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.

1 **Exploring the potential biomarkers for prognosis of**

2 **glioblastoma via Weighted Gene Co-expression**

3 **Network Analysis**

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

5   1   Department of Neurology and Stroke Center, The First Hospital of China Medical

6   University, Shenyang 110000, PR China.

7   2   Department of Geriatrics, The First Affiliated Hospital, China Medical University,

8   Shenyang 110000, PR China.

10 **Corresponding Author:** Chuansheng Zhao

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

12 Shenyang 110000, PR China

13 Email Adress: cszhao@cmu.edu.cn

15 **Abstract:**

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

Introduction

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

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

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

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

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

Materials and methods

Data information and construction of WGCNA

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

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

Enrichment analysis of gene module function

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

Identifying Key Genes

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

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

97   ***Download and Pre-processing Data From TCGA***

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

102   ***Co-expression analysis***

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

107   ***GSCALite and GDSC***

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

119   ***TIMER database analysis***

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

126   ***Gene functional analysis***

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

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

134   ***Statistical analysis***

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

137

138   # Result

139   ***Identification of gene co-expression modules***

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

152   ***Correlation of modules and clinical traits***

153   In order to study the relationship between these modules and the prognosis of GBM, we
154   investigated the correlation between each module and the prognosis of GBM.We found that the
155   darkturquoise module had the highest correlation with disease relapse(r = -0.62, p=0.01)(Fig.
156   2A, B). We used Metascape to analyze the function and pathway of the darkturquoise module.
157   Metascape can identify the enrichment process in the gene list and the association between
158   enrichment processes [14, 15] by querying many databases,such as GO functional, Hallmark Gene
159   Sets, and KEGG pathways (Tripathi et al. 2015; Zheng et al. 2018). Based on GO enrichment
160   analysis, it was found that the co-expressed genes within the modules of interest s mainly related
161   to the steady state of cellular transition metal ion homeostasis, protein hydroxylation, intrinsic
162   apoptotic signaling pathway, ncRNA metabolic process(Fig. 2C).The KEGG pathway analysis
163   revealed that the co-expressed genes within the modules of interest was mostly enriched in the
164   'Ferroptosis'(Fig. 2C). In addition, the enrichment processes were highly connected and could be
165   clustered into a complete network (Fig. 2C) . These results indicated that these functions were
166   related in the occurrence and development of GBM.

167 ***Identification of key genes in darkturquoise module***

168 According to the WGCNA theory, key genes have the highest connectivity in the module, which
169 determine the biological significance of the module and therefore influence the survial of patients
170 intensively(Chen et al. 2019).Therefore, we searched for hub genes in the darkturquoise
171 module.We analyzed the Kaplan-Meier survival curves of 139 genes in the darkturquoise module.
172 14 genes with significant survival analysis results (p<0.05) were selected for sequencing(Table.
173 1). Further determination of the key gene requires a combination of its expression, typicality in
174 previous studies, and previous research experience in our lab. Finally, we found that four genes
175 (DARS/GDI2/P4HA2/TRUB1) had an impact on the survival rate of patients with GBM, and were
176 also confirmed to be related to tumorigenesis. At the same time, they were not confirmed to
177 participate in the occurrence and development of GBM, which met the conditions for our further
178 research (Fig. 3). Survival analysis showed that the patients with low expression of DARS/GDI2/
179 TRUB1 and high expression of P4HA2 had poor prognosis.

180

181 ***Analysis of the co-expression of key genes***

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

198

199 ***Cancer-related pathways and drug sensitivity analysis of key genes.***

200 First, we investigated the role of key genes in all well-known cancer-related pathways, as the
201 following: TSC/mTOR, RTK, RAS/MAPK, PI3K/AKT, Hormone ER, Hormone AR, EMT, DNA
202 Damage Response, Cell Cycle, Apoptosis pathways. The results found that DARS participated in
203 the TSC/mTOR pathway activation; GDI2 was involved in Apoptosis, TSC/mTOR pathway
204 activation; P4HA2 was involved in DNA Damage Response, EMT, Hormone AR, Hormone ER,
205 RAS/MAPK and TSC/mTOR pathway; TRUB1 was involved in Apoptosis, DNA Damage

206 Response, EMT, Hormone AR, PI3K/AKT and TSC/mTOR pathway (Fig. 5A). To investigate
207 whether the expression of DARS / GDI2 / P4HA2 / TRUB1 in GBM had an impact on treatment
208 (e.g. chemotherapies), we constructed a predictive model on six commonly used chemo drugs (i.e.
209 AKT.inhibitor, Cisplatin, Dasatinib, Erlotinib, Gefitinib, and Gemcitabine) and confirmed that
210 high expression of DARS was less sensitive to Cisplatin(p=0.00026) and Gemcitabine(p=0.0024),
211 high expression of GDI2 was less sensitive to AKT.inhibitor(p=5.2e−05), Cisplatin(p=0.00067),
212 Dasatinib(p=0.012) and Gemcitabine(p=3.5e−06), low expression of P4HA2 was less sensitive to
213 Cisplatin(p=0.0032), and Gemcitabine(p=0.00017),and high expression of TRUB1 was less
214 sensitive to AKT.inhibitor(p=0.00043)(Fig.5B).

215

Immune cells infiltration analysis

217 In view of the obvious prognostic value of DARS/GDI2/P4HA2/TRUB1 gene, we used the
218 TIMER database to determine whether there was an association between tumor-infiltrating and
219 immune cells and DARS/GDI2/P4HA2/TRUB1 expression.Results showed that DARS expression
220 was positively correlated with tumor purity, P4HA2 and B cells (partial cor=-0.239, p=7.89e-07),
221 P4HA2 and CD8+ T cells (partial cor=-0.158, p=1.19e-03), TRUB1 and CD8+ T cells (partial
222 cor=-0.206, p=1.87e-02), DARS and neutrophils (partial cor= 0.245, p=3.73e-07), GDI2 and
223 neutrophils (partial cor=0.184, p=1.62e-04), GDI2 and Dendritic cells (partial cor=0.167, p=6.21e-
224 04) P4HA2 and B cells (partial cor=-0.239, p=7.89e-07), P4HA2 and CD8+ T cells (partial cor=-
225 0.158, p=1.19e-03), TRUB1 and CD8+ T cells (partial cor=-0.206, p=1.87e-02), DARS and
226 neutrophils (partial cor= 0.245, p=3.73e-07), GDI2 and neutrophils (partial cor=0.184, p=1.62e-
227 04), GDI2 and Dendritic cells (partial cor=0.167, p=6.21e-04) (Fig. 6A). We also explored the
228 correlation between tumor immune cell infiltration and somatic copy number alterations (SCNAs).
229 The samples were divided into four types according to the copy number of genes.The distribution
230 of infiltrating immune cells among the four types of samples was compared, as shown in Fig. 6B.
231 We found that the various forms of mutations carried by the DARS / GDI2 / P4HA2 / TRUB1
232 gene can usually suppress immune infiltration, including CD8+T cells, neutrophils, dendritic cells,
233 macrophages, CD4+T cells, and B cells. Also, we found that these four pivotal genes had a greater
234 effect on immune infiltration than alterations in the genes.

