GRHL2 is a Candidate Prognostic and Immunotherapy Biomarker in Breast Cancer

Xiaoyu Bai
Tianjin Medical University
Yue Li
Tianjin Medical University
Yanlei Li
Tianjin Medical University
Fan Li
Tianjin Medical University
Na Che
Tianjin Medical University
Chunsheng Ni
Tianjin Medical University
Nan Zhao
Tianjin Medical University
Xiulan Zhao
Tianjin Medical University
Tieju Liu (mango616@163.com)
Tianjin Medical University

Research Article

Keywords: GRHL2, Breast cancer, Prognosis, immunotherapy, Tumor aggressiveness

Posted Date: October 25th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-999774/v1

Abstract

Background:Breast cancer (BC) is the most frequent malignancy in women worldwide and the leading cause of female cancer-associated death in the world. Thus, we need to identify new biomarkers to predict the prognosis of BC. In mammals, the Grainyhead-like (GRHL) family of transcription factors has three members, called Grainyhead-like 1 (GRHL1), Grainyhead-like 2 (GRHL2) and Grainyhead-like 3 (GRHL3). GRHL2 contributes to epithelial morphogenesis and differentiation.

Materials & Methods: In this study, we explored the expresstion of GRHL2 across multiple databases, including Oncomine, the Human Protein Atlas (HPA), UALCAN, and the Cancer Cell Line Encyclopedia (CCLE). PrognoScan, GEPIA and Kaplan-Meier plotter were utilized to analyze the prognostic value of GRHL2 in certain cancers. Based on TCGA data, a receiver operating characteristic (ROC) curve was used to evaluate the diagnostic performance of GRHL2 expression. Using the GEPIA and TIMER databases, we investigated the correlations between GRHL2 expression and immune cell infiltration in breast cancer. In addition, the gene set enrichment analysis was performed to unveil the potential molecular mechanism of GRHL2 promoting breast cancer aggressiveness by using LinkedOmics and METASCAPE.

Results: We demonstrated that tumor tissue had the higher expression level of GRHL2, compared with that in normal tissue. High expression of GRHL2 was associated with worse prognosis of breast cancer patients. ROC analysis indicated that GRHL2 had significant diagnostic value. Importantly, there were significant positive correlations between GRHL2 expression and immune infiltrates, including CD8+ T cells and macrophages in breast cancer. GRHL2 expression is regulated by methylation. Furthermore, KEGG and GO analysis showed GRHL2 related signaling pathways in breast cancer are related to tumor cells proliferation, invasion and angiogenesis.

Conclusion:In summary, we demonstrate that GRHL2 can be used as a prognostic and immunotherapy biomarker for breast cancer.

Introduction

Breast cancer is a global public health problem. It is currently the most common tumor in the world, and it is also the main cause of cancer deaths in women worldwide[1, 2]. In 2020, the number of new breast cancers is 2.26 million, surpassing lung cancer to become the most common cancer in the world. Due to the improved screening accuracy of mammography, screening mammography had led to a 19% overall reduction in breast cancer mortality [3]. However, there are still other ways to be found out to improve the survival rate and prognosis of breast cancer. At present, other research on the underlying pathogenesis and etiology of breast cancer need to be done and try to find ways that may lead to the discovery of advanced treatment methods and effective biomarