Exploring the potential biomarkers for prognosis of glioblastoma via Weighted Gene Co-expression Network Analysis

Mengyuan Zhang [1], **Zhike Zhou** [2], **Zhouyang Liu** [1], **Fangxi Liu** [1], **Chuansheng Zhao** [Corresp. 1]

[1] Department of Neurology and Stroke Center, The First Hospital of China Medical University, Shenyang, China

[2] Department of Geriatrics, The First Hospital of China Medical University, Shenyang, China

Corresponding Author: Chuansheng Zhao
Email address: cszhao@cmu.edu.cn

Background: Glioblastoma (GBM) is the most common malignant tumor in the central system with a poor prognosis. Due to the complexity of its molecular mechanism, the recurrence rate and mortality rate of GBM patients are still high. Therefore, there is an urgent need to screen GBM biomarkers to prove the therapeutic effect and improve the prognosis. Results: We extracted data from GBM patients from the Gene Expression Integration Database (GEO), analyzed differentially expressed genes in GEO and identified key modules by weighted gene co-expression network analysis (WGCNA). GSE145128 data was obtained from the GEO database, and the darkturquoise module was determined to be the most relevant to the GBM prognosis by WGCNA (r = -0.62, p=0.01). We performed enrichment analysis of gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) to reveal the interaction activity in the selected modules. Then Kaplan-Meier survival curve analysis was used to extract genes closely related to GBM prognosis. We used Kaplan-Meier survival curves to analyze the 139 genes in the darkturquoise module, identified four genes (DARS / GDI2 / P4HA2 / TRUB1) associated with prognostic GBM. Low expression of DARS/GDI2/TRUB1 and high expression of P4HA2 had a poor prognosis. Finally, we used tumor genome map (TCGA) data, verified the characteristics of hub genes through Co-expression analysis, Drug sensitivity analysis, TIMER database analysis and GSVA analysis. We downloaded the data of GBM from the TCGA database, the results of co-expression analysis showed that DARS/GDI2/P4HA2/TRUB1 could regulate the development of GBM by affecting genes such as CDC73/CDC123/B4GALT1/CUL2. Drug sensitivity analysis showed that genes are involved in many classic Cancer-related pathways including TSC/mTOR, RAS/MAPK.TIMER database analysis showed DARS expression is positively correlated with tumor purity(cor=0.125,p=1.07e-02)), P4HA2 expression is negatively correlated with tumor purity(cor=-0.279,p=6.06e-09). Finally, GSVA analysis found that DARS/GDI2/P4HA2/TRUB1 gene sets are closely related to the

occurrence of cancer. Conclusion: We used two public databases to identify four valuable biomarkers for GBM prognosis, namely DARS/GDI2/P4HA2/TRUB1, which have potential clinical application value and can be used as prognostic markers for GBM.

Exploring the potential biomarkers for prognosis of glioblastoma via Weighted Gene Co-expression Network Analysis

Mengyuan Zhang [1], Zhike Zhou[2], Zhouyang Liu[1], Fangxi Liu[1] and Chuansheng Zhao [1]*

1 Department of Neurology and Stroke Center, The First Hospital of China Medical University, Shenyang 110000, PR China.

2 Department of Geriatrics, The First Affiliated Hospital, China Medical University, Shenyang 110000, PR China.

Corresponding Author: Chuansheng Zhao

The First Hospital of China Medical University, No. 155 Nanjing North Street, Heping District Shenyang 110000, PR China

Email Adress: cszhao@cmu.edu.cn

Abstract:

Glioblastoma (GBM) is the most common malignant tumor of the central system, with a poor prognosis.Due to the complexity of its molecular mechanism, the recurrence rate and mortality rate of GBM patients remain high. Therefore, screening of biomarkers for GBM is urgently needed to demonstrate therapeutic efficacy and improve prognosis. In this study, we extracted GBM patients' data from gene expression integration database (GEO), analyzed the differentially expressed genes in GEO by Weighted gene co-expression network analysis (WGCNA), constructed the co-expression network, and determined the correlation with the key recurrent modules of GBM. At the same time, based on Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG), the selected modules were analyzed. Then four genes (DARS / GDI2 / P4HA2 / TRUB1) which are closely related to the prognosis of GBM were extracted by Kaplan-Meier survival curve analysis. The characteristics of these four genes were verified by tumor genome atlas (TCGA) data, Co-expression analysis, Drug sensitivity analysis, TIMER database analysis and Gene set variation analysis (GSVA) analysis. It was found that these four genes were differentially expressed genes in the initiation and progression of GBM, which could provide reference and basis for the observation of the clinical treatment and prognosis of GBM.

Introduction

Glioblastoma (GBM) is the most common primary neurogenic tumor, and the prognosis of most subtypes is poor(Tan et al. 2020). Despite of aggressive treatment strategies such as surgery followed by irradiation and chemotherapy, the prognoses of GBM patients remained unsatisfactory (Wu et al. 2020). According to the existing data, GBM patients have a survival of only 12–15 months after the standard treatment, with the 5-year survival rate of 3–5% (Gong et al. 2020; Szopa et al. 2017). The main reasons for the poor prognosis of GBM are due to tumor metastasis and postoperative recurrence(Tij et al. 2021). Given that tumors invade the brain aggressively, GBM tumors can rarely be completely removed by surgery(Reichel et al. 2020). And the resulting network by GBM enables multicellular communication through microtube-associated gap junctions, and increases tumor resistance to cell ablation and radiotherapy(Li et al. 2017a). Actively searching for biological markers related to the treatment and prognosis of GBM patients is of great significance for improving the survival rate of GBM patients.

In the past few decades, gene sequencing and bioinformatics analysis have been widely used for genetic variation screening at the gene level(Tingting et al. 2019), which helps us to identify differentially expressed genes (DEG) and functional pathways in the development of GBM. It has been found that the increased expression of SPRY2 mRNA indicates the decreased survival rate of GBM patients(Li et al. 2017a). Another study showed that the mRNA levels of NOTCH and Epidermal Growth Factor Receptor (EGFR) genes were increased in GBM tissues, which was related to the survival of patients(Irshad et al. 2015; Xing et al. 2015). However, most of these studies are single gene analysis, which may limit the analysis of the pathogenesis and prognosis of GBM.

Weighted gene co-expression network analysis (WGCNA) is a platform to identify hub genes or therapeutic targets based on the interconnectivity of gene subsets and the association between gene subsets and phenotypes(Wang et al. 2020; Zhang & Horvath 2005). WGCNA can use the information of thousands of genes to identify the gene modules of interest and perform important

association analysis on phenotypes. Recently, many journals have published relevant studies using WGCNA(Schafer et al. 2019; Wang et al. 2020; Zhou et al. 2018).

In this study, we extracted four GBM related biomarkers (DARS / GDI2 / P4HA2 / TRUB1) by extracting data from GBM patients from the gene expression integrated database (GEO) and using WGCNA and Kaplan-Meier survival curves analysis. Then, we established GBM gene markers in the tumor genome atlas (TCGA), and confirmed the characteristics of these four genes by means of Co-expression analysis, Drug sensitivity analysis, TIMER database analysis , and GSVA analysis. In summary, our purpose is to find reliable biomarkers related to the prognosis of GBM by analyzing the relationship between DARS / GDI2 / P4HA2 / TRUB1 gene and GBM, so as to provide reference and basis for clinical treatment and prognosis observation of GBM.

Materials and methods

Data information and construction of WGCNA

The Series Matrix File data File of GSE145128 was downloaded from the NCBI GEO public database, which were contained 15 GBM patients and sets of transcriptional data, including untreated group (n=7) and recurrent group (n=8), for the construction of WGCNA co-expression network.

We constructed a weighted gene co-expression network to find co-expressed gene modules, and clarified the relationship between the gene network and phenotype and hub genes. The WGCNA-R package was used to construct a co-expression network of genes in the GSE145128 dataset, where the soft-thresholding power was set to 16. The weighted adjacency matrix is converted into a topological overlap matrix (TOM) to estimate the network connectivity, and a hierarchical clustering method is used to construct a clustering tree structure of the TOM matrix. Different gene modules are represented by different cluster tree branches and colors. All genes are divided into multiple modules through gene expression patterns, and genes with similar expression patterns are divided into one module based on weighted correlation coefficients and expression patterns.

Enrichment analysis of gene module function

In order to obtain the biological functions and signaling pathways involved in the interest module of WGCNA, the Metascape database (www.metascape.org) was used for annotation and visualization, and Gene Ontology (GO) analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses were performed on the genes in the specific module. Min overlap$\geq$3 & p$\leq$0.01 was considered statistically significant.

Identifying Hub Genes

To determine hub genes, the most important thing is whether they have an impact on tumor prognosis. According to WGCNA theory, hub genes have the highest connectivity in the module, which determines the biological significance of the module(Chen et al. 2019). So we think that hub genes must exist in the interested module of WGCNA. Combined with the above two points, we analyzed the Kaplan-Meier survival rate of all genes in the interest module of WGCNA. We

believe that the genes that can affect the prognosis of GBM patients are the hub genes. And then,the next step is to explore and verify the specific molecular mechanism of hub genes.

Download and Pre-processing Data From TCGA

TCGA database as the biggest cancer gene information database, including gene expression data, the miRNA expression data and copy number variation, DNA methylation, SNPS and other data. We downloaded the processed original mRNA expression data of GBM. A total of 159 specimens were collected(Supplementary table S1).

Co-expression analysis

The co-expression of the hub genes were analyzed. The correlation coefficient filter condition was 0.3 and the p-value was 0.001. After screening the genes with the most significant expression of hub genes, the correlation analysis circles of hub genes were plotted using "corrplot" and "circlize" packages.

GSCALite and GDSC

GSCALite is a set analysis platform for cancer genes. It integrates cancer genomics data of 33 cancer types from TCGA, drug response data from GDSC and CTRP, and normal tissue data from GTEx, and conducts gene set analysis in the data analysis process. Our study through this analysis was carried out on the hub genes.Then base on the largest publicly available pharmacogenomics database (GDSC, the Genomics of Drug Sensitivity in Cancer, https://www.cancerrxgene.org/), we used the R packet "pRophetic" to predict the chemosensitivity of each tumor sample, and the estimated IC_ (50) of each specific chemotherapeutic drug treatment was obtained by ridge regression. The prediction accuracy was measured by 10-fold cross-validation with the GDSC training set. Select default values for all parameters, including "combats" to remove batch effects, "allSoldTumours" for tissue types, and average values for summarizing repetitive gene expressions(Liu et al. 2018).

TIMER database analysis

TIMER is a website for systematically testing the molecular characteristics of tumor-immune interactions(Li et al. 2017b). This website has incorporated 10,897 samples ranging from 32 different kinds of cancer types from the TCGA dataset(Shi et al. 2020). In this study, TIMER was used to explore the relationship between hub genes and the contents of immune cells and to compare the infiltration levels between tumors with different somatic copy number changes of hub genes.

Gene functional analysis

GSVA uses a non-parametric and unsupervised method, and bypasses the traditional method of explicitly modeling phenotypes in affluent scoring algorithms(Hanzelmann et al. 2013). By

comprehensively scoring the gene set of interest, GSVA converts the gene level change into the pathway level change, and then judges the biological function of the sample. In this study, the gene sets were downloaded from the Molecular signatures database (v7.0 version),and used the algorithm of GSVA comprehensive score of each gene set, evaluating the potential biological function change different samples.

Statistical analysis

All statistical analyses were performed in R language (version 3.6). All the statistical tests were bilateral, and $p < 0.05$ was statistically significant.

Result

Identification of gene co-expression modules

The Series Matrix File data File of GSE145128 was downloaded from the NCBI GEO public database. A total of 15 transcriptional data sets, including untreated group (n=7) and recurrent group (n=8), were used to construct the WGCNA co-expression network. In order to determine whether the 15 samples in GSE145128 were suitable for network analysis, a sample dendrogram and similar clinical features were studied. We confirmed that all samples were included in the group (Fig. 1A). The soft-thresholding power was set as 16 for the subsequent construction of co-expression network (Fig. 1B, Fig. 1C). The clustering tree structure of TOM matrix was constructed by hierarchical clustering method. The different branches and colors represent different gene modules (Fig. 1D). The network heatmap was used to analyze the interaction of 41 modules (Fig. 1E). The results showed that each module was independent of each other, indicating that each module was highly personalized and the gene expression of each module was relatively independent.