235

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

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

244

Gene functional analysis

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

252

Discussion

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

261   GBM, as a highly heterogeneous tumor harboring multiple genetic alterations(Harter et al.
262 2014), molecular heterogeneity affects the effectiveness of single-molecule markers in predicting
263 prognosis(Tonry 2020). At the same time, some studies have found that the high recurrence rate
264 of GBM is related to the expression of strong proliferation genes of cells(Lara-Velazquez et al.
265 2020). And these processes usually involve multiple genes(Malik et al. 2020). Therefore, we
266 believe that multi-gene markers have a higher predictive power for GBM prognosis than single-
267 gene marker. We built a multi-gene markers model for predicting GBM prognosis, and validated
268 the multi-gene markers model through strategies including training, testing, and independent cross-
269 validation. The above strategies significantly improve the predictive ability of genetic markers (Li
270 et al. 2019b).

271   In our research and analysis, the results of GO and KEGG analysis indicate that cell transition
272 metal ion homeostasis, protein hydroxylation, intrinsic apoptotic signaling pathway and other
273 processes may play an important role in GBM. Among them,transition metals are critical for
274 many metabolic processes(Nelson & N. 2014), and their steady state is vital to life. Aberrations in
275 the cellular metal ion concentrations may lead to cell death and severe diseases such as cancer(Pi
276 et al. 2020). Hydroxylation is a post-translational modification affecting protein stability, activity
277 or interactome(Zurlo & Zhang 2020). Many cancers are related to protein hydroxylation, such as
278 breast cancer(Zurlo & Zhang 2020), gastric cancer(Li et al. 2020), and prostate cancer(Della-Flora
279 et al. 2020). For example, a study found that a set of enzymes PLOD1、PLOD2 and PLOD3
280 involved in the hydroxylation of lysine and stabilization of collagen by crosslinks, which up-
281 regulated expression in gastric cancer patients(Li et al. 2020). Similarly, intrinsic apoptotic
282 signaling pathway can activate or inactivate multiple signaling pathways and inhibit multiple
283 tumor suppressor genes, thereby promoting tumor progression. Almost all cancers involve intrinsic
284 apoptotic signaling pathway, including renal cell carcinoma(Chae et al. 2020) and multiple

285 myeloma(Chen et al. 2020a). Combined with the above results, we believe that DARS / GDI2 /
286 P4HA2 / TRUB1 may be involved in these processes to affect the occurrence and development of
287 GBM disease, which is also consistent with our Drug sensitivity analysis results. Among them, the
288 DARS gene encodes the aspartyl-tRNA synthetase(Dominik et al. 2018), which pairs aspartate
289 with its corresponding tRNA. Missense mutations in the gene encoding DARS can lead to
290 leukocyte dystrophy, accompanied by a marked reduction in myelin sheath, abnormal movement
291 and cognitive impairment (Fröhlich et al. 2018). There are no related reports about the relations
292 between DARS and GBM.According to our research, DARS may participate in TSC/mTOR
293 signaling, by regulating GBM cell growth process. GDI2 controls the activity of Rho GTPase's
294 pathway to regulatory guanine nucleotide exchange factor and GTPase activating protein, and may
295 play a role in tumor cell apoptosis. This is also in line with our results. At the same time, a recent
296 study shows that RhoGDI2 suppresses lung metastasis in mice by reducing tumor versican
297 expression and macrophage infiltration. The expression of P4HA2 increased in head and neck
298 squamous cell carcinoma (HNSCC)(Kisoda et al. 2020), Oral Squamous Cell Carcinoma
299 (OSCC)(Reis et al. 2020), cervical cancer(Cao et al. 2020b) and other cancers. Especially,we
300 found that P4HA2 are markedly upregulated in cervical cancer tissues and upregulation of P4HA2
301 was associated with shorter overall survival (OS) and relapse-free survival (RFS)(Cao et al.
302 2020b). In GBM, we found that P4HA2 is mainly involved in the process of inhibiting DNA
303 damage, and is also related to EMT, Hormone AR, Hormone ER, RAS / MAPK, TSC / mTOR
304 and other pathways. TRUB1 mRNA is widely expressed in various human tissues (especially
305 heart, skeletal muscle and liver),but there are few studies on its relationship with cancer(Zucchini
306 et al. 2003). In our research, we analyzed that TRUB1 is mainly involved in Apoptosis, DNA
307 damage, EMT, PI3K / AKT and other processes.

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

323 Further TIMER analysis indicated that the immune system had a good effect on tumor
324 microenvironment, and that the mutations of DARS / GDI2 / P4HA2 / TRUB1 had important
325 application value in tumor immunology. Finally, we conducted a comprehensive evaluation of
326 gene sets using GSVA and we found that the DARS/GDI2/P4HA2/TRUB1 gene sets are closely
327 related to the occurrence of cancer.For instance, the APICAL_ JUNCTION in the DARS gene set
328 is more common in highly differentiated epithelial cells, such as colon cancer cells(Nair-Menon et
329 al. 2020).MITOTIC_SPINDLE in the GDI2 gene set, the mitotic spindle inhibitor is one of the

330 most commonly used chemotherapeutics now(Bukowski et al. 2020). DNA_REPAIR in the
331 P4HA2 gene set and ANGIOGENESIS in the TRUB1 gene set are also two important mechanisms
332 of cancer development .

333   In recent years, with the GBM genes related to the occurrence and prognosis of feature
334 recognition in many studies. Such as Chen X found the ASPMexpression pattern from the database
335 showed that it is highly expressed in GBM tissue, and patients with high expression of ASPM have
336 a poor prognosis(Chen et al. 2020b). Recently, a bioinformatic analysis of 123 GBM patients has
337 established a 14-mRNA prognostic signature, which could be used to classify GBM patients into
338 low   and   high   risk   groups(Arimappamagan   et   al.   2013).To   our   knowledge,   the
339 DARS/GDI2/P4HA2/TRUB1 that we identified are new GBM biomarkers because they have
340 never been reported to be associated with the development and progression of GBM(Lu et al.
341 2020). At the same time, compared with the traditional typing methods, the multi-gene markers
342 model   has   many   advantages,   such   as   high   prediction   accuracy   and   personalized   detection
343 results(Albuquerque et al. 2012). Therefore, multi-gene markers have a good application prospect
344 in clinical practice. In our study, we built and verified the characteristic of the four genes through
345 analyzing   the   two   independent   data   sets.   More   reasonable   use   of   biometrics   and   multiple
346 independent data sets of mutual verification makes our results more reliable.

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

Conclusion

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

363

Acknowledgement

365 None

Reference

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

370 Arimappamagan A, Somasundaram K, Thennarasu K, Peddagangannagari S, Srinivasan H,
371 Shailaja BC, Samuel C, Patric IR, Shukla S, Thota B, Prasanna KV, Pandey P, Balasubramaniam
372 A, Santosh V, Chandramouli BA, Hegde AS, Kondaiah P, and Sathyanarayana Rao MR. 2013. A
373 fourteen gene GBM prognostic signature identifies association of immune response pathway and
374 mesenchymal subtype with high risk group. *PLoS ONE* 8:e62042. 10.1371/journal.pone.0062042

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

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

379 Cao Y, Han Q, Li J, Jia Y, Zhang R, and Shi H. 2020b. P4HA2 contributes to cervical cancer
380 progression via inducing epithelial-mesenchymal transition. *J Cancer* 11:2788-2799.
381 10.7150/jca.38401

382 Cetani F, Marcocci C, Torregrossa L, and Pardi E. 2019. Atypical parathyroid adenomas:
383 challenging lesions in the differential diagnosis of endocrine tumors. *Endocr Relat Cancer*
384 26:R441-R464. 10.1530/ERC-19-0135

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

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

392 Chen L, Peng T, Luo Y, Zhou F, Wang G, Qian K, Xiao Y, and Wang X. 2019. ACAT1 and
393 Metabolism-Related Pathways Are Essential for the Progression of Clear Cell Renal Cell
394 Carcinoma (ccRCC), as Determined by Co-expression Network Analysis. *Front Oncol* 9:957.
395 10.3389/fonc.2019.00957

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

399 Della-Flora A, Wilde ML, Pinto IDF, Lima EC, and Sirtori C. 2020. Degradation of the anticancer
400 drug flutamide by solar photo-Fenton treatment at near-neutral pH: Identification of transformation
401 products and in silico (Q)SAR risk assessment. *Environ Res* 183:109223.
402 10.1016/j.envres.2020.109223

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

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

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

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

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

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

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

423 Irshad K, Mohapatra SK, Srivastava C, Garg H, Mishra S, Dikshit B, Sarkar C, Gupta D, Chandra
424 PS, Chattopadhyay P, Sinha S, and Chosdol K. 2015. A combined gene signature of hypoxia and
425 notch pathway in human glioblastoma and its prognostic relevance. *PLoS ONE* 10:e0118201.
426 10.1371/journal.pone.0118201

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

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

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

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

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

442 Li T, Fan J, Wang B, Traugh N, Chen Q, Liu JS, Li B, and Liu XS. 2017b. TIMER: A Web Server
443 for Comprehensive Analysis of Tumor-Infiltrating Immune Cells. *Cancer Research* 77:e108-e110.
444 10.1158/0008-5472.Can-17-0307

445 Li W, Lu J, Ma Z, Zhao J, and Liu J. 2019b. An Integrated Model Based on a Six-Gene Signature
446 Predicts Overall Survival in Patients With Hepatocellular Carcinoma. *Front Genet* 10:1323.
447 10.3389/fgene.2019.01323

448 Liu A, Zhang S, Shen Y, Lei R, and Wang Y. 2019. Association of mRNA expression levels of
449 Cullin family members with prognosis in breast cancer. *Medicine* 98.
450 10.1097/md.0000000000016625

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

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

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

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

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

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

465 Pi H, Wendel BM, and Helmann JD. 2020. Dysregulation of Magnesium Transport Protects
466 Bacillus subtilis against Manganese and Cobalt Intoxication. *J Bacteriol* 202. 10.1128/JB.00711-
467 19

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

471 Reis PP, Tokar T, Goswami RS, Xuan Y, Sukhai M, Seneda AL, Moz LES, Perez-Ordonez B,
472 Simpson C, Goldstein D, Brown D, Gilbert R, Gullane P, Irish J, Jurisica I, and Kamel-Reid S.
473 2020. A 4-gene signature from histologically normal surgical margins predicts local recurrence in
474 patients with oral carcinoma: clinical validation. *Sci Rep* 10:1713. 10.1038/s41598-020-58688-y

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

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

482 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.
483 CMA1 is potent prognostic marker and associates with immune infiltration in gastric cancer.
484 *Autoimmunity*:1-8. 10.1080/08916934.2020.1735371

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

487 syndrome. *Scientific Reports*.

488 Szopa W, Burley TA, Kramer-Marek G, and Kaspera W. 2017. Diagnostic and Therapeutic
489 Biomarkers in Glioblastoma: Current Status and Future Perspectives. *Biomed Res Int*
490 2017:8013575. 10.1155/2017/8013575

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

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

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

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

501 Tripathi S, Pohl Marie O, Zhou Y, Rodriguez-Frandsen A, Wang G, Stein David A, Moulton
502 Hong M, DeJesus P, Che J, Mulder Lubbertus CF, Yánguez E, Andenmatten D, Pache L,
503 Manicassamy B, Albrecht Randy A, Gonzalez Maria G, Nguyen Q, Brass A, Elledge S, White M,
504 Shapira S, Hacohen N, Karlas A, Meyer Thomas F, Shales M, Gatorano A, Johnson Jeffrey R,
505 Jang G, Johnson T, Verschueren E, Sanders D, Krogan N, Shaw M, König R, Stertz S, García-
506 Sastre A, and Chanda Sumit K. 2015. Meta- and Orthogonal Integration of Influenza "OMICs"
507 Data Defines a Role for UBR4 in Virus Budding. *Cell Host & Microbe* 18:723-735.
508 10.1016/j.chom.2015.11.002

509 Wang W, Xing H, Huang C, Pan H, and Li D. 2020. Identification of pancreatic cancer type related
510 factors by Weighted Gene Co-Expression Network Analysis. *Med Oncol* 37:33. 10.1007/s12032-
511 020-1339-0

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

515 Xing Z-y, Sun L-g, and Guo W-j. 2015. Elevated expression of Notch-1 and EGFR induced
516 apoptosis in glioblastoma multiforme patients. *Clinical Neurology and Neurosurgery* 131:54-58.
517 10.1016/j.clineuro.2015.01.018

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

520 Zhang L, Zhang Z, and Yu Z. 2019. Identification of a novel glycolysis-related gene signature for
521 predicting metastasis and survival in patients with lung adenocarcinoma. *J Transl Med* 17:423.
522 10.1186/s12967-019-02173-2

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

526   10.1002/jcb.27677

527   Zhou Z, Cheng Y, Jiang Y, Liu S, Zhang M, Liu J, and Zhao Q. 2018. Ten hub genes associated
528   with progression and prognosis of pancreatic carcinoma identified by co-expression analysis.
529   *International Journal of Biological Sciences* 14:124-136. 10.7150/ijbs.22619

530   Zucchini C, Strippoli P, Biolchi A, Solmi R, Lenzi L, D'Addabbo P, Carinci P, and Valvassori L.
531   2003. The human TruB family of pseudouridine synthase genes, including the Dyskeratosis
532   Congenita 1 gene and the novel member TRUB1. *Int J Mol Med* 11:697-704.

533   Zurlo G, and Zhang Q. 2020. Adenylosuccinate lyase hydroxylation contributes to triple negative
534   breast cancer via the activation of cMYC. *Mol Cell Oncol* 7:1707045.
535   10.1080/23723556.2019.1707045

536
537

Figure Caption

538

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

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

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

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

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

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

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

581 **Fig. 7. Genomic alterations of DARS/GDI2/P4HA2/TRUB1 in GBM.** OncoPrint of
582 DARS/GDI2/P4HA2/TRUB1 alterations in GBM cohort. The different types of genetic alterations
583 are highlighted in different colors. Expression profiles of mRNAs showing different expression
584 (≥3-fold) compared to control were considered to be mRNA high, and vice versa for low.

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

588

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

B

Scale independence

C

Mean connectivity

E

Network heatmap plot, selected genes

D

Cluster Dendrogram

Height

Module colors
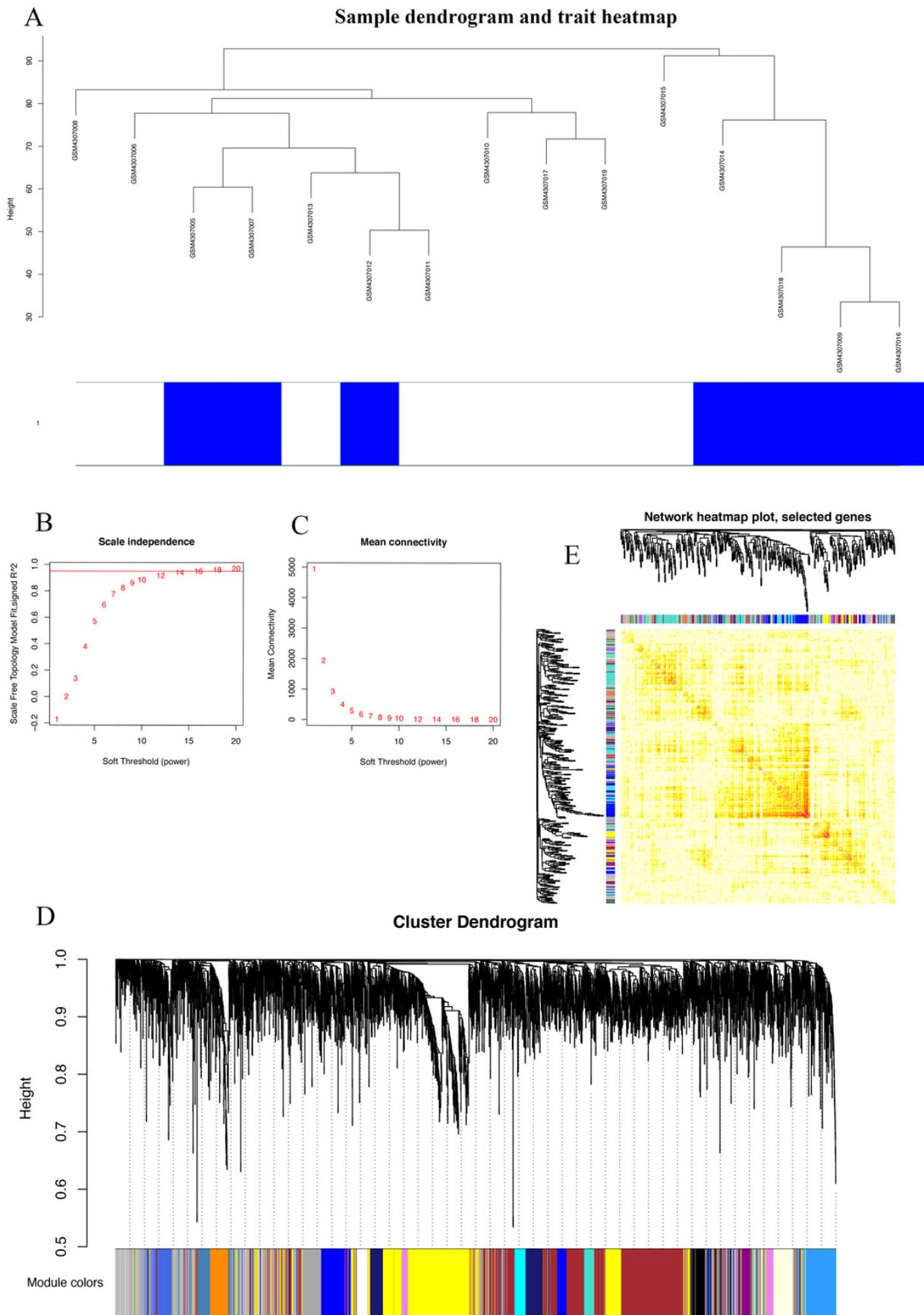
Figure 2

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.
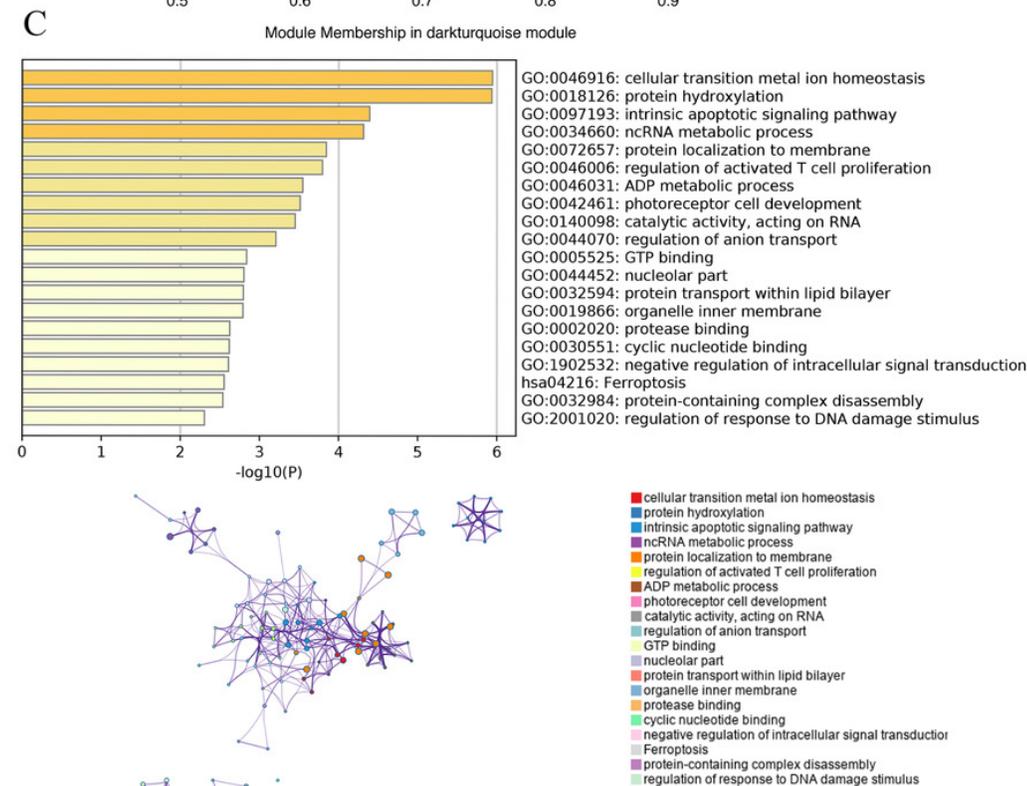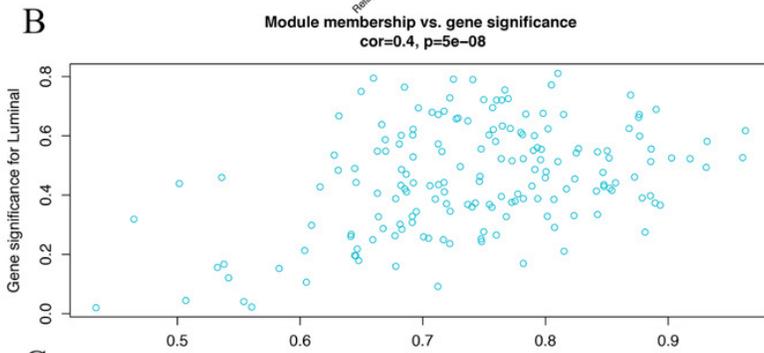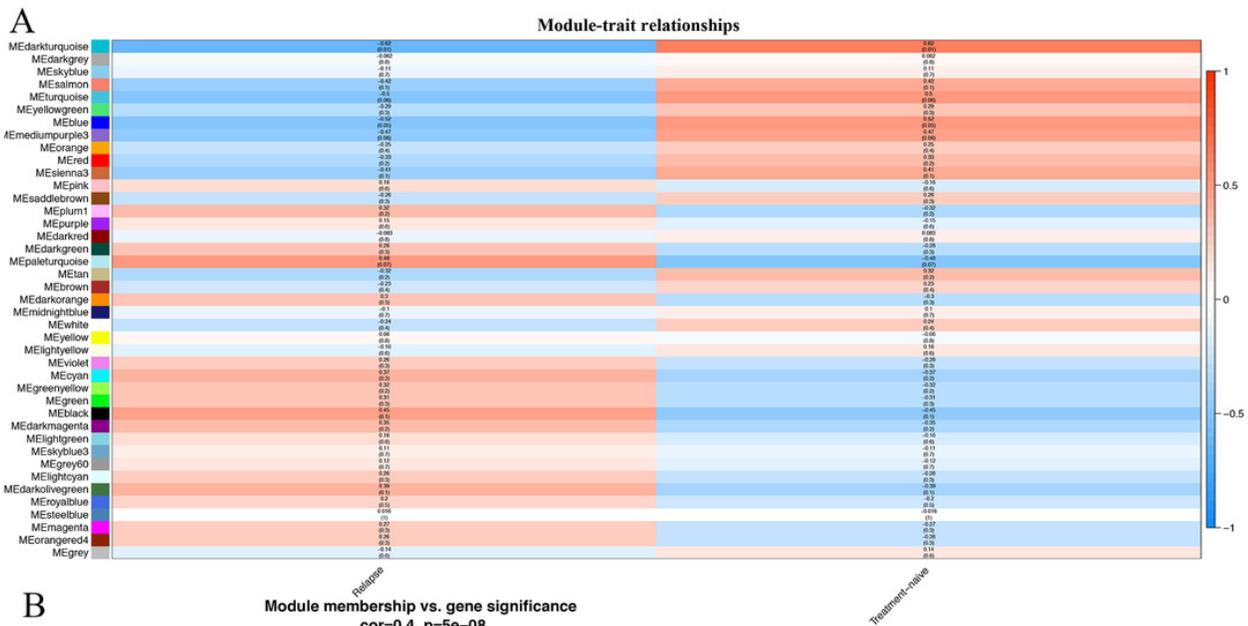
Figure 3

