 TFs that might tightly participate in regulating hub gene expression. The biological processes of the 20 TFs by GO analysis were mainly enriched in gene expression regulation processes, such as regulation of transcription, DNA-templation, and regulation of RNA biosynthetic processes.

The correlation between hub gene expression and drug sensitivity was also identified, from which we found that the five hub genes might be targets for pharmacological treatment in cancers. Several previous studies have identified the resistant role of V-ATPase in mediating cancer treatment. Our drug sensitivity analysis identified the potential drugs that might modulate the hub genes, among which the expression of ATP6V1G2 showed consistently negative correlation with drug IC50, while other genes showed distinct correlation with drug IC50. Although the roles of ATP6V1G2 in cancers were diverse by

previous studies [53,54], our findings still supported that ATP6V1G2 was a protective factor in glioma patients and worthy of further investigation.

Many reported gene signatures could compose risk score models that have shown better performance in predicting prognoses and molecular phenotypes, and even as an indicator for evaluating immunotherapeutic effects in many cancers. Using univariate, LASSO, and multivariate Cox regression analyses, we identified these five hub genes and created a risk score model that effectively reflected malignant phenotypes of glioma patients, such as IDH-1 wild type, WHO grade and 1p19q non-codeletion status, and fully predicted prognoses. Otherwise, we used this risk score model in combination with other clinical parameters to create a novel nomogram that could be used by clinicians to predict glioma patients' 1-, 3- and 5-year OS conveniently. Importantly, these functions of risk score model were validated by multi-independent datasets, which could guarantee the accuracy of our risk score model in clinical applications.

The positive correlation between risk scores and most immune checkpoints and immunoinhibitory molecular expression highlighted the potential value of applying this valuable model to predict ICB treatment. The higher the risk scores, the higher immune cells infiltration levels and immune scores. In the setting of different risk scores (high vs. low), we found that immune gene expression profiles in the TCGA dataset were divided into two sections by PCA, indicating that the immune status of glioma patients in the low-risk group was quite different from those in the high-risk group, which further confirmed that our risk score model was an accurate indicator for evaluating local immune infiltration and a predictor for immunotherapy in glioma patients. The results from the ImmuCellAI analysis also supported our hypothesis.

Compared with previous studies, our study had several strengths. We comprehensively analyzed the V-ATPase in glioma patients and created a prognosis-related gene signature, the underlying mechanisms in regulation of hub genes expression and drug sensitivity analysis probably offered us novel targets for glioma therapy, importantly, this gene signature composed a risk score model that indicated an effective method in predicting gliomas' therapeutic resistance, as well immune cells infiltration level and ICB treatment.

Although the results were inspiring, several limitations still exist. First, protein level validation of the five hub genes was not adequate, which was only based on the HPA and CPTAC platforms; therefore, more clinical sample validation was urgently needed. Second, the retrospective studies with distinct sequencing platforms might inevitably introduce bias. The inconsistent results from HPA, TCGA-GTEx and CPTAC pointedly highlighted this problem. More prospective studies should be performed to overcome these difficulties. Third, the complex regulation system of five hub genes and their roles in TME should be explored by *in vivo* and *in vitro* experiments, which will constitute the focus of our next step research in the future. In addition, whether the value of our risk score model in predicting the prognosis of glioma patients relies on the immune system remains unknown. These unresolved challenges will require in-depth validation of the underlying mechanisms for the selected genes and their risk score system in the future.

Disclosure statement

Funding

Data availability statement

The data that support the findings of this study are openly available in the UCSC Xena browser (https://xenabrowser.net/datapages/), GlioVis (http://gliovis.bioinfo.cnio.es), CGGA (http://www.cgga.org.cn/) and other online platforms aforementioned.

References

[1] DeAngelis LM. Brain tumors. N Engl J Med. 2001;344 (2):114–123.

[2] Stupp R, Mason WP, van den Bent MJ, et al. Radiotherapy plus concomitant and adjuvant

temozolomide for glioblastoma. N Engl J Med. 2005;352(10):987–996.

[3] 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):803–820.

[4] Jia D, Li S, Li D, et al. Mining TCGA database for genes of prognostic value in glioblastoma microenvironment. Aging (Albany NY). 2018;10 (4):592–605.

[5] Lin W, Wu S, Chen X, et al. Characterization of hypoxia signature to evaluate the tumor immune microenvironment and predict prognosis in glioma groups. Front Oncol. 2020;10:796.

[6] Hirata E, Sahai E. Tumor Microenvironment and Differential Responses to Therapy. Cold Spring Harb Perspect Med. 2017;7(7):a026781.

[7] Osipov A, Saung MT, Zheng L, et al. Small molecule immunomodulation: the tumor microenvironment and overcoming immune escape. J Immunother Cancer. 2019;7(1):224.