Correlation of modules and clinical traits

In order to study the relationship between these modules and the prognosis of GBM, we investigated the correlation between each module and the prognosis of GBM.We found that the darkturquoise module had the highest correlation with disease relapse（$r = -0.62$, $p=0.01$）（Fig. 2A, B）. We used Metascape to analyze the function and pathway of the darkturquoise module. Metascape can identify the enrichment process in the gene list and the association between enrichment processes [14, 15] by querying many databases,such as GO functional, Hallmark Gene Sets, and KEGG pathways (Tripathi et al. 2015; Zheng et al. 2018). Based on GO enrichment analysis, it was found that the co-expressed genes within the modules of interest s mainly related to the steady state of cellular transition metal ion homeostasis, protein hydroxylation, intrinsic apoptotic signaling pathway, ncRNA metabolic process(Fig. 2C).The KEGG pathway analysis revealed that the co-expressed genes within the modules of interest was mostly enriched in the 'Ferroptosis'(Fig. 2C). In addition, the enrichment processes were highly connected and could be clustered into a complete network (Fig. 2C) . These results indicated that these functions were related in the occurrence and development of GBM.

Identification of hub genes in darkturquoise module

According to the WGCNA theory, hub genes have the highest connectivity in the module, which determine the biological significance of the module (Chen et al. 2019).Therefore, we searched for hub genes in the darkturquoise module.We analyzed the Kaplan-Meier survival curves of 139 genes in the darkturquoise module. 14 genes with significant survival analysis results ($p<0.05$) were selected for sequencing(Table. 1). Finally, we found that only four genes (DARS/GDI2/P4HA2/TRUB1) had an impact on the prognosis of GBM patients (Fig. 3). Survival analysis showed that the patients with low expression of DARS/GDI2/ TRUB1 and high expression of P4HA2 had poor prognosis.

Analysis of the co-expression of hub genes

It was clear that the hub genes can affect the process of disease progression by regulating related genes. It can be assumed that DARS/GDI2/P4HA2/TRUB1 was associated with the most abundant pathways and genes and could regulate more biological processes. In order to assess the gene correlation of DARS / GDI2 / P4HA2 / TRUB1, we analyzed the co-expression of DARS/GDI2/P4HA2/TRUB1 through Pearson correlation analysis($cor>0.3$, $p<0.001$). We screened the 10 genes with the strongest correlation with the expression of DARS / GDI2 / P4HA2 / TRUB1, drew the correlation analysis map and heat map of DARS / GDI2 / P4HA2 / TRUB1 (Fig. 4A-4H), and found that the correlation between DARS and CDC73 was the highest, and the correlation between GDI2 and CDC123 was the highest. P4HA2 and B4GALT1 have the highest correlation, TRUB1 and CUL2 have the highest correlation (Fig. 4I-L). Among them, CDC73 and CDC123 are cyclins of cell division(Sun et al.), B4GALT1 is one of seven β - 1, 4-galactosyltransferases (B4GALT). CUL2 contributes to form E3 ubiquitin ligase that can recognize numerous substrates and is involved in a variety of cellular processes(Liu et al. 2019). These four genes have been shown to have a close relationship with many kinds of cancers(Cao et al. 2020a; Dou et al. 2020; Li et al. 2019a), such as thyroid carcinoma (Sarquis et al. 2019), breast cancer, etc.

Cancer-related pathways and drug sensitivity analysis of hub genes.

First, we investigated the role of hub genes in all well-known cancer-related pathways, as the following: TSC/mTOR, RTK, RAS/MAPK, PI3K/AKT, Hormone ER, Hormone AR, EMT, DNA Damage Response, Cell Cycle, Apoptosis pathways. The results found that DARS participated in the TSC/mTOR pathway activation; GDI2 was involved in Apoptosis, TSC/mTOR pathway activation; P4HA2 was involved in DNA Damage Response, EMT, Hormone AR, Hormone ER, RAS/MAPK and TSC/mTOR pathway; TRUB1 was involved in Apoptosis, DNA Damage Response, EMT, Hormone AR, PI3K/AKT and TSC/mTOR pathway (Fig. 5A). To investigate whether the expression of DARS / GDI2 / P4HA2 / TRUB1 in GBM had an impact on treatment (e.g. chemotherapies), we constructed a predictive model on six commonly used chemo drugs (i.e.

AKT.inhibitor, Cisplatin, Dasatinib, Erlotinib, Gefitinib, and Gemcitabine) and confirmed that high expression of DARS was less sensitive to Cisplatin(p=0.00026) and Gemcitabine(p=0.0024), high expression of GDI2 was less sensitive to AKT.inhibitor(p=5.2e−05), Cisplatin(p=0.00067), Dasatinib(p=0.012) and Gemcitabine(p=3.5e−06), low expression of P4HA2 was less sensitive to Cisplatin(p=0.0032), and Gemcitabine(p=0.00017),and high expression of TRUB1 was less sensitive to AKT.inhibitor(p=0.00043)(Fig.5B).

Immune cells infiltration analysis

In view of the obvious prognostic value of DARS/GDI2/P4HA2/TRUB1 gene, we used the TIMER database to determine whether there was an association between tumor-infiltrating and immune cells and DARS/GDI2/P4HA2/TRUB1 expression.Results showed that DARS expression was positively correlated with tumor purity, P4HA2 and B cells (partial cor=-0.239, p=7.89e-07), P4HA2 and CD8+ T cells (partial cor=-0.158, p=1.19e-03), TRUB1 and CD8+ T cells (partial cor=-0.206, p=1.87e-02), DARS and neutrophils (partial cor= 0.245, p=3.73e-07), GDI2 and neutrophils (partial cor=0.184, p=1.62e-04), GDI2 and Dendritic cells (partial cor=0.167, p=6.21e-04) P4HA2 and B cells (partial cor=-0.239, p=7.89e-07), P4HA2 and CD8+ T cells (partial cor=-0.158, p=1.19e-03), TRUB1 and CD8+ T cells (partial cor=-0.206, p=1.87e-02), DARS and neutrophils (partial cor= 0.245, p=3.73e-07), GDI2 and neutrophils (partial cor=0.184, p=1.62e-04), GDI2 and Dendritic cells (partial cor=0.167, p=6.21e-04) (Fig. 6A). We also explored the correlation between tumor immune cell infiltration and somatic copy number alterations (SCNAs). The samples were divided into four types according to the copy number of genes.The distribution of infiltrating immune cells among the four types of samples was compared, as shown in Fig. 6B. We found that the various forms of mutations carried by the DARS / GDI2 / P4HA2 / TRUB1 gene can usually suppress immune infiltration, including CD8+T cells, neutrophils, dendritic cells, macrophages, CD4+T cells, and B cells. Also, we found that these four pivotal genes had a greater effect on immune infiltration than alterations in the genes.

Genomic alterations of DARS/GDI2/P4HA2/TRUB1 in GBM

We then used the cBioPortal tool to determine the types and frequency of DARS/GDI2/P4HA2/TRUB1 alterations based on DNA sequencing data from GBM patients. The genetic variation rates of DARS/GDI2/P4HA2/TRUB1 ranged from 0% to 4%(DARS was 4%, GDI2 was 1.4%, P4HA2 was 4%, TRUB1 was 0.0%.) These alterations include Missense Mutation, mRNA High, mRNA Low, Amplification (AMP), and Deep Deletion.(Fig. 7) In view of this, DARS and P4HA2 show potentially stronger cancer-driving properties at a higher mutation frequency. In contrast, TRUB1 is genetically stable and could potentially act as a stable biomarker.

Gene functional analysis

We downloaded the DARS/GDI2/P4HA2/TRUB1 gene sets from the Molecular signatures database (v7.0 version) and comprehensively evaluated the gene sets through GSVA. Our analysis showed that in the DARS gene set, 17 gene sets were up-regulated (t>, 1) and 14 gene sets were down-regulated (t<1). In GDI2, 13 gene sets were up-regulated and 21 gene sets were down-regulated. In P4HA2, 11 gene sets were up-regulated and 30 gene sets were down-regulated. In TRUB1, 21 gene sets were down-regulated and 14 gene sets were down-regulated. (Fig. 8A - D).

Discussion

Due to the complex mechanisms of GBM,it is one of the most threatening CNS malignancies. Therefore, it is an urgent need to find biomarkers related to the occurrence and prognosis of GBM to reveal the possible pathogenesis or predict the prognosis of patients, and then develop personalized treatment plans for GBM patients. Based on gene sequencing technology, we have discovered some biological markers with predictive value for patients including GBM. However, the role of these markers are still limited. In order to better understand GBM, there is an urgent need to screen out more biomarkers to improve the efficacy of GBM treatment and prognosis.

GBM, as a highly heterogeneous tumor harboring multiple genetic alterations(Harter et al. 2014), molecular heterogeneity affects the effectiveness of single-molecule markers in predicting prognosis(Tonry 2020). At the same time, some studies have found that the high recurrence rate of GBM is related to the expression of strong proliferation genes of cells(Lara-Velazquez et al. 2020). And these processes usually involve multiple genes(Malik et al. 2020). Therefore, we believe that multi-gene markers have a higher predictive power for GBM prognosis than single-gene marker. We built a multi-gene markers model for predicting GBM prognosis, and validated the multi-gene markers model through strategies including training, testing, and independent cross-validation. The above strategies significantly improve the predictive ability of genetic markers (Li et al. 2019b).

In our research and analysis, the results of GO and KEGG analysis indicate that cell transition metal ion homeostasis, protein hydroxylation, intrinsic apoptotic signaling pathway and other processes may play an important role in GBM. Among them， transition metals are critical for many metabolic processes(Nelson & N. 2014), and their steady state is vital to life. Aberrations in the cellular metal ion concentrations may lead to cell death and severe diseases such as cancer(Pi et al. 2020). Hydroxylation is a post-translational modification affecting protein stability, activity or interactome(Zurlo & Zhang 2020). Many cancers are related to protein hydroxylation, such as breast cancer(Zurlo & Zhang 2020), gastric cancer(Li et al. 2020), and prostate cancer(Della-Flora et al. 2020). For example, a study found that a set of enzymes PLOD1、PLOD2 and PLOD3 involved in the hydroxylation of lysine and stabilization of collagen by crosslinks, which up-regulated expression in gastric cancer patients(Li et al. 2020). Similarly, intrinsic apoptotic signaling pathway can activate or inactivate multiple signaling pathways and inhibit multiple tumor suppressor genes, thereby promoting tumor progression. Almost all cancers involve intrinsic apoptotic signaling pathway, including renal cell carcinoma(Chae et al. 2020) and multiple myeloma(Chen et al. 2020a). Combined with the above results, we believe that DARS / GDI2 / P4HA2 / TRUB1 may be involved in these processes to affect the occurrence and development of GBM disease, which is also consistent with our Drug sensitivity analysis results. Among them, the DARS gene encodes the aspartyl-tRNA synthetase(Dominik et al. 2018), which pairs aspartate

with its corresponding tRNA. Missense mutations in the gene encoding DARS can lead to leukocyte dystrophy, accompanied by a marked reduction in myelin sheath, abnormal movement and cognitive impairment (Fröhlich et al. 2018). There are no related reports about the relations between DARS and GBM.According to our research, DARS may participate in TSC/mTOR signaling, by regulating GBM cell growth process. GDI2 controls the activity of Rho GTPase's pathway to regulatory guanine nucleotide exchange factor and GTPase activating protein, and may play a role in tumor cell apoptosis. This is also in line with our results. At the same time, a recent study shows that RhoGDI2 suppresses lung metastasis in mice by reducing tumor versican expression and macrophage infiltration. The expression of P4HA2 increased in head and neck squamous cell carcinoma (HNSCC)(Kisoda et al. 2020), Oral Squamous Cell Carcinoma (OSCC)(Reis et al. 2020), cervical cancer(Cao et al. 2020b) and other cancers. Especially，we found that P4HA2 are markedly upregulated in cervical cancer tissues and upregulation of P4HA2 was associated with shorter overall survival (OS) and relapse-free survival (RFS)(Cao et al. 2020b). In GBM, we found that P4HA2 is mainly involved in the process of inhibiting DNA damage, and is also related to EMT, Hormone AR, Hormone ER, RAS / MAPK, TSC / mTOR and other pathways. TRUB1 mRNA is widely expressed in various human tissues (especially heart, skeletal muscle and liver)，but there are few studies on its relationship with cancer(Zucchini et al. 2003). In our research, we analyzed that TRUB1 is mainly involved in Apoptosis, DNA damage, EMT, PI3K / AKT and other processes.

In the analysis of hub genes co-expression, we found the four genes (CDC73 / CDC123 / B4GALT1 / CUL2) are most relevant to the expression of hub genes and also related to the occurrence of many cancers. For example, CDC73 is a tumor suppressor, which can prevent cells from growing and dividing too fast or uncontrolled, and is closely related to parathyroid carcinoma(Cetani et al. 2019). CDC123 is a cell division cycle protein, and the regulatory effects of the entire cell cycle process can be stopped in one of the normal stages (G1, S, G2, M).CDC123 is highly expressed in choriocarcinoma(Hussain et al. 2018). B4GALT1 is one of the seven β-1,4-galactosyltransferase (beta4galt) genes. The β1,4-galactosylation of glycans is very important for many biological events, including the development of cancer. In a variety of cancers, the B4GALTs family is associated with cancer cell proliferation, invasion, metastasis, and drug resistance.B4GALT1 is highly expressed in patients with lung adenocarcinoma(Zhang et al. 2019). CUL2is one of the seven members of Cullin family. It can participate in the regulation of cell cycle, proliferation, apoptosis, differentiation, gene expression, transcription regulation, signal transmission, damage repair, inflammation and immunity.CUL2 affects the occurrence of renal cell carcinoma by promoting the substrate ubiquitination and degradation(Liu et al. 2020).