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)

C

Survival curve (p=1.019e-02)

B

Survival curve (p=4.42e-03)

D

Survival curve (p=2.247e-02)
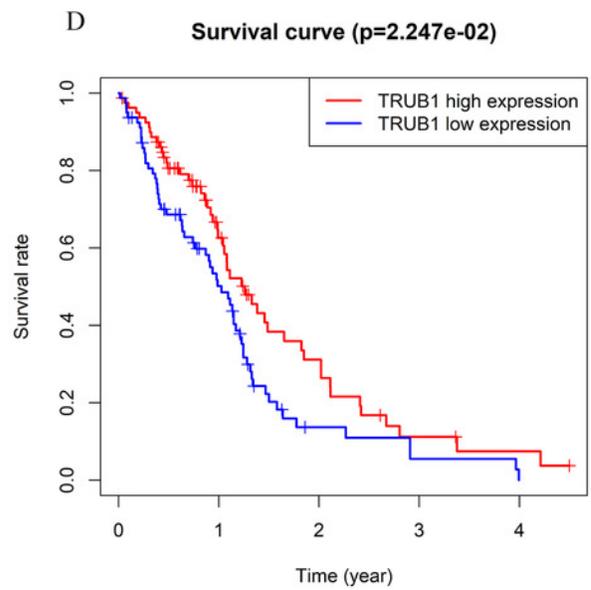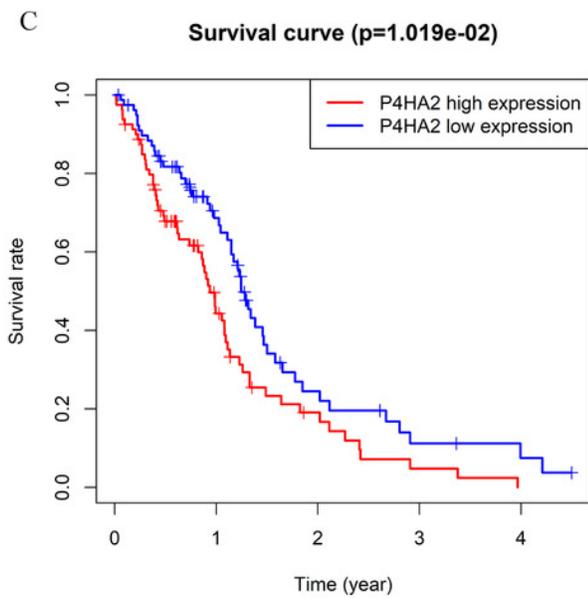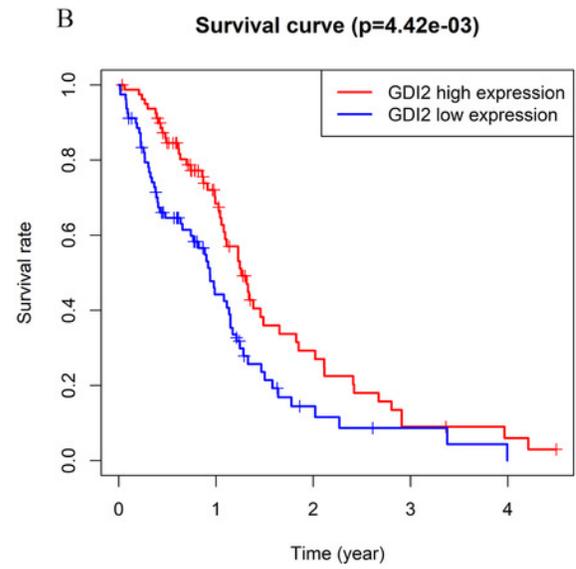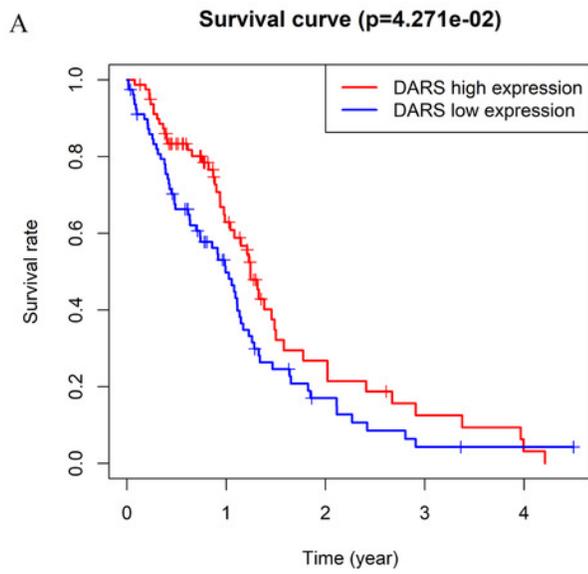
Figure 4