[8] Brüx A, Liu TY, Krebs M, et al. Reduced V-ATPase activity in the trans-Golgi network causes oxylipin-dependent hypocotyl growth Inhibition in Arabidopsis. Plant Cell. 2008;20(4):1088–1100.

[9] Santos-Pereira C, Rodrigues LR, Côrte-Real M. Emerging insights on the role of V-ATPase in human diseases: therapeutic challenges and opportunities. Med Res Rev. 2021;41(4):1927–1964.

[10] Whitton B, Okamoto H, Packham G, et al. Vacuolar ATPase as a potential therapeutic target and mediator of treatment resistance in cancer. Cancer Med. 2018;7 (8):3800–3811.

[11] Qi C, Lei L, Hu J, et al. T cell immune regulator 1 is a prognostic marker associated with immune infiltration in glioblastoma multiforme. Oncol Lett. 2021;21(4):252.

[12] Couto-Vieira J, Nicolau-Neto P, Costa EP, et al. Multi-cancer V-ATPase molecular signatures: a distinctive balance of subunit C isoforms in esophageal carcinoma. EBioMedicine. 2020;51:102581.

[13] Terrasi A, Bertolini I, Martelli C, et al. Specific V-ATPase expression sub-classifies IDHwt lower-grade gliomas and impacts glioma growth in vivo. EBioMedicine. 2019;41:214–224.

[14] Chu A, Zirngibl RA, Manolson MF. The V-ATPase a3 subunit: structure, function and therapeutic potential of an essential biomolecule in osteoclastic bone resorption. Int J Mol Sci. 2021;22(13):6934.

[15] Goldman MJ, Craft B, Hastie M, et al. Visualizing and interpreting cancer genomics data via the Xena platform. Nat Biotechnol. 2020;38(6):675–678.

[16] Zhao Z, Zhang KN, Wang Q, et al. Chinese Glioma Genome Atlas (CGGA): a comprehensive resource with functional genomic data from Chinese Glioma Patients. Genomics Proteomics Bioinformatics. 2021;19(1):1–12.

[17] Bowman RL, Wang Q, Carro A, et al. GlioVis data portal for visualization and analysis of brain tumor expression datasets. Neuro Oncol. 2017;19(1):139–141.

[18] Wang Y, Qian T, You G, et al. Localizing seizure-susceptible brain regions associated with low-grade gliomas using voxel-based lesion-symptom mapping. Neuro Oncol. 2015;17(2):282–288.

[19] Liu X, Li Y, Qian Z, et al. A radiomic signature as a non-invasive predictor of progression-free survival in patients with lower-grade gliomas. Neuroimage Clin. 2018;20:1070–1077.

[20] Bao ZS, Chen HM, Yang MY, et al. RNA-seq of 272 gliomas revealed a novel, recurrent PTPRZ1-MET fusion transcript in secondary glioblastomas. Genome Res. 2014;24(11):1765–1773.

[21] Zhao Z, Meng F, Wang W, et al. Comprehensive RNA-seq transcriptomic profiling in the malignant progression of gliomas. Sci Data. 2017;4(1):170024.

[22] Fang S, Liang J, Qian T, et al. Anatomic Location of Tumor Predicts the Accuracy of Motor Function Localization in Diffuse Lower-Grade Gliomas Involving the Hand Knob Area. AJNR Am J Neuroradiol. 2017;38(10):1990–1997.

[23] Wang Y, Wang Y, Fan X, et al. Putamen involvement and survival outcomes in patients with insular low-grade gliomas. J Neurosurg. 2017;126(6):1788–1794.

[24] Gravendeel LA, Kouwenhoven MC, Gevaert O, et al. Intrinsic gene expression profiles of gliomas are a better predictor of survival than histology. Cancer Res. 2009;69(23):9065–9072.

[25] Madhavan S, Zenklusen JC, Kotliarov Y, et al. Rembrandt: helping personalized medicine become a reality through integrative translational research. Mol Cancer Res. 2009;7(2):157–167.

[26] Louis DN, Ohgaki H, Wiestler OD, et al. The 2007 WHO classification of tumours of the central nervous system. Acta Neuropathol. 2007;114(2):97–109.

[27] Thorsson V, Gibbs DL, Brown SD, et al. The immune landscape of cancer. Immunity. 2018;48(4):812–830.e14.

[28] Thul PJ, Åkesson L, Wiking M, et al. A subcellular map of the human proteome. Science. 2017;356(6340):eaal3321.

[29] Chandrashekar DS, Bashel B, Balasubramanya SAH, et al. UALCAN: a portal for facilitating tumor sub-group gene expression and survival analyses. Neoplasia. 2017;19(8):649–658.

[30] Sun D, Wang J, Han Y, et al. TISCH: a comprehensive web resource enabling interactive single-cell transcriptome visualization of tumor microenvironment. Nucleic Acids Res. 2021;49(D1):D1420–d1430.