Further TIMER analysis indicated that the immune system had a good effect on tumor microenvironment, and that the mutations of DARS / GDI2 / P4HA2 / TRUB1 had important application value in tumor immunology. Finally, we conducted a comprehensive evaluation of gene sets using GSVA and we found that the DARS/GDI2/P4HA2/TRUB1 gene sets are closely related to the occurrence of cancer.For instance，the APICAL_ JUNCTION in the DARS gene set is more common in highly differentiated epithelial cells, such as colon cancer cells(Nair-Menon et al. 2020).MITOTIC_SPINDLE in the GDI2 gene set, the mitotic spindle inhibitor is one of the most commonly used chemotherapeutics now(Bukowski et al. 2020). DNA_REPAIR in the P4HA2 gene set and ANGIOGENESIS in the TRUB1 gene set are also two important mechanisms of cancer development .

In recent years, with the GBM genes related to the occurrence and prognosis of feature recognition in many studies. Such as Chen X found the ASPMexpression pattern from the database showed that it is highly expressed in GBM tissue, and patients with high expression of ASPM have a poor prognosis(Chen et al. 2020b). Recently, a bioinformatic analysis of 123 GBM patients has established a 14-mRNA prognostic signature, which could be used to classify GBM patients into low and high risk groups(Arimappamagan et al. 2013).To our knowledge, the DARS/GDI2/P4HA2/TRUB1 that we identified are new GBM biomarkers because they have never been reported to be associated with the development and progression of GBM(Lu et al. 2020). At the same time, compared with the traditional typing methods, the multi-gene markers model has many advantages, such as high prediction accuracy and personalized detection results(Albuquerque et al. 2012). Therefore, multi-gene markers have a good application prospect in clinical practice. In our study, we built and verified the characteristic of the four genes through analyzing the two independent data sets. More reasonable use of biometrics and multiple independent data sets of mutual verification makes our results more reliable.

However, our study had some limitations. Associated with disease, for example, age, race, sex, and some unknown prognostic factors may not be included in the model, which limits the prediction ability of the model. In the future, we plan to establish a more reasonable model of biological information analysis. Meanwhile, it should be acknowledged that the single gene analysis in this study does have limitations, and in future studies we will combine all the hub genes or other factors together to find a biomarker with better sensitivity and accuracy using a multi-omics approach. In summary, our results had shown that DARS/GDI2/P4HA2/TRUB1 can be used as a new biological marker for GBM, which is related to the occurrence and prognosis of GBM, how to rationally apply various genetic characteristics at specific stages of GBM for diagnose and prediction of prognosis.

Conclusion

The molecular biological characteristics of GBM has changed the classification and treatment of tumors and become an important part of diagnosis and oncologic therapy. This study used public databases to identify four valuable biomarkers for GBM prognosis, namely DARS / GDI2 / P4HA2 / TRUB1, which have potential and clinical application values to act as prognostic markers of GBM.

Acknowledgement

None

Reference

Albuquerque AD, Kubisch I, Breier G, Stamminger G, Fersis N, Eichler A, Kaul S, and St?Lzel U. 2012. Multimarker gene analysis of circulating tumor cells in pancreatic cancer patients: a feasibility study. *Oncology* 82:3-10.

Arimappamagan A, Somasundaram K, Thennarasu K, Peddagangannagari S, Srinivasan H, Shailaja BC, Samuel C, Patric IR, Shukla S, Thota B, Prasanna KV, Pandey P, Balasubramaniam A, Santosh V, Chandramouli BA, Hegde AS, Kondaiah P, and Sathyanarayana Rao MR. 2013. A

fourteen gene GBM prognostic signature identifies association of immune response pathway and mesenchymal subtype with high risk group. *PLoS ONE* 8:e62042. 10.1371/journal.pone.0062042

Bukowski K, Kciuk M, and Kontek R. 2020. Mechanisms of Multidrug Resistance in Cancer Chemotherapy. *Int J Mol Sci* 21. 10.3390/ijms21093233

Cao Y, Han Q, Li J, Jia Y, and Shi H. 2020a. P4HA2 contributes to cervical cancer progression via inducing epithelial-mesenchymal transition. *Journal of Cancer* 11:2788-2799.

Cao Y, Han Q, Li J, Jia Y, Zhang R, and Shi H. 2020b. P4HA2 contributes to cervical cancer progression via inducing epithelial-mesenchymal transition. *J Cancer* 11:2788-2799. 10.7150/jca.38401

Cetani F, Marcocci C, Torregrossa L, and Pardi E. 2019. Atypical parathyroid adenomas: challenging lesions in the differential diagnosis of endocrine tumors. *Endocr Relat Cancer* 26:R441-R464. 10.1530/ERC-19-0135

Chae IG, Song NY, Kim DH, Lee MY, Park JM, and Chun KS. 2020. Thymoquinone induces apoptosis of human renal carcinoma Caki-1 cells by inhibiting JAK2/STAT3 through pro-oxidant effect. *Food Chem Toxicol* 139:111253. 10.1016/j.fct.2020.111253

Chen G, Hu K, Sun H, Zhou J, Song D, Xu Z, Gao L, Lu Y, Cheng Y, Feng Q, Zhang H, Wang Y, Hu L, Lu K, Wu X, Li B, Zhu W, and Shi J. 2020a. A novel phosphoramide compound, DCZ0847, displays in vitro and in vivo anti-myeloma activity, alone or in combination with bortezomib. *Cancer Lett* 478:45-55. 10.1016/j.canlet.2020.03.006

Chen L, Peng T, Luo Y, Zhou F, Wang G, Qian K, Xiao Y, and Wang X. 2019. ACAT1 and Metabolism-Related Pathways Are Essential for the Progression of Clear Cell Renal Cell Carcinoma (ccRCC), as Determined by Co-expression Network Analysis. *Front Oncol* 9:957. 10.3389/fonc.2019.00957

Chen X, Huang L, Yang Y, Chen S, Sun J, Ma C, Xie J, Song Y, and Yang J. 2020b. ASPM promotes glioblastoma growth by regulating G1 restriction point progression and Wnt-beta-catenin signaling. *Aging (Albany NY)* 12:224-241. 10.18632/aging.102612

Della-Flora A, Wilde ML, Pinto IDF, Lima EC, and Sirtori C. 2020. Degradation of the anticancer drug flutamide by solar photo-Fenton treatment at near-neutral pH: Identification of transformation products and in silico (Q)SAR risk assessment. *Environ Res* 183:109223. 10.1016/j.envres.2020.109223

Dominik F, Suchowerska AK, Carola V, He R, Ernst W, Georg VJ, Cas S, Thomas F, Housley GD, and Matthias K. 2018. Expression Pattern of the Aspartyl-tRNA Synthetase DARS in the Human Brain. *Frontiers in Molecular Neuroscience* 11:81-.

Dou B, Jiang Z, Chen X, Wang C, and Sheng G. 2020. Oncogenic Long Noncoding RNA DARS-AS1 in Childhood Acute Myeloid Leukemia by Binding to microRNA-425. *Technology in Cancer Research & Treatment* 19:153303382096558.

Fröhlich D, Suchowerska AK, Voss C, He R, Wolvetang E, von Jonquieres G, Simons C, Fath T, Housley GD, and Klugmann M. 2018. Expression Pattern of the Aspartyl-tRNA Synthetase DARS in the Human Brain. *Frontiers in Molecular Neuroscience* 11. 10.3389/fnmol.2018.00081

Gong Z, Hong F, Wang H, Zhang X, and Chen J. 2020. An eight-mRNA signature outperforms the lncRNA-based signature in predicting prognosis of patients with glioblastoma. *BMC Med Genet* 21:56. 10.1186/s12881-020-0992-7

Hanzelmann S, Castelo R, and Guinney J. 2013. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* 14:7. 10.1186/1471-2105-14-7

Harter D, Wilson T, and Karajannis M. 2014. Glioblastoma multiforme: State of the art and future therapeutics. *Surgical Neurology International* 5. 10.4103/2152-7806.132138

Hussain S, Saxena S, Shrivastava S, Mohanty AK, Kumar S, Singh RJ, Kumar A, Wani SA, Gandham RK, Kumar N, Sharma AK, Tiwari AK, and Singh RK. 2018. Gene expression profiling of spontaneously occurring canine mammary tumours: Insight into gene networks and pathways linked to cancer pathogenesis. *PLoS ONE* 13:e0208656. 10.1371/journal.pone.0208656

Irshad K, Mohapatra SK, Srivastava C, Garg H, Mishra S, Dikshit B, Sarkar C, Gupta D, Chandra PS, Chattopadhyay P, Sinha S, and Chosdol K. 2015. A combined gene signature of hypoxia and notch pathway in human glioblastoma and its prognostic relevance. *PLoS ONE* 10:e0118201. 10.1371/journal.pone.0118201

Kisoda S, Shao W, Fujiwara N, Mouri Y, Tsunematsu T, Jin S, Arakaki R, Ishimaru N, and Kudo Y. 2020. Prognostic value of partial EMT-related genes in head and neck squamous cell carcinoma by a bioinformatic analysis. *Oral Dis*. 10.1111/odi.13351

Lara-Velazquez M, Zarco N, Carrano A, Phillipps J, and Guerrero-Cazares H. 2020. 543: Cerebrospinal Fluid-Responsive Factor SERPINA3 Promotes Proliferation, Migration and Invasion of Glioblastoma. 543: Cerebrospinal Fluid-Responsive Factor SERPINA3 Promotes Proliferation, Migration and Invasion of Glioblastoma.

Li C, Tan J, Chang J, Li W, Liu Z, Li N, and Ji Y. 2017a. Radioiodine-labeled anti-epidermal growth factor receptor binding bovine serum albumin-polycaprolactone for targeting imaging of glioblastoma. *Oncol Rep* 38:2919-2926. 10.3892/or.2017.5937

Li Q, Wang Q, Zhang Q, Zhang J, and Zhang J. 2019a. Collagen prolyl 4-hydroxylase 2 predicts worse prognosis and promotes glycolysis in cervical cancer. *American Journal of Translational Research* 11:6938-6951.

Li SS, Lian YF, Huang YL, Huang YH, and Xiao J. 2020. Overexpressing PLOD family genes predict poor prognosis in gastric cancer. *J Cancer* 11:121-131. 10.7150/jca.35763

Li T, Fan J, Wang B, Traugh N, Chen Q, Liu JS, Li B, and Liu XS. 2017b. TIMER: A Web Server for Comprehensive Analysis of Tumor-Infiltrating Immune Cells. *Cancer Research* 77:e108-e110. 10.1158/0008-5472.Can-17-0307

Li W, Lu J, Ma Z, Zhao J, and Liu J. 2019b. An Integrated Model Based on a Six-Gene Signature Predicts Overall Survival in Patients With Hepatocellular Carcinoma. *Front Genet* 10:1323. 10.3389/fgene.2019.01323

Liu A, Zhang S, Shen Y, Lei R, and Wang Y. 2019. Association of mRNA expression levels of Cullin family members with prognosis in breast cancer. *Medicine* 98. 10.1097/md.0000000000016625

Liu CJ, Hu FF, Xia MX, Han L, Zhang Q, and Guo AY. 2018. GSCALite: a web server for gene set cancer analysis. *Bioinformatics* 34:3771-3772. 10.1093/bioinformatics/bty411

Liu X, Zurlo G, and Zhang Q. 2020. The Roles of Cullin-2 E3 Ubiquitin Ligase Complex in Cancer. *Adv Exp Med Biol* 1217:173-186. 10.1007/978-981-15-1025-0_11

Lu WC, Xie H, Yuan C, Li JJ, and Wu AH. 2020. Identification of potential biomarkers and candidate small molecule drugs in glioblastoma. *Cancer Cell International* 20.

Malik V, Garg S, Afzal S, Dhanjal JK, and Wadhwa R. 2020. Bioinformatics and Molecular Insights to Anti-Metastasis Activity of Triethylene Glycol Derivatives. *International Journal of Molecular Sciences* 21.