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.
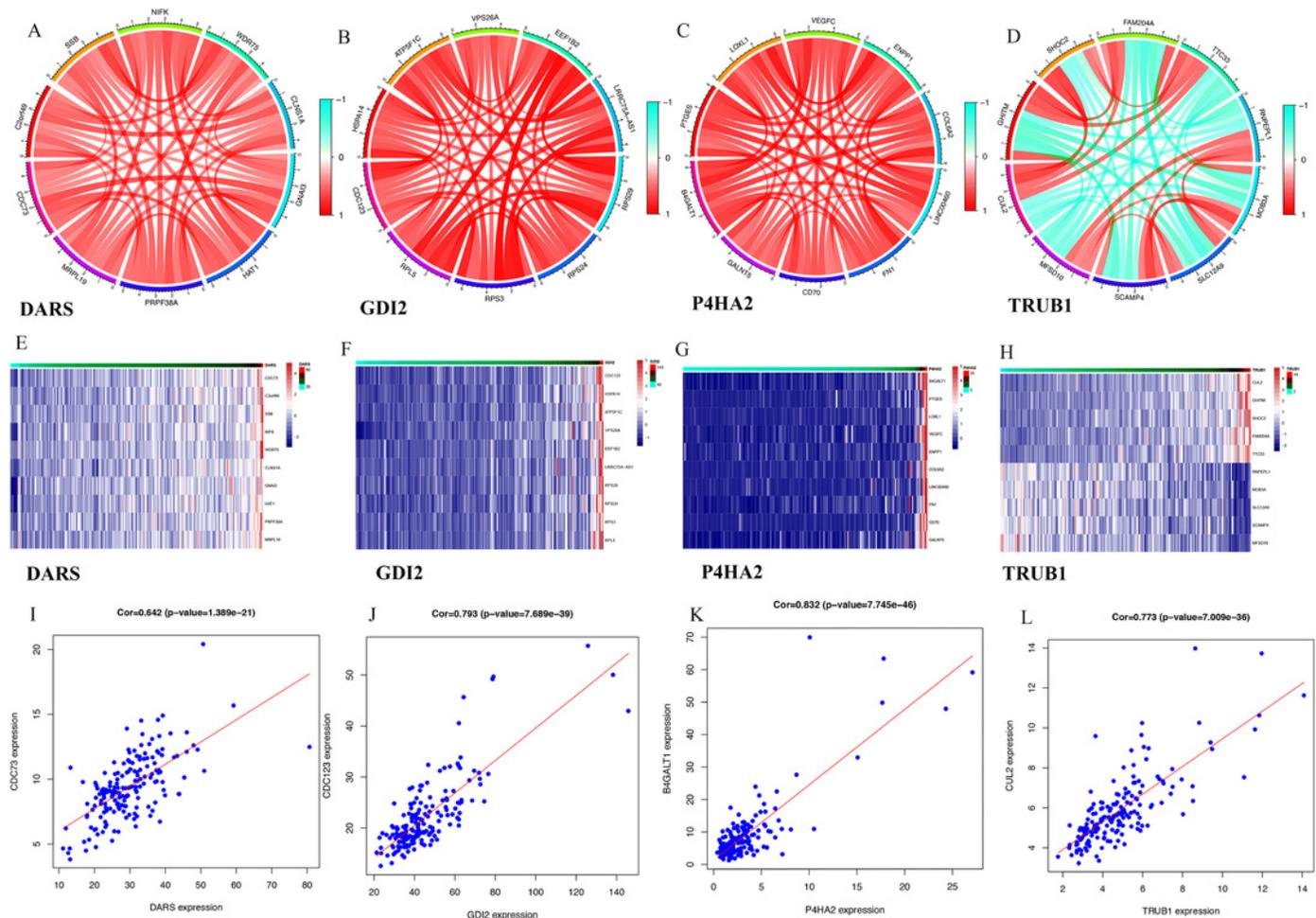
Figure 5

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.
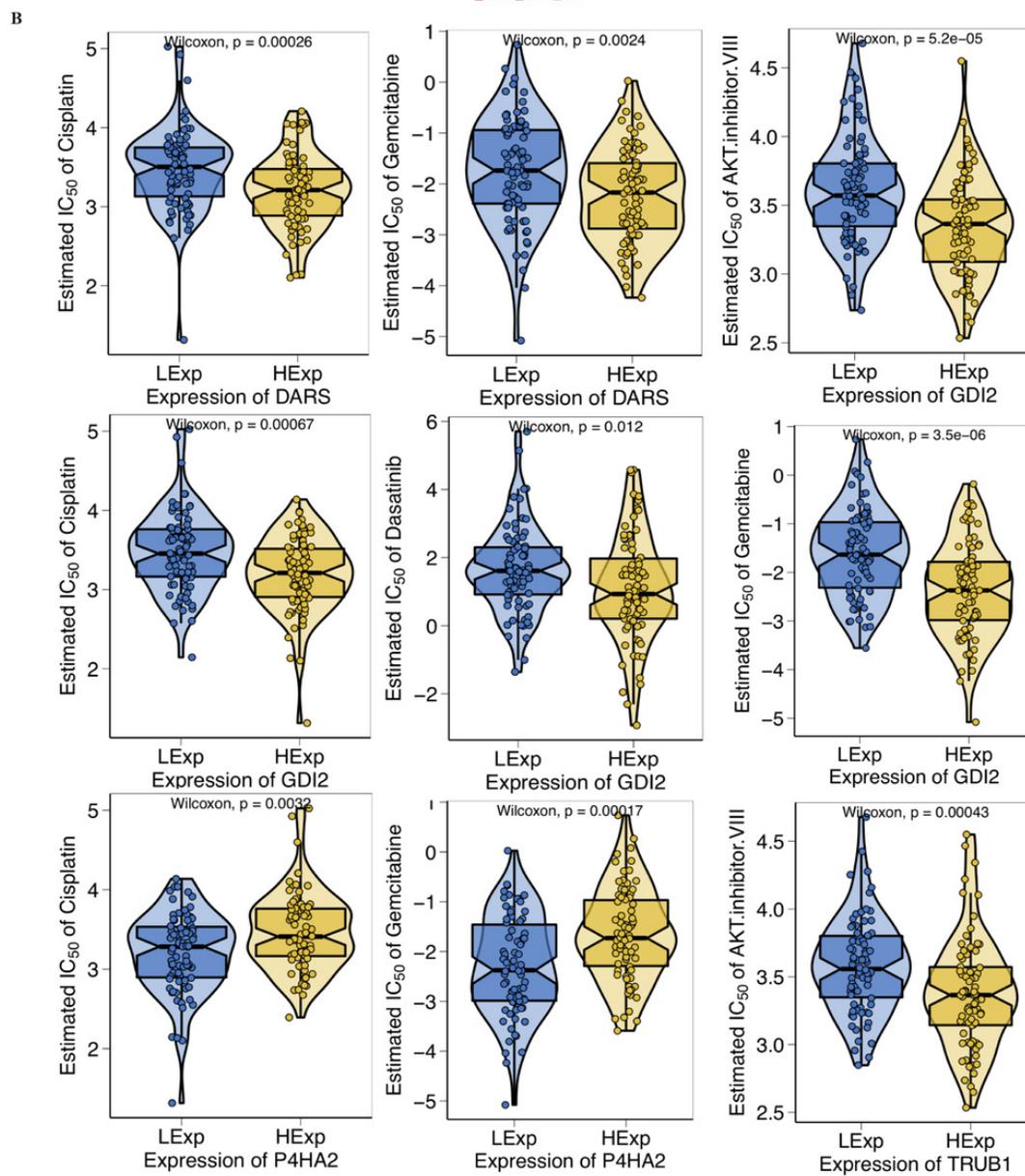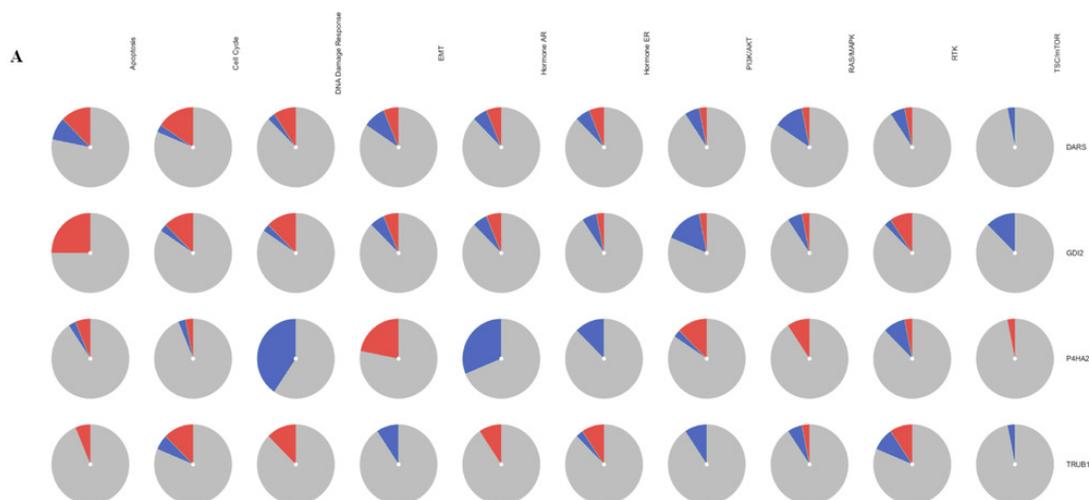
Figure 6