[31] Neftel C, Laffy J, Filbin MG, et al. An integrative model of cellular states, plasticity, and genetics for glioblastoma. Cell. 2019;178(4):835–849.e21.

[32] Yoshihara K, Shahmoradgoli M, Martínez E, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. Nat Commun. 2013;4 (1):2612.

[33] Miao YR, Zhang Q, Lei Q, et al. ImmuCellAI: a unique method for comprehensive T-cell subsets abundance prediction and its application in cancer immunotherapy. Adv Sci (Weinh). 2020;7(7):1902880.

[34] Ru B, Wong CN, Tong Y, et al. TISIDB: an integrated repository portal for tumor-immune system interactions. Bioinformatics. 2019;35(20):4200–4202.

[35] Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009;4 (1):44–57.

[36] Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 2009;37(1):1–13.

[37] Metsalu T, Vilo J. ClustVis: a web tool for visualizing clustering of multivariate data using Principal Component Analysis and heatmap. Nucleic Acids Res. 2015;43(W1):W566–W570.

[38] Liu CJ, Hu FF, Xia MX, et al. GSCALite: a web server for gene set cancer analysis. Bioinformatics. 2018;34 (21):3771–3772.

[39] Warde-Farley D, Donaldson SL, Comes O, et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. Nucleic Acids Res. 2010;38(suppl_2):W214–W220.

[40] Keenan AB, Torre D, Lachmann A, et al. ChEA3: transcription factor enrichment analysis by orthogonal omics integration. Nucleic Acids Res. 2019;47(W1): W212–W224.

[41] Xu S, Tang L, Li X, et al. Immunotherapy for glioma: current management and future application. Cancer Lett. 2020;476:1–12.

[42] Pectasides E. Genomic alterations and targeted therapy in gastric and esophageal adenocarcinoma. Clin Ther. 2016;38(7):1589–1599.

[43] Jenzer M, Keß P, Nientiedt C, et al. The BRCA2 mutation status shapes the immune phenotype of prostate cancer. Cancer Immunol Immunother. 2019;68 (10):1621–1633.

[44] Wang H, You S, Fang M, et al. Recognition of immune microenvironment landscape and immune-related prognostic genes in breast cancer. Biomed Res Int. 2020;2020:3909416.

[45] Brat DJ, Verhaak RG, Aldape KD, et al. Comprehensive, Integrative Genomic Analysis of Diffuse Lower-Grade Gliomas. N Engl J Med. 2015;372(26):2481–2498.

[46] Katara GK, Kulshrestha A, Jaiswal MK, et al. Inhibition of vacuolar ATPase subunit in tumor cells delays tumor growth by decreasing the essential macrophage population in the tumor microenvironment. Oncogene. 2016;35(8):1058–1065.

[47] Yang HD, Eun JW, Lee KB, et al. T-cell immune regulator 1 enhances metastasis in hepatocellular carcinoma. Exp Mol Med. 2018;50(1):e420.

[48] Hinton A, Sennoune SR, Bond S, et al. Function of a subunit isoforms of the V-ATPase in pH homeostasis and in vitro invasion of MDA-MB231 human breast cancer cells. J Biol Chem. 2009;284(24):16400–16408.

[49] Pareja F, Brandes AH, Basili T, et al. Loss-of-function mutations in ATP6AP1 and ATP6AP2 in granular cell tumors. Nat Commun. 2018;9(1):3533.

[50] Cotter K, Liberman R, Sun-Wada G, et al. The a3 isoform of subunit a of the vacuolar ATPase localizes to the plasma membrane of invasive breast tumor cells and is overexpressed in human breast cancer. Oncotarget. 2016;7(29):46142–46157.

[51] Li X, Li H, Yang C, et al. Comprehensive Analysis of ATP6V1s Family Members in Renal Clear Cell Carcinoma With Prognostic Values. Front Oncol. 2020;10:567970.

[52] McConnell M, Feng S, Chen W, et al. Osteoclast proton pump regulator Atp6v1c1 enhances breast cancer growth by activating the mTORC1 pathway and bone metastasis by increasing V-ATPase activity. Oncotarget. 2017;8(29):47675–47690.

[53] Shao Y, Jia H, Huang L, et al. An Original ferroptosis-related gene signature effectively predicts the prognosis and clinical status for colorectal cancer patients. Front Oncol. 2021;11:711776.

[54] Huang C, Zhao J, Zhu Z. Prognostic nomogram of prognosis-related genes and clinicopathological characteristics to predict the 5-year survival rate of colon cancer patients. Front Surg. 2021;8:681721.

[55] Pérez RF, Tejedor JR, Bayón GF, et al. Distinct chromatin signatures of DNA hypomethylation in aging and cancer. Aging Cell. 2018;17(3):e12744.