Nair-Menon J, Daulagala AC, Connor DM, Rutledge L, Penix T, Bridges MC, Wellslager B, Spyropoulos DD, Timmers CD, Broome AM, and Kourtidis A. 2020. Predominant Distribution of the RNAi Machinery at Apical Adherens Junctions in Colonic Epithelia Is Disrupted in Cancer. *Int J Mol Sci* 21. 10.3390/ijms21072559

Nelson, and N. 2014. Metal ion transporters and homeostasis. *Embo Journal* 18:4361-4371.

Pi H, Wendel BM, and Helmann JD. 2020. Dysregulation of Magnesium Transport Protects Bacillus subtilis against Manganese and Cobalt Intoxication. *J Bacteriol* 202. 10.1128/JB.00711-19

Reichel D, Sagong B, Teh J, Zhang Y, and Perez JM. 2020. Near Infrared Fluorescent Nanoplatform for Targeted Intraoperative Resection and Chemotherapeutic Treatment of Glioblastoma. *ACS Nano* XXXX.

Reis PP, Tokar T, Goswami RS, Xuan Y, Sukhai M, Seneda AL, Moz LES, Perez-Ordonez B, Simpson C, Goldstein D, Brown D, Gilbert R, Gullane P, Irish J, Jurisica I, and Kamel-Reid S. 2020. A 4-gene signature from histologically normal surgical margins predicts local recurrence in patients with oral carcinoma: clinical validation. *Sci Rep* 10:1713. 10.1038/s41598-020-58688-y

Sarquis M, Marx SJ, Beckers A, Bradwell AR, Simonds WF, Bicalho MAC, Daly AF, Betea D, Friedman E, and De Marco L. 2019. Long-term remission of disseminated parathyroid cancer following immunotherapy. *Endocrine* 67:204-208. 10.1007/s12020-019-02136-z

Schafer ST, Paquola ACM, Stern S, Gosselin D, Ku M, Pena M, Kuret TJM, Liyanage M, Mansour AA, Jaeger BN, Marchetto MC, Glass CK, Mertens J, and Gage FH. 2019. Pathological priming causes developmental gene network heterochronicity in autistic subject-derived neurons. *Nature Neuroscience* 22:243-255. 10.1038/s41593-018-0295-x

Shi S, Ye S, Mao J, Ru Y, Lu Y, Wu X, Xu M, Zhu T, Wang Y, Chen Y, Tang X, and Xi Y. 2020. CMA1 is potent prognostic marker and associates with immune infiltration in gastric cancer. *Autoimmunity*:1-8. 10.1080/08916934.2020.1735371

Sun W, Kuang XL, Liu YP, Tian LF, Yan XX, and Xu W. Crystal structure of the N-terminal domain of human CDC73 and its implications for the hyperparathyroidism-jaw tumor (HPT-JT) syndrome. *Scientific Reports*.

Szopa W, Burley TA, Kramer-Marek G, and Kaspera W. 2017. Diagnostic and Therapeutic Biomarkers in Glioblastoma: Current Status and Future Perspectives. *Biomed Res Int*

2017:8013575. 10.1155/2017/8013575

Tan AC, Ashley DM, López G, Malinzak M, and Khasraw M. 2020. Management of glioblastoma: State of the art and future directions. *CA A Cancer Journal for Clinicians* 70.

Tij A, Pr A, Acbc D, Is A, e FMMbcd, Rk F, Tka G, Dszbc H, Mkbc D, and Rm I. 2021. Frontiers in the treatment of glioblastoma: Past, present and emerging - ScienceDirect. *Advanced Drug Delivery Reviews*.

Tingting, Long, Zijing, Liu, Xing, Zhou, Shuang, Yu, Hui, and Tian. 2019. Identification of differentially expressed genes and enriched pathways in lung cancer using bioinformatics analysis. *Molecular Medicine Reports*.

Tonry C. 2020. Clinical proteomics for prostate cancer: understanding prostate cancer pathology and protein biomarkers for improved disease management. *Clinical Proteomics* 17.

Tripathi S, Pohl Marie O, Zhou Y, Rodriguez-Frandsen A, Wang G, Stein David A, Moulton Hong M, DeJesus P, Che J, Mulder Lubbertus CF, Yángüez E, Andenmatten D, Pache L, Manicassamy B, Albrecht Randy A, Gonzalez Maria G, Nguyen Q, Brass A, Elledge S, White M, Shapira S, Hacohen N, Karlas A, Meyer Thomas F, Shales M, Gatorano A, Johnson Jeffrey R, Jang G, Johnson T, Verschueren E, Sanders D, Krogan N, Shaw M, König R, Stertz S, García-Sastre A, and Chanda Sumit K. 2015. Meta- and Orthogonal Integration of Influenza "OMICs" Data Defines a Role for UBR4 in Virus Budding. *Cell Host & Microbe* 18:723-735. 10.1016/j.chom.2015.11.002

Wang W, Xing H, Huang C, Pan H, and Li D. 2020. Identification of pancreatic cancer type related factors by Weighted Gene Co-Expression Network Analysis. *Med Oncol* 37:33. 10.1007/s12032-020-1339-0

Wu J, Su HK, Yu ZH, Xi SY, Guo CC, Hu ZY, Qu Y, Cai HP, Zhao YY, Zhao HF, Chen FR, Huang YF, To ST, Feng BH, Sai K, Chen ZP, and Wang J. 2020. Skp2 modulates proliferation, senescence and tumorigenesis of glioma. *Cancer Cell Int* 20:71. 10.1186/s12935-020-1144-z

Xing Z-y, Sun L-g, and Guo W-j. 2015. Elevated expression of Notch-1 and EGFR induced apoptosis in glioblastoma multiforme patients. *Clinical Neurology and Neurosurgery* 131:54-58. 10.1016/j.clineuro.2015.01.018

Zhang B, and Horvath S. 2005. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol* 4:Article17. 10.2202/1544-6115.1128

Zhang L, Zhang Z, and Yu Z. 2019. Identification of a novel glycolysis-related gene signature for predicting metastasis and survival in patients with lung adenocarcinoma. *J Transl Med* 17:423. 10.1186/s12967-019-02173-2

Zheng W, Zou Z, Lin S, Chen X, Wang F, Li X, and Dai J. 2018. Identification and functional analysis of spermatogenesis-associated gene modules in azoospermia by weighted gene coexpression network analysis. *Journal of Cellular Biochemistry* 120:3934-3944. 10.1002/jcb.27677

Zhou Z, Cheng Y, Jiang Y, Liu S, Zhang M, Liu J, and Zhao Q. 2018. Ten hub genes associated with progression and prognosis of pancreatic carcinoma identified by co-expression analysis.

International Journal of Biological Sciences 14:124-136. 10.7150/ijbs.22619

Zucchini C, Strippoli P, Biolchi A, Solmi R, Lenzi L, D'Addabbo P, Carinci P, and Valvassori L. 2003. The human TruB family of pseudouridine synthase genes, including the Dyskeratosis Congenita 1 gene and the novel member TRUB1. *Int J Mol Med* 11:697-704.

Zurlo G, and Zhang Q. 2020. Adenylosuccinate lyase hydroxylation contributes to triple negative breast cancer via the activation of cMYC. *Mol Cell Oncol* 7:1707045. 10.1080/23723556.2019.1707045

Figure Caption

Fig. 1. Identification of gene co-expression modules. (A) cluster samples to detect outliers. All samples are located in the cluster and pass the critical threshold at the same time. The green highlighting means that the samples are in strong trait relationships established by correlation analyses. (B) The scale-free fit index was analyzed under the background of different soft-thresholding power (β).(C) Analyze average connectivity when using different soft-thresholding powers.(D) Dendrogram clustering of all genomic genes in GBM samples.(E) Heatmap of co-expressed genes.Different modules on the X and Y axis have different colors.The connection degree of different modules is indicated by the yellow intensity.

Fig. 2. Correlation of modules and clinical traits. (A) Module intrinsic genes and relapse Heatmap of the correlation between. (B) Scatter plot of the correlation between the darkturquoise module and relapse. All modules can be correlated with genes, and all continuous traits can be correlated with gene expression values. The two correlation matrices are combined and the vertical coordinate is the Gene significance for luminal when the module of interest is specified for analysis. (C) Enrichment analysis of the function and pathway of the darkturquoise module.The rich biological process terms in the selected modules are described as interactive networks and listed according to their P-value. The size of the dots represents the number of genes that are co-expressed, the larger the dot, the more genes are co-expressed, presumably the more important they are and the more important they are as core genes in the network graph. Each node is a gene. The size of the node means degree of gene enrichment. Set $P < 0.01$ as the cutoff criterion.Enrich the term network, colored with cluster-ID, where nodes sharing the same cluster ID are usually close to each other.

Fig. 3. The Kaplan-Meier survival curve can evaluate the prognostic performance of core genes based on the expression status of selected biomarkers in the database. (A) DARS. (B) GDI2. (C) P4HA2. (D)TRUB1. All patients in each group were divided into high expression group and low expression group by gene expression. The cutoff for low versus high expression is 3-fold expression of controls.

Fig. 4. Gene co-expression. (A-H) In the TCGA dataset, selected the mRNA expression levels of DARS / GDI2 / P4HA2 / TRUB1 related genes, analyzed the correlation of these genes through R, and visualize them with the circus and heatmap graph. (I-L) The four genes with the highest correlation with DARS / GDI2 / P4HA2 / TRUB1, drew scatter plots.

Fig. 5. Drug sensitivity analysis. (A) The role of DARS/GDI2/P4HA2/TRUB1 in the famous cancer related pathways (GSCALite). The size of an area in the pie chart represents the extent of the role of DARS/GDI2/P4HA2/TRUB1 in the well-known cancer-related pathway (GSCALite). (B) In the GDSC training set,high expression of DARS/GDI2/P4HA2/TRUB1 was inferred to be less sensitive to commonly used chemotherapy drugs.The test for association between paired samples used Pearson's correlation coefficient. Two-tailed statistical P values were calculated by a two-sample Mann-Whitney test or Student's t test when appropriate.

Fig. 6. Genetic and transcriptional changes and connections with immune cell populations. (A) Correlation of DARS/GDI2/P4HA2/TRUB1 expression with immune infiltration level in GBM. (B) DARS/GDI2/P4HA2/TRUB1 copy number alterations (CNV) affects the level of

infiltration of B cells, CD8+ T cells, CD4+T celles, Macrophages, Neutrophils, and Dendritic cells in GBM.

Fig. 7. Genomic alterations of DARS/GDI2/P4HA2/TRUB1 in GBM. OncoPrint of DARS/GDI2/P4HA2/TRUB1 alterations in GBM cohort. The different types of genetic alterations are highlighted in different colors. Expression profiles of mRNAs showing different expression (≥3-fold) compared to control were considered to be mRNA high, and vice versa for low.

Fig. 8. GSVA analysis. GSVA of DARS/GDI2/P4HA2/TRUB1 gene sets in GBM. (A) DARS. (B) GDI2. (C) P4HA2. (D) TRUB1.A t value > 1 or < − 1 represents statistically significant changes.

Figure 1

Identification of gene co-expression modules.

(A) cluster samples to detect outliers. All samples are located in the cluster and pass the critical threshold at the same time. The green highlighting means that the samples are in strong trait relationships established by correlation analyses. (B) The scale-free fit index was analyzed under the background of different soft-thresholding power (β).(C) Analyze average connectivity when using different soft-thresholding powers.(D) Dendrogram clustering of all genomic genes in GBM samples.(E) Heatmap of co-expressed genes.Different modules on the X and Y axis have different colors.The connection degree of different modules is indicated by the yellow intensity.