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.
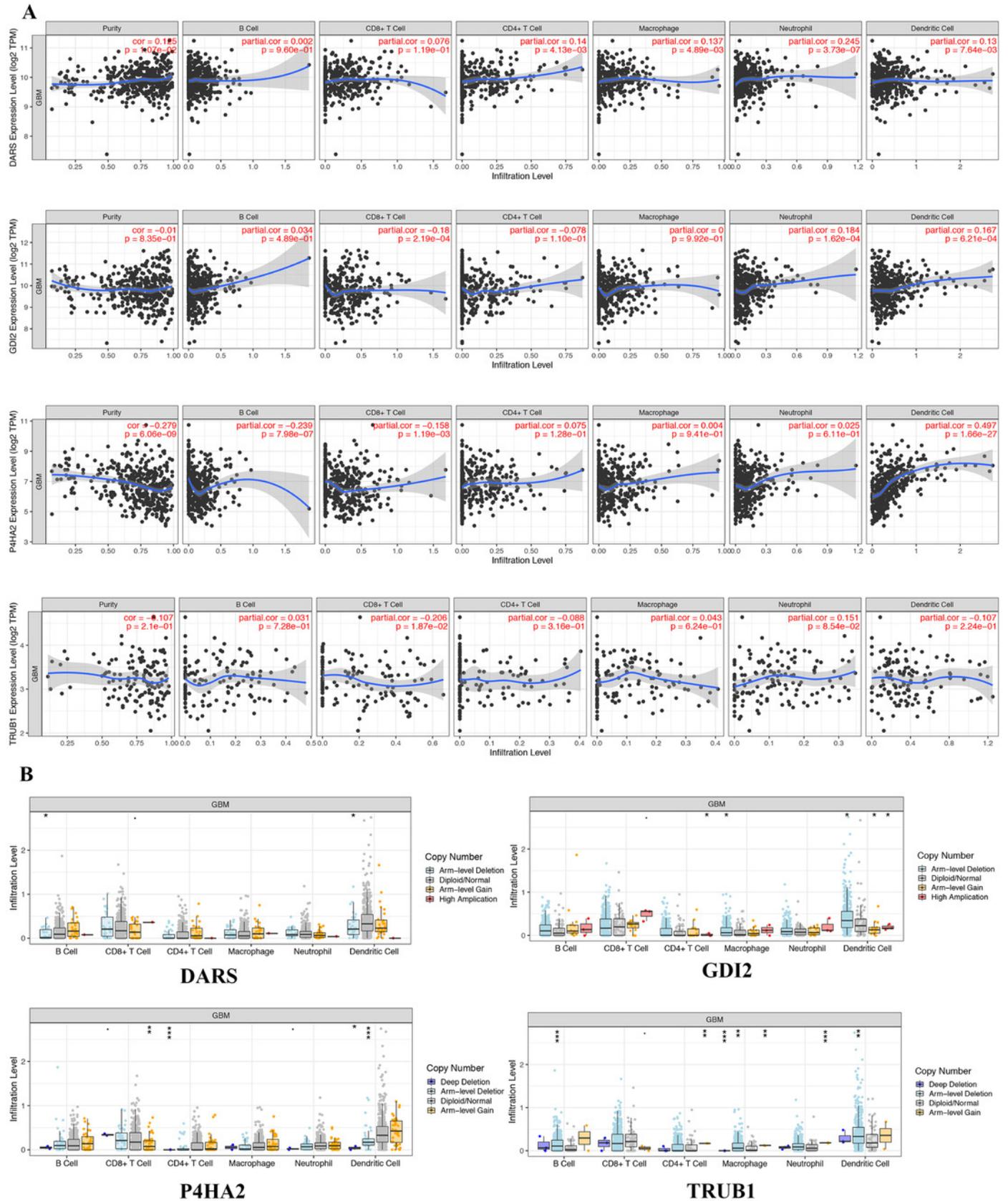
Figure 7

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.
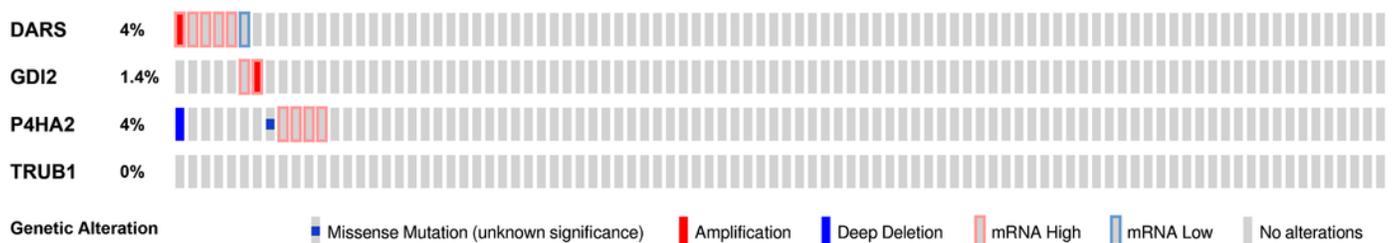
Figure 8