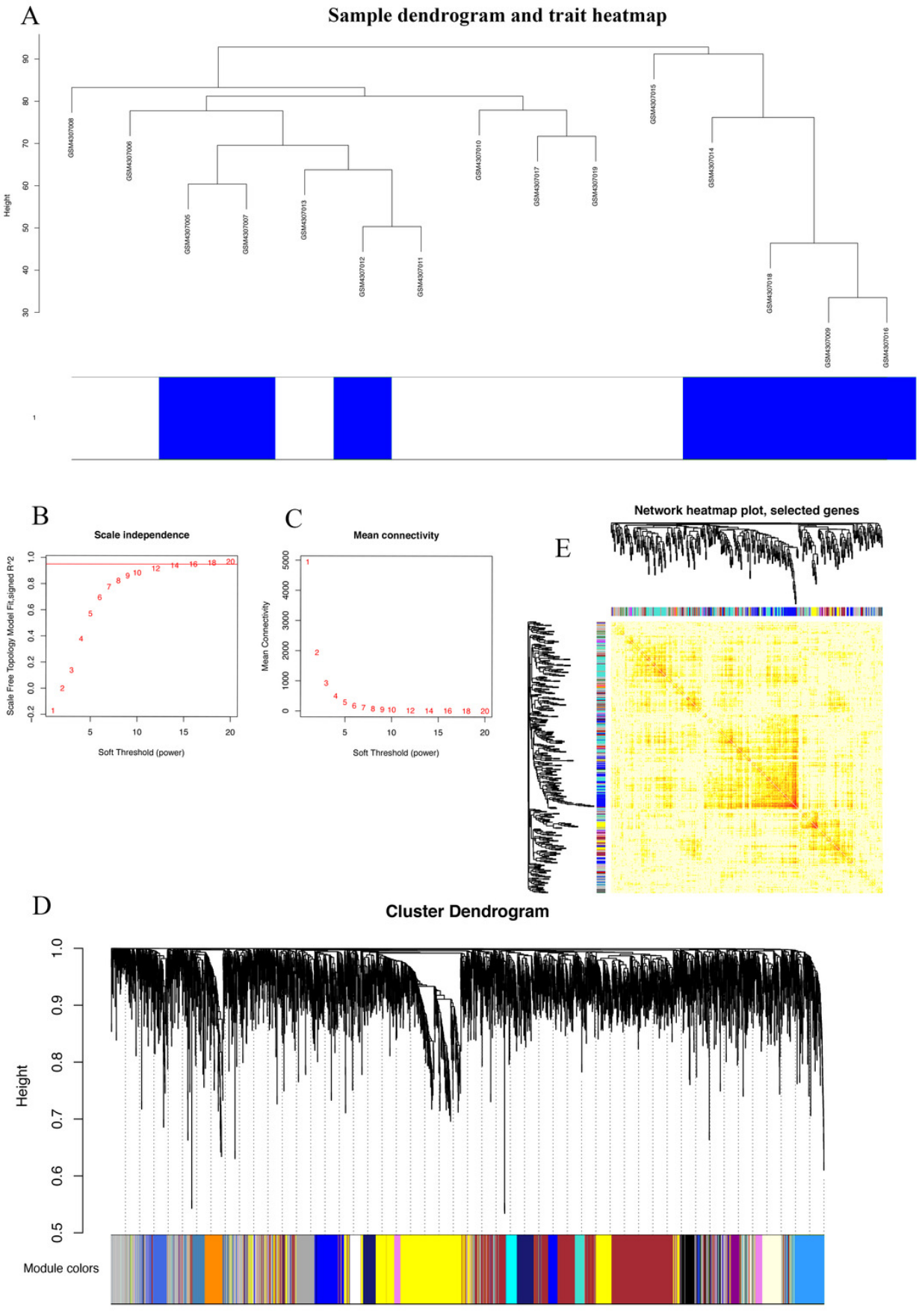
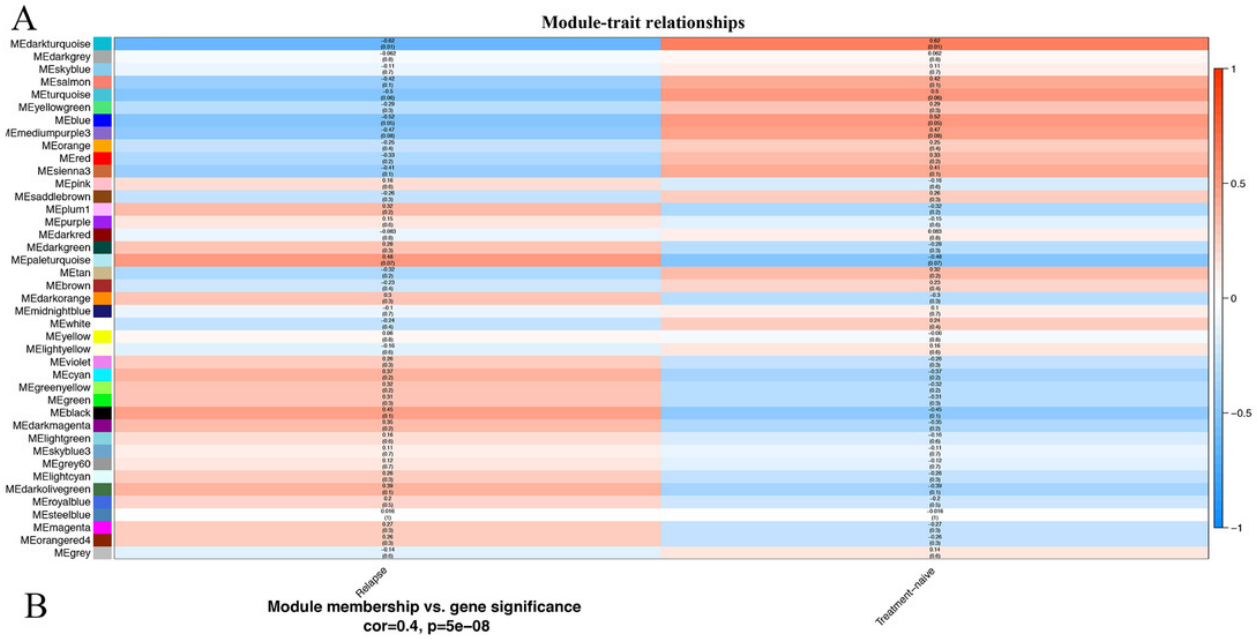
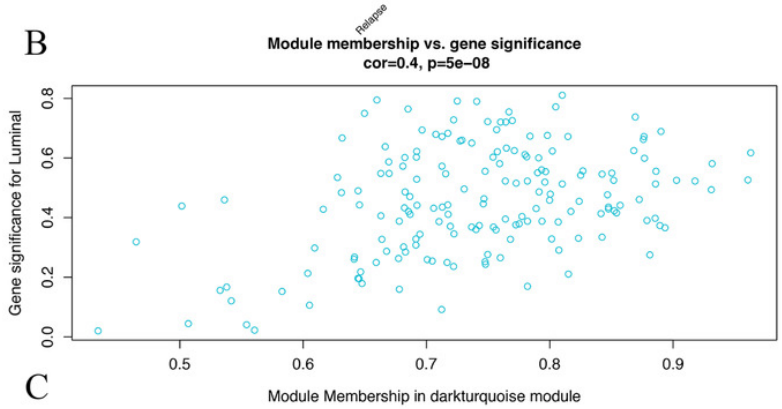

A
Sample dendrogram and trait heatmap
Height
GSM4307008
GSM4307006
GSM4307005
GSM4307007
GSM4307013
GSM4307012
GSM4307011
GSM4307010
GSM4307017
GSM4307019
GSM4307015
GSM4307014
GSM4307018
GSM4307009
GSM4307016
B
Scale independence
Scale Free Topology Model Fit,signed R^2
Soft Threshold (power)
C
Mean connectivity
Mean Connectivity
Soft Threshold (power)
E
Network heatmap plot, selected genes
D
Cluster Dendrogram
Height
Module colors

Figure 2

Correlation of modules and clinical traits.

(A) Module intrinsic genes and relapse Heatmap of the correlation between. (B) Scatter plot of the correlation between the darkturquoise module and relapse. All modules can be correlated with genes, and all continuous traits can be correlated with gene expression values. The two correlation matrices are combined and the vertical coordinate is the Gene significance for luminal when the module of interest is specified for analysis. (C) Enrichment analysis of the function and pathway of the darkturquoise module.The rich biological process terms in the selected modules are described as interactive networks and listed according to their P-value. The size of the dots represents the number of genes that are co-expressed, the larger the dot, the more genes are co-expressed, presumably the more important they are and the more important they are as core genes in the network graph. Each node is a gene. The size of the node means degree of gene enrichment. Set $P < 0.01$ as the cutoff criterion.Enrich the term network, colored with cluster-ID, where nodes sharing the same cluster ID are usually close to each other.


A
Module-trait relationships
MEdarkturquoise
MEdarkgrey
MEskyblue
MEsalmon
MEturquoise
MEyellowgreen
MEblue
MEmediumpurple3
MEorange
MEred
MEsienna3
MEpink
MEsaddlebrown
MEplum1
MEpurple
MEdarkred
MEdarkgreen
MEpaleturquoise
MEtan
MEbrown
MEdarkorange
MEmidnightblue
MEwhite
MEyellow
MElightyellow
MEviolet
MEcyan
MEgreenyellow
MEgreen
MEblack
MEdarkmagenta
MElightgreen
MEskyblue3
MEgrey60
MElightcyan
MEdarkolivegreen
MEroyalblue
MEsteelblue
MEmagenta
MEorangered4
MEgrey
Relapse
Treatment-naive
1
0.5
0
-0.5
-1


B
Module membership vs. gene significance
cor=0.4, p=5e−08
Gene significance for Luminal
0.0
0.2
0.4
0.6
0.8
0.5
0.6
0.7
0.8
0.9
Module Membership in darkturquoise module

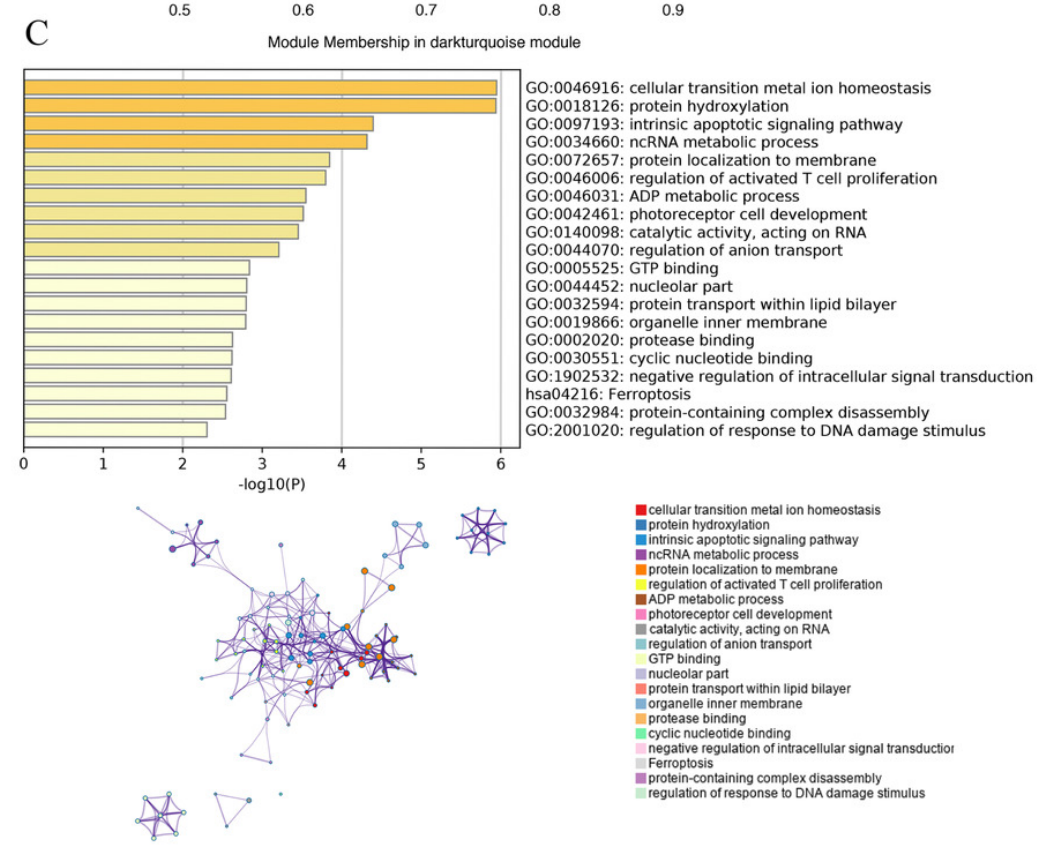
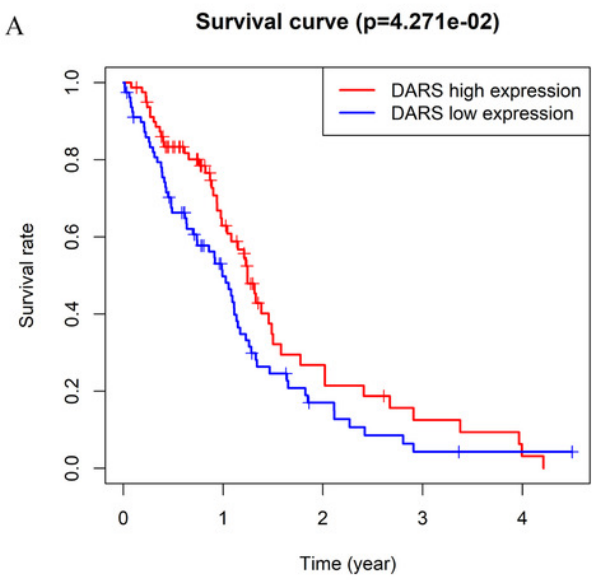
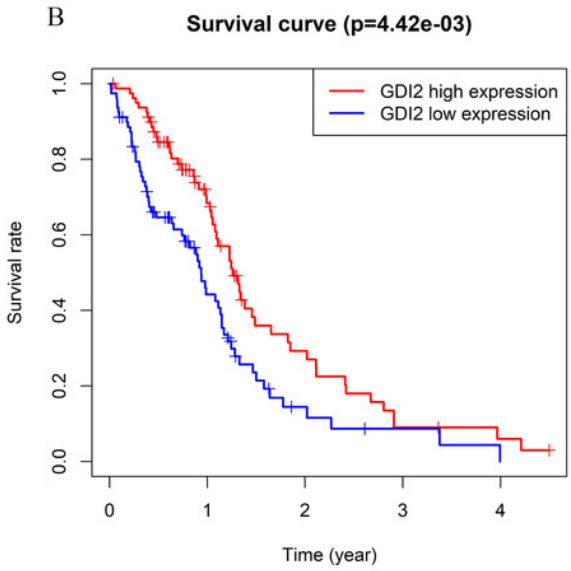

C
GO:0046916: cellular transition metal ion homeostasis
GO:0018126: protein hydroxylation
GO:0097193: intrinsic apoptotic signaling pathway
GO:0034660: ncRNA metabolic process
GO:0072657: protein localization to membrane
GO:0046006: regulation of activated T cell proliferation
GO:0046031: ADP metabolic process
GO:0042461: photoreceptor cell development
GO:0140098: catalytic activity, acting on RNA
GO:0044070: regulation of anion transport
GO:0005525: GTP binding
GO:0044452: nucleolar part
GO:0032594: protein transport within lipid bilayer
GO:0019866: organelle inner membrane
GO:0002020: protease binding
GO:0030551: cyclic nucleotide binding
GO:1902532: negative regulation of intracellular signal transduction
hsa04216: Ferroptosis
GO:0032984: protein-containing complex disassembly
GO:2001020: regulation of response to DNA damage stimulus
0
1
2
3
4
5
6
-log10(P)
cellular transition metal ion homeostasis
protein hydroxylation
intrinsic apoptotic signaling pathway
ncRNA metabolic process
protein localization to membrane
regulation of activated T cell proliferation
ADP metabolic process
photoreceptor cell development
catalytic activity, acting on RNA
regulation of anion transport
GTP binding
nucleolar part
protein transport within lipid bilayer
organelle inner membrane
protease binding
cyclic nucleotide binding
negative regulation of intracellular signal transductior
Ferroptosis
protein-containing complex disassembly
regulation of response to DNA damage stimulus

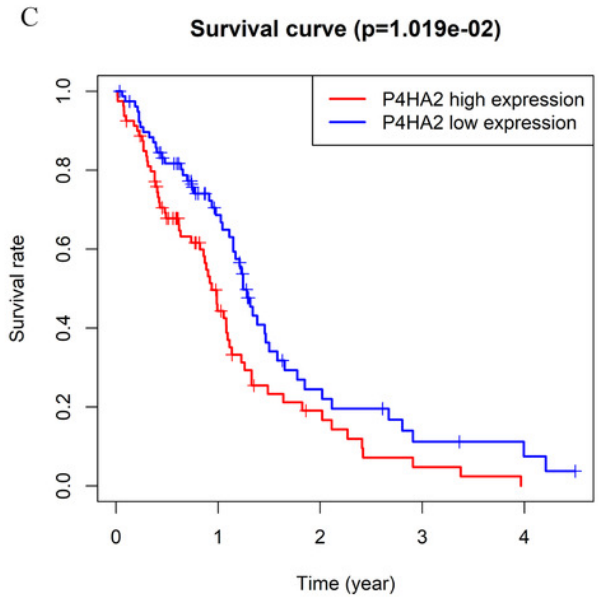
Figure 3

The Kaplan-Meier survival curve can evaluate the prognostic performance of core genes based on the expression status of selected biomarkers in the database.

(A) DARS. (B) GDI2. (C) P4HA2. (D)TRUB1. All patients in each group were divided into high expression group and low expression group by gene expression. The cutoff for low versus high expression is 3-fold expression of controls.


A
Survival curve (p=4.271e-02)
DARS high expression
DARS low expression
Survival rate
Time (year)


B
Survival curve (p=4.42e-03)
GDI2 high expression
GDI2 low expression
Survival rate
Time (year)


C
Survival curve (p=1.019e-02)
P4HA2 high expression
P4HA2 low expression
Survival rate
Time (year)

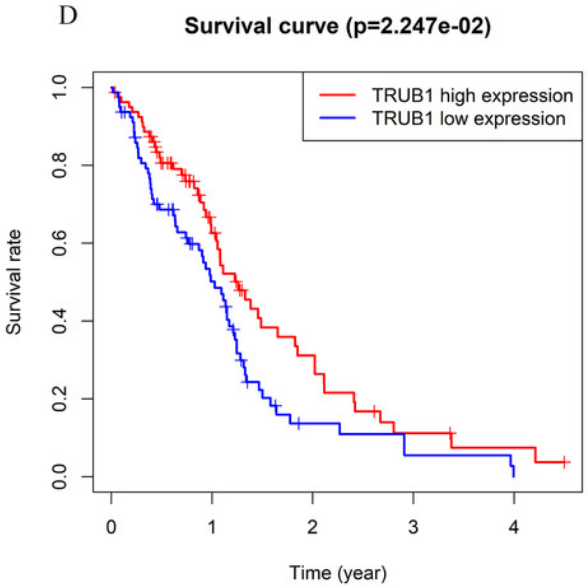

D
Survival curve (p=2.247e-02)
TRUB1 high expression
TRUB1 low expression
Survival rate
Time (year)

Figure 4

Gene co-expression.

(A-H) In the TCGA dataset, selected the mRNA expression levels of DARS / GDI2 / P4HA2 / TRUB1 related genes, analyzed the correlation of these genes through R, and visualize them with the circus and heatmap graph. (I-L) The four genes with the highest correlation with DARS / GDI2 / P4HA2 / TRUB1, drew scatter plots.

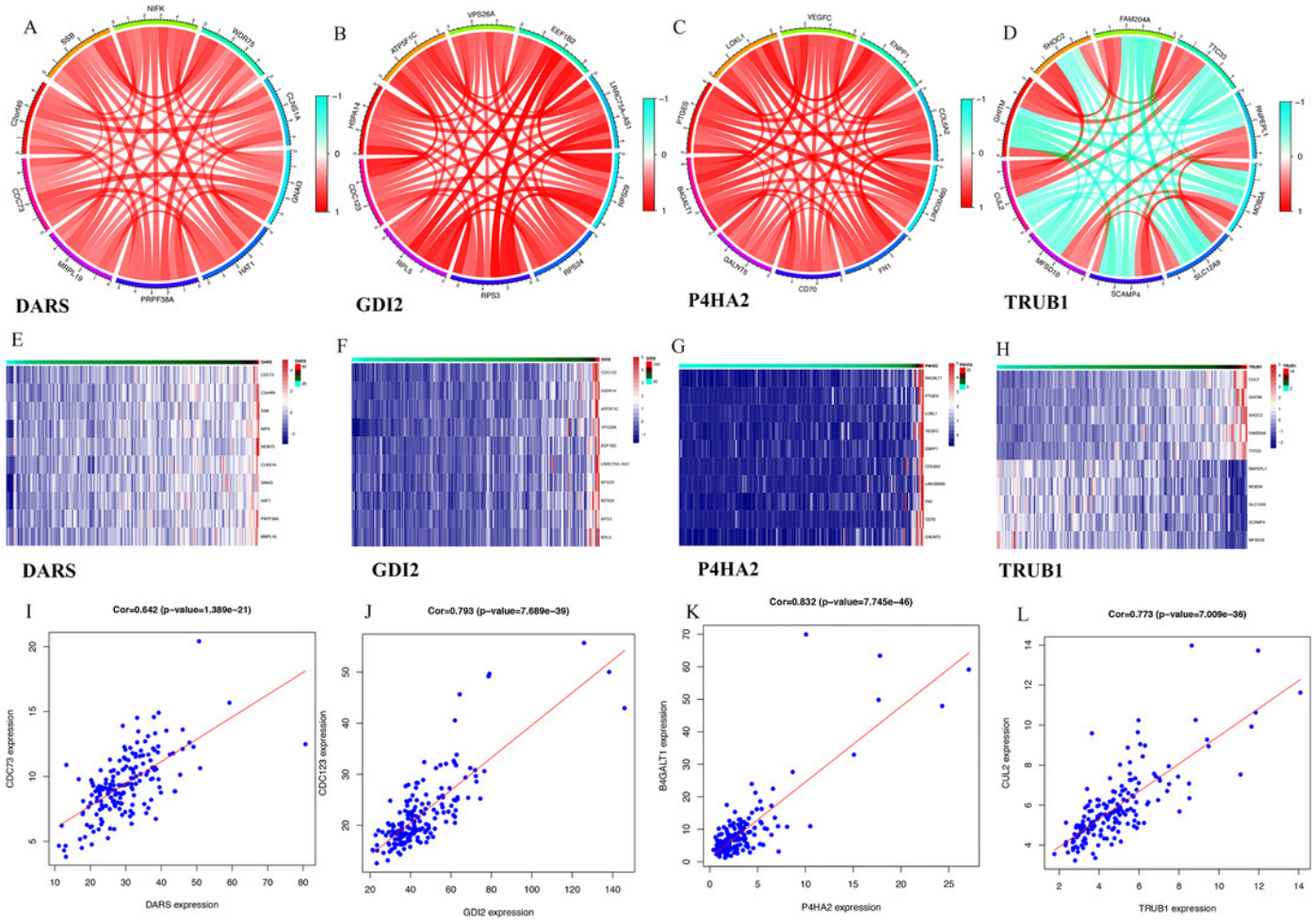


Figure 5

Drug sensitivity analysis.

(A) The role of DARS/GDI2/P4HA2/TRUB1 in the famous cancer related pathways (GSCALite). The size of an area in the pie chart represents the extent of the role of DARS/GDI2/P4HA2/TRUB1 in the well-known cancer-related pathway (GSCALite). (B) In the GDSC training set,high expression of DARS/GDI2/P4HA2/TRUB1 was inferred to be less sensitive to commonly used chemotherapy drugs.The test for association between paired samples used Pearson’s correlation coefficient. Two-tailed statistical P values were calculated by a two-sample Mann-Whitney test or Student’s t test when appropriate.