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.
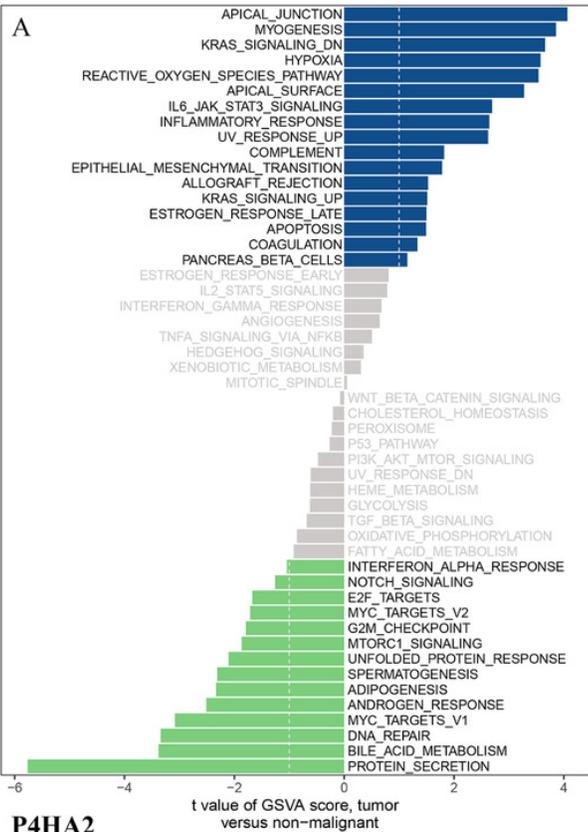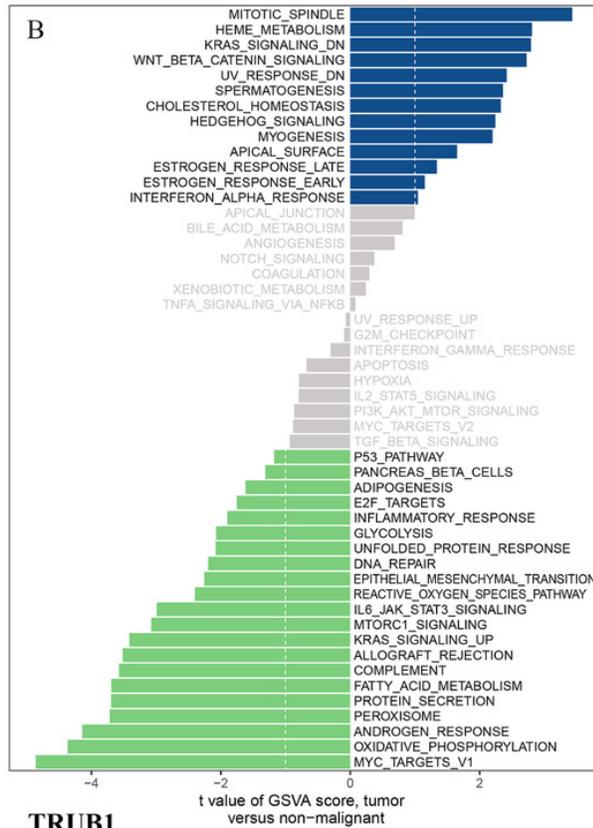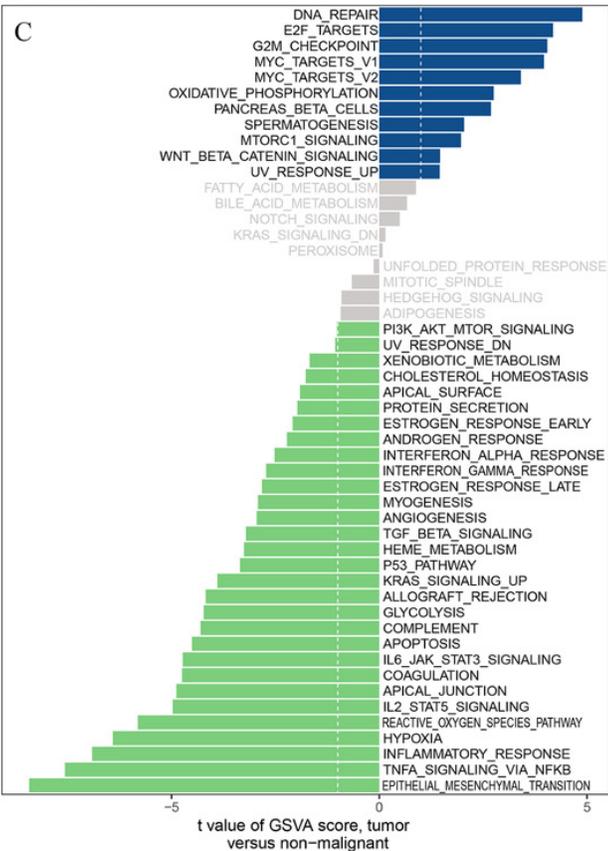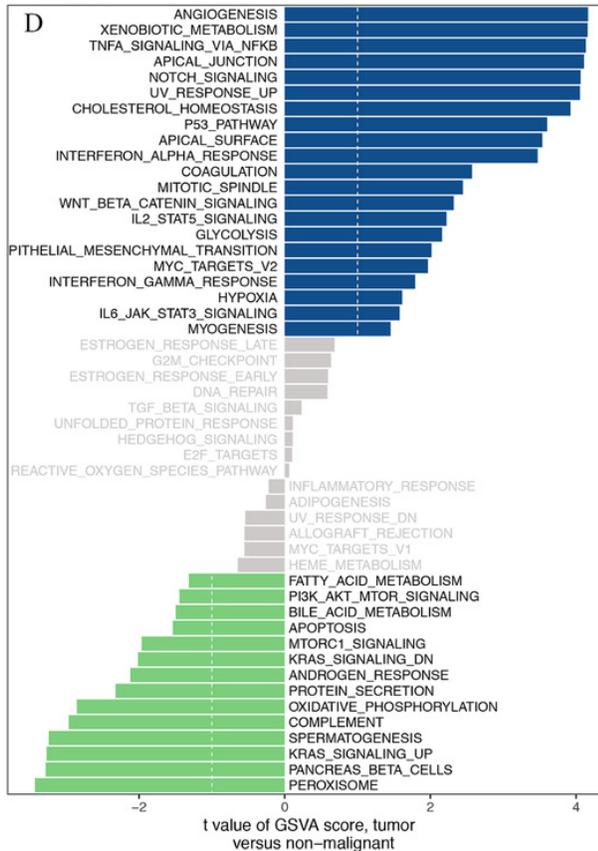
Table1: Statistics of genes in darkturquoise modules.

Table1 Statistics of genes in darkturquoise modules.

1   Table1 Statistics of genes in darkturquoise modules.

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