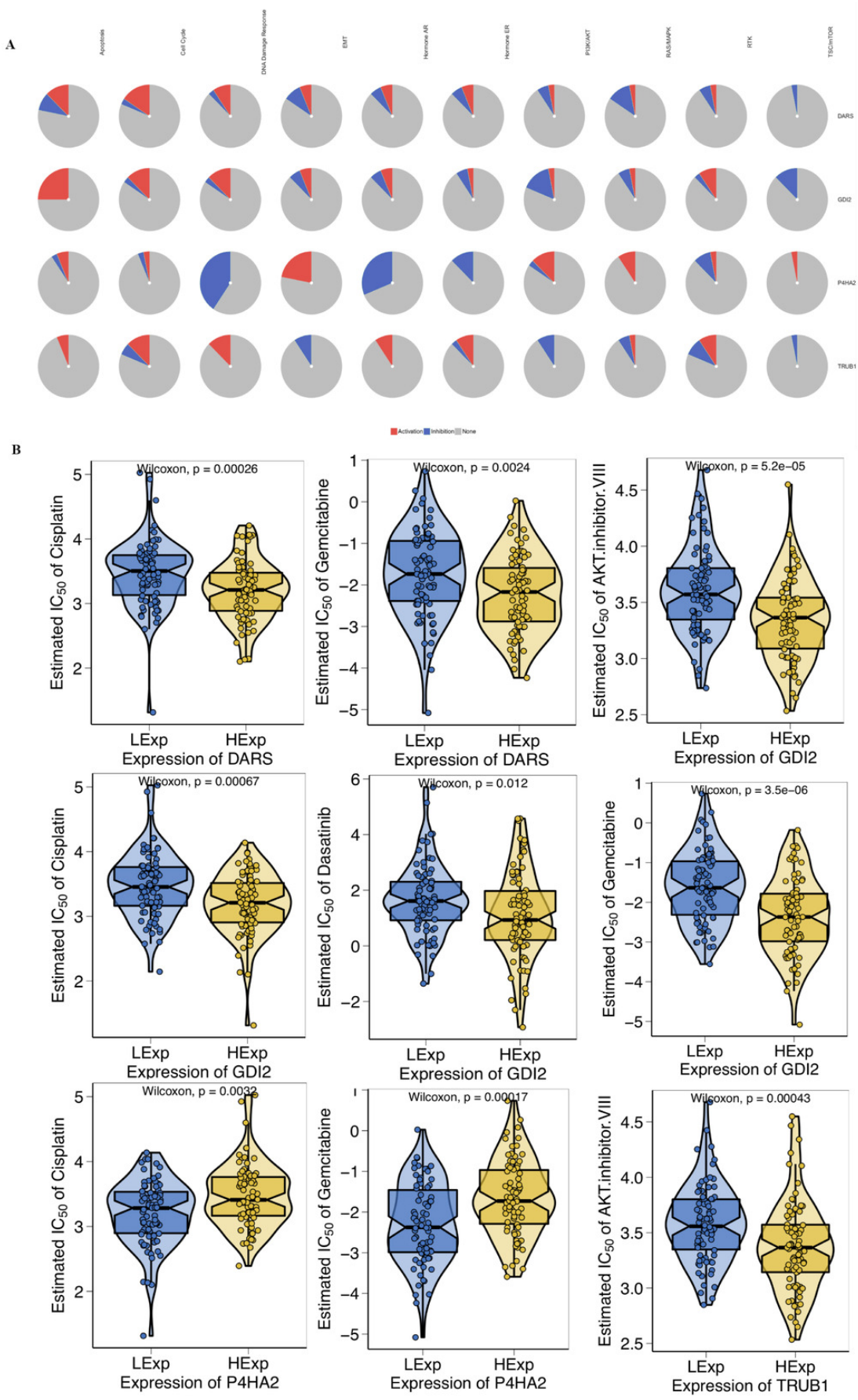
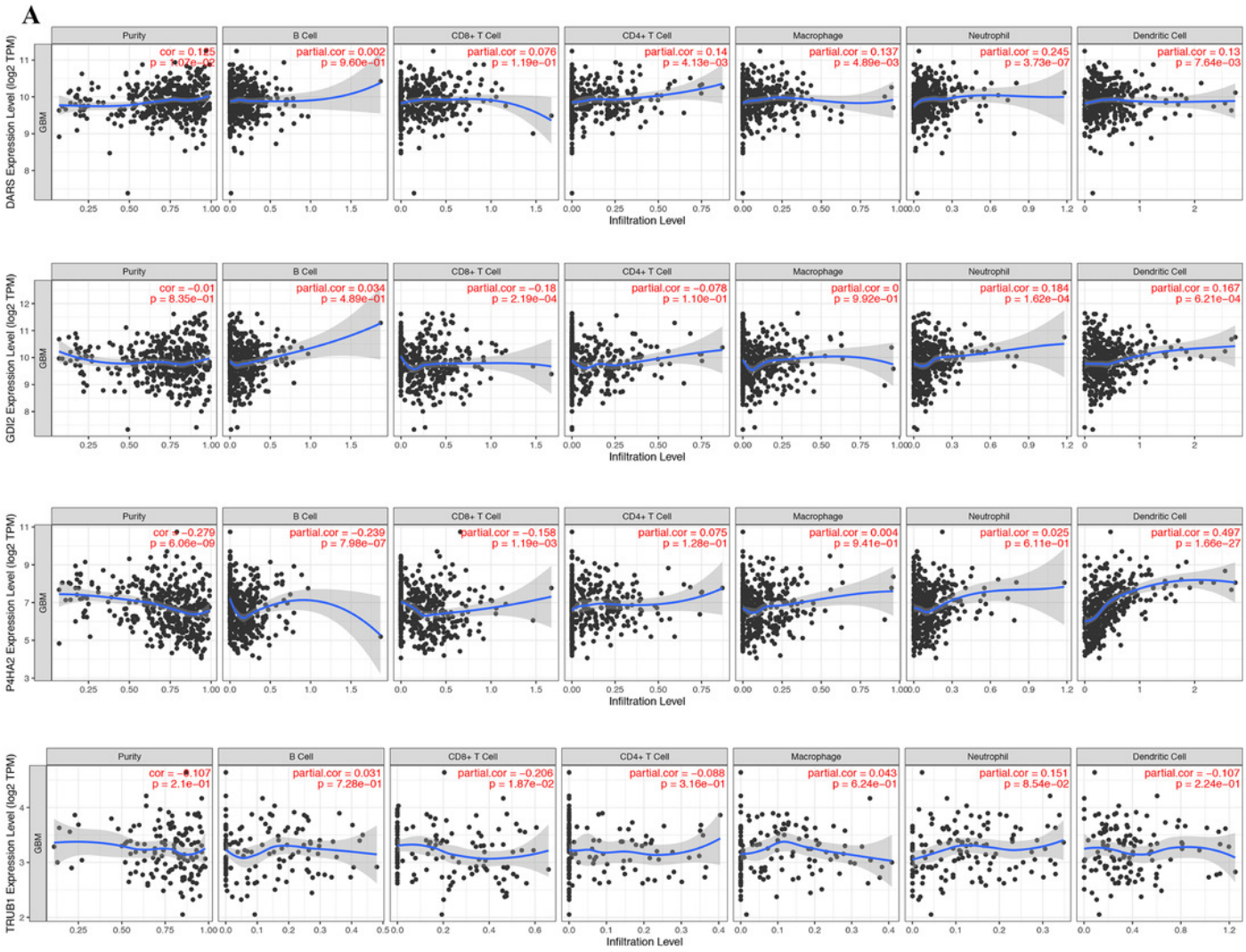

A
Apoptosis
Cell Cycle
DNA Damage Response
EMT
Hormone AR
Hormone ER
PI3K/AKT
RAS/MAPK
RTK
TSC/mTOR
DARS
GDI2
P4HA2
TRUB1
Activation
Inhibition
None
B
Wilcoxon, p = 0.00026
Estimated IC50 of Cisplatin
LExp
HExp
Expression of DARS
Wilcoxon, p = 0.0024
Estimated IC50 of Gemcitabine
LExp
HExp
Expression of DARS
Wilcoxon, p = 5.2e−05
Estimated IC50 of AKT.inhibitor.VIII
LExp
HExp
Expression of GDI2
Wilcoxon, p = 0.00067
Estimated IC50 of Cisplatin
LExp
HExp
Expression of GDI2
Wilcoxon, p = 0.012
Estimated IC50 of Dasatinib
LExp
HExp
Expression of GDI2
Wilcoxon, p = 3.5e−06
Estimated IC50 of Gemcitabine
LExp
HExp
Expression of GDI2
Wilcoxon, p = 0.0032
Estimated IC50 of Cisplatin
LExp
HExp
Expression of P4HA2
Wilcoxon, p = 0.00017
Estimated IC50 of Gemcitabine
LExp
HExp
Expression of P4HA2
Wilcoxon, p = 0.00043
Estimated IC50 of AKT.inhibitor.VIII
LExp
HExp
Expression of TRUB1

Figure 6

Genetic and transcriptional changes and connections with immune cell populations.

(A) Correlation of DARS/GDI2/P4HA2/TRUB1 expression with immune infiltration level in GBM.

(B) DARS/GDI2/P4HA2/TRUB1 copy number alterations (CNV) affects the level of infiltration of B cells, CD8+ T cells, CD4+T celles, Macrophages, Neutrophils, and Dendritic cells in GBM.


A
Purity
B Cell
CD8+ T Cell
CD4+ T Cell
Macrophage
Neutrophil
Dendritic Cell
DARS Expression Level (log2 TPM)
GDI2 Expression Level (log2 TPM)
P4HA2 Expression Level (log2 TPM)
TRUB1 Expression Level (log2 TPM)
GBM
Infiltration Level
cor = 0.125
p = 5.07e−02
partial.cor = 0.002
p = 9.60e−01
partial.cor = 0.076
p = 1.19e−01
partial.cor = 0.14
p = 4.13e−03
partial.cor = 0.137
p = 4.89e−03
partial.cor = 0.245
p = 3.73e−07
partial.cor = 0.13
p = 7.64e−03
cor = −0.01
p = 8.35e−01
partial.cor = 0.034
p = 4.89e−01
partial.cor = −0.18
p = 2.19e−04
partial.cor = −0.078
p = 1.10e−01
partial.cor = 0
p = 9.92e−01
partial.cor = 0.184
p = 1.62e−04
partial.cor = 0.167
p = 8.21e−04
cor = −0.279
p = 6.06e−09
partial.cor = −0.239
p = 7.98e−07
partial.cor = −0.158
p = 1.19e−03
partial.cor = 0.075
p = 1.28e−01
partial.cor = 0.004
p = 9.41e−01
partial.cor = 0.025
p = 6.11e−01
partial.cor = 0.497
p = 1.66e−27
cor = −0.107
p = 2.1e−01
partial.cor = 0.031
p = 7.28e−01
partial.cor = −0.206
p = 1.87e−02
partial.cor = −0.088
p = 3.16e−01
partial.cor = 0.043
p = 6.24e−01
partial.cor = 0.151
p = 8.54e−02
partial.cor = −0.107
p = 2.24e−01

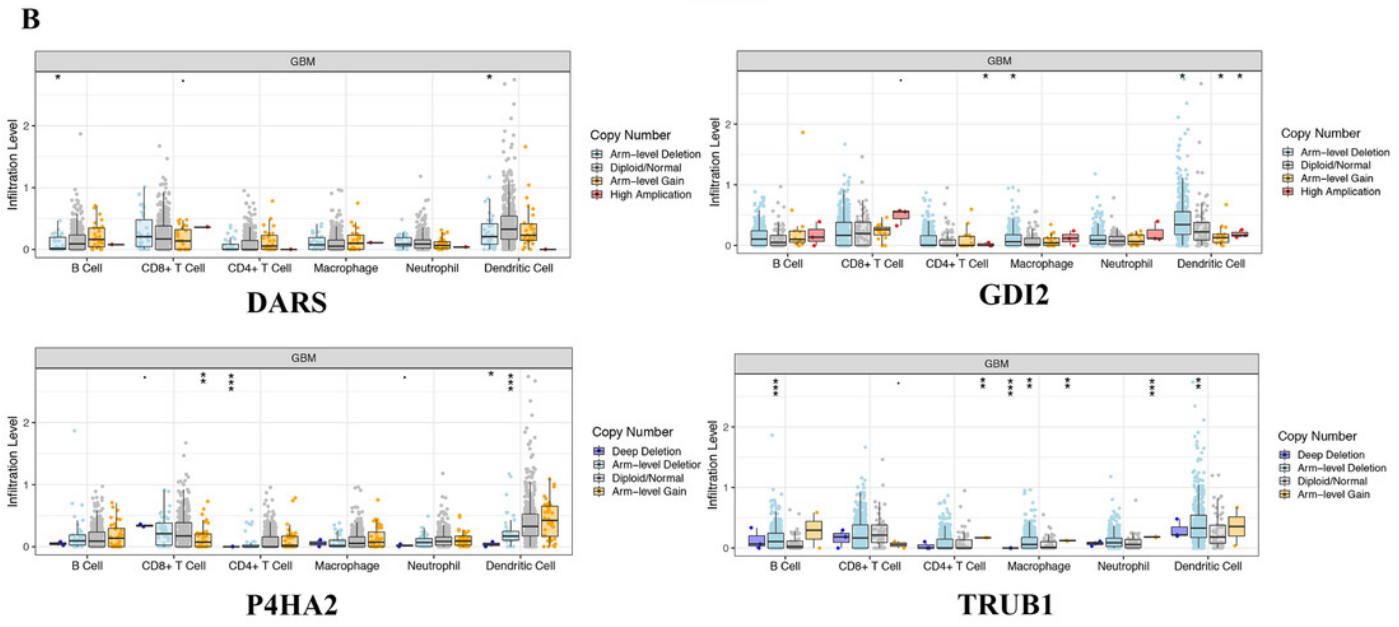

B
GBM
Infiltration Level
B Cell
CD8+ T Cell
CD4+ T Cell
Macrophage
Neutrophil
Dendritic Cell
Copy Number
Arm−level Deletion
Diploid/Normal
Arm−level Gain
High Amplication
Deep Deletion
DARS
GDI2
P4HA2
TRUB1

Figure 7

Genomic alterations of DARS/GDI2/P4HA2/TRUB1 in GBM.

OncoPrint of DARS/GDI2/P4HA2/TRUB1 alterations in GBM cohort. The different types of genetic alterations are highlighted in different colors. Expression profiles of mRNAs showing different expression (≥3-fold) compared to control were considered to be mRNA high, and vice versa for low.

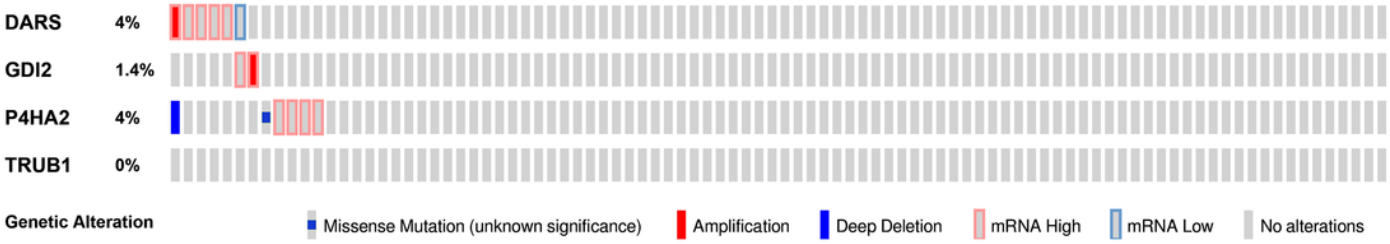


Figure 8

Fig. 8. GSVA analysis. GSVA of DARS/GDI2/P4HA2/TRUB1 gene sets in GBM.

(A) DARS. (B) GDI2. (C) P4HA2. (D) TRUB1.A t value > 1 or < − 1 represents statistically significant changes.

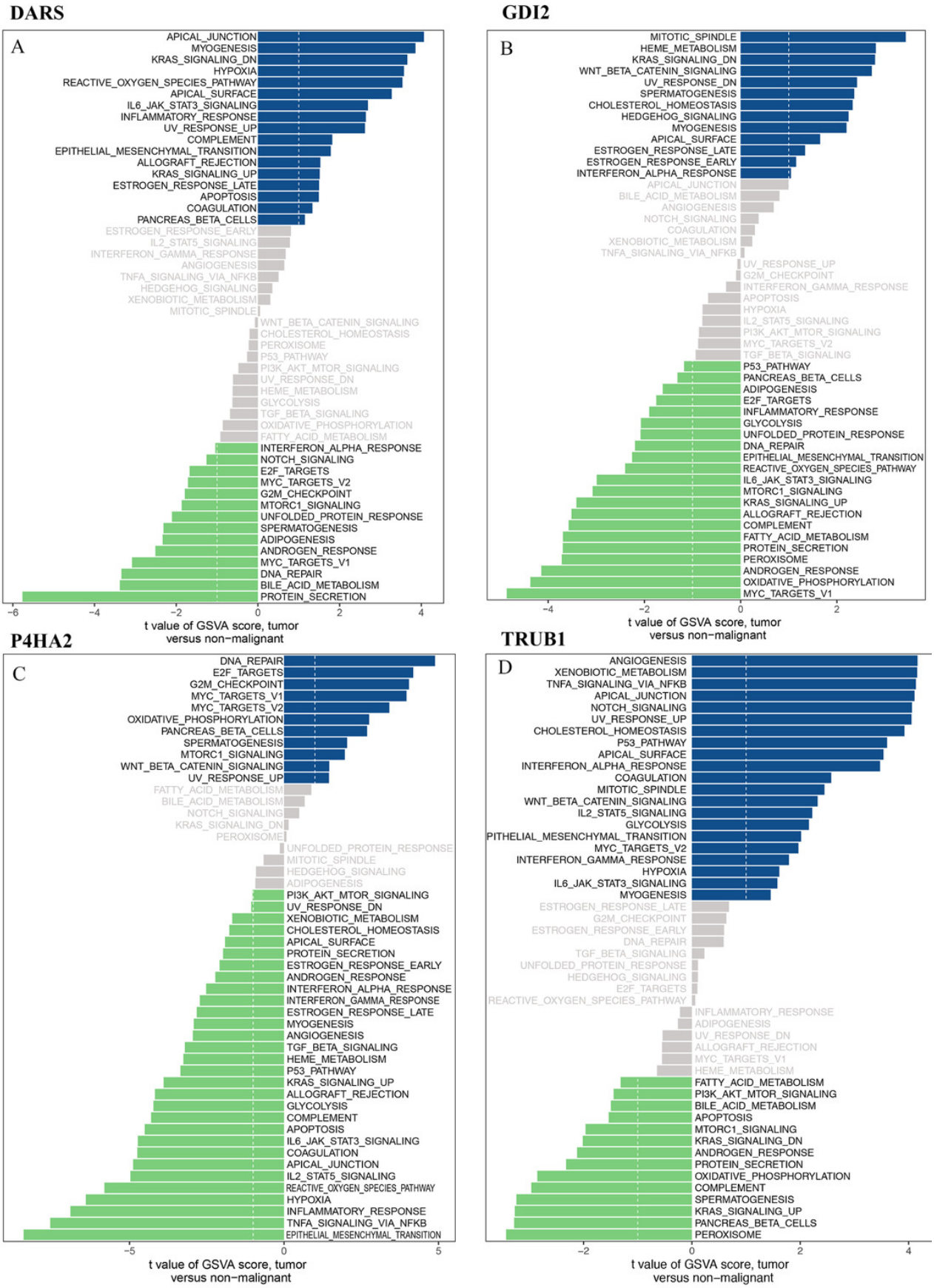

DARS
A
APICAL_JUNCTION
MYOGENESIS
KRAS_SIGNALING_DN
HYPOXIA
REACTIVE_OXYGEN_SPECIES_PATHWAY
APICAL_SURFACE
IL6_JAK_STAT3_SIGNALING
INFLAMMATORY_RESPONSE
UV_RESPONSE_UP
COMPLEMENT
EPITHELIAL_MESENCHYMAL_TRANSITION
ALLOGRAFT_REJECTION
KRAS_SIGNALING_UP
ESTROGEN_RESPONSE_LATE
APOPTOSIS
COAGULATION
PANCREAS_BETA_CELLS
ESTROGEN_RESPONSE_EARLY
IL2_STAT5_SIGNALING
INTERFERON_GAMMA_RESPONSE
ANGIOGENESIS
TNFA_SIGNALING_VIA_NFKB
HEDGEHOG_SIGNALING
XENOBIOTIC_METABOLISM
MITOTIC_SPINDLE
WNT_BETA_CATENIN_SIGNALING
CHOLESTEROL_HOMEOSTASIS
PEROXISOME
P53_PATHWAY
PI3K_AKT_MTOR_SIGNALING
UV_RESPONSE_DN
HEME_METABOLISM
GLYCOLYSIS
TGF_BETA_SIGNALING
OXIDATIVE_PHOSPHORYLATION
FATTY_ACID_METABOLISM
INTERFERON_ALPHA_RESPONSE
NOTCH_SIGNALING
E2F_TARGETS
MYC_TARGETS_V2
G2M_CHECKPOINT
MTORC1_SIGNALING
UNFOLDED_PROTEIN_RESPONSE
SPERMATOGENESIS
ADIPOGENESIS
ANDROGEN_RESPONSE
MYC_TARGETS_V1
DNA_REPAIR
BILE_ACID_METABOLISM
PROTEIN_SECRETION
−6
−4
−2
0
2
4
t value of GSVA score, tumor versus non−malignant
GDI2
B
MITOTIC_SPINDLE
HEME_METABOLISM
KRAS_SIGNALING_DN
WNT_BETA_CATENIN_SIGNALING
UV_RESPONSE_DN
SPERMATOGENESIS
CHOLESTEROL_HOMEOSTASIS
HEDGEHOG_SIGNALING
MYOGENESIS
APICAL_SURFACE
ESTROGEN_RESPONSE_LATE
ESTROGEN_RESPONSE_EARLY
INTERFERON_ALPHA_RESPONSE
APICAL_JUNCTION
BILE_ACID_METABOLISM
ANGIOGENESIS
NOTCH_SIGNALING
COAGULATION
XENOBIOTIC_METABOLISM
TNFA_SIGNALING_VIA_NFKB
UV_RESPONSE_UP
G2M_CHECKPOINT
INTERFERON_GAMMA_RESPONSE
APOPTOSIS
HYPOXIA
IL2_STAT5_SIGNALING
PI3K_AKT_MTOR_SIGNALING
MYC_TARGETS_V2
TGF_BETA_SIGNALING
P53_PATHWAY
PANCREAS_BETA_CELLS
ADIPOGENESIS
E2F_TARGETS
INFLAMMATORY_RESPONSE
GLYCOLYSIS
UNFOLDED_PROTEIN_RESPONSE
DNA_REPAIR
EPITHELIAL_MESENCHYMAL_TRANSITION
REACTIVE_OXYGEN_SPECIES_PATHWAY
IL6_JAK_STAT3_SIGNALING
MTORC1_SIGNALING
KRAS_SIGNALING_UP
ALLOGRAFT_REJECTION
COMPLEMENT
FATTY_ACID_METABOLISM
PROTEIN_SECRETION
PEROXISOME
ANDROGEN_RESPONSE
OXIDATIVE_PHOSPHORYLATION
MYC_TARGETS_V1
−4
−2
0
2
t value of GSVA score, tumor versus non−malignant
P4HA2
C
DNA_REPAIR
E2F_TARGETS
G2M_CHECKPOINT
MYC_TARGETS_V1
MYC_TARGETS_V2
OXIDATIVE_PHOSPHORYLATION
PANCREAS_BETA_CELLS
SPERMATOGENESIS
MTORC1_SIGNALING
WNT_BETA_CATENIN_SIGNALING
UV_RESPONSE_UP
FATTY_ACID_METABOLISM
BILE_ACID_METABOLISM
NOTCH_SIGNALING
KRAS_SIGNALING_DN
PEROXISOME
UNFOLDED_PROTEIN_RESPONSE
MITOTIC_SPINDLE
HEDGEHOG_SIGNALING
ADIPOGENESIS
PI3K_AKT_MTOR_SIGNALING
UV_RESPONSE_DN
XENOBIOTIC_METABOLISM
CHOLESTEROL_HOMEOSTASIS
APICAL_SURFACE
PROTEIN_SECRETION
ESTROGEN_RESPONSE_EARLY
ANDROGEN_RESPONSE
INTERFERON_ALPHA_RESPONSE
INTERFERON_GAMMA_RESPONSE
ESTROGEN_RESPONSE_LATE
MYOGENESIS
ANGIOGENESIS
TGF_BETA_SIGNALING
HEME_METABOLISM
P53_PATHWAY
KRAS_SIGNALING_UP
ALLOGRAFT_REJECTION
GLYCOLYSIS
COMPLEMENT
APOPTOSIS
IL6_JAK_STAT3_SIGNALING
COAGULATION
APICAL_JUNCTION
IL2_STAT5_SIGNALING
REACTIVE_OXYGEN_SPECIES_PATHWAY
HYPOXIA
INFLAMMATORY_RESPONSE
TNFA_SIGNALING_VIA_NFKB
EPITHELIAL_MESENCHYMAL_TRANSITION
−5
0
5
t value of GSVA score, tumor versus non−malignant
TRUB1
D
ANGIOGENESIS
XENOBIOTIC_METABOLISM
TNFA_SIGNALING_VIA_NFKB
APICAL_JUNCTION
NOTCH_SIGNALING
UV_RESPONSE_UP
CHOLESTEROL_HOMEOSTASIS
P53_PATHWAY
APICAL_SURFACE
INTERFERON_ALPHA_RESPONSE
COAGULATION
MITOTIC_SPINDLE
WNT_BETA_CATENIN_SIGNALING
IL2_STAT5_SIGNALING
GLYCOLYSIS
PITHELIAL_MESENCHYMAL_TRANSITION
MYC_TARGETS_V2
INTERFERON_GAMMA_RESPONSE
HYPOXIA
IL6_JAK_STAT3_SIGNALING
MYOGENESIS
ESTROGEN_RESPONSE_LATE
G2M_CHECKPOINT
ESTROGEN_RESPONSE_EARLY
DNA_REPAIR
TGF_BETA_SIGNALING
UNFOLDED_PROTEIN_RESPONSE
HEDGEHOG_SIGNALING
E2F_TARGETS
REACTIVE_OXYGEN_SPECIES_PATHWAY
INFLAMMATORY_RESPONSE
ADIPOGENESIS
UV_RESPONSE_DN
ALLOGRAFT_REJECTION
MYC_TARGETS_V1
HEME_METABOLISM
FATTY_ACID_METABOLISM
PI3K_AKT_MTOR_SIGNALING
BILE_ACID_METABOLISM
APOPTOSIS
MTORC1_SIGNALING
KRAS_SIGNALING_DN
ANDROGEN_RESPONSE
PROTEIN_SECRETION
OXIDATIVE_PHOSPHORYLATION
COMPLEMENT
SPERMATOGENESIS
KRAS_SIGNALING_UP
PANCREAS_BETA_CELLS
PEROXISOME
−2
0
2
4
t value of GSVA score, tumor versus non−malignant

Table 1(on next page)

Table1: Statistics of genes in darkturquoise modules.

Table1 Statistics of genes in darkturquoise modules.

1 Table1 Statistics of genes in darkturquoise modules.

2

Gene	P value
TRUB1	2.25E-02
P4HA2	1.02E-02
DARS	4.27E-02
FKBP1B	6.13E-03
NRL	2.20E-02
CORO6	1.83E-02
LRRC43	4.65E-02
GAS6	3.63E-02
SPAG4	2.07E-03
PRKAR2B	1.48E-02
CAMSAP2	1.31E-02
CD24	2.52E-02
GDI2	4.42E-03
DLEU1	1.45E-02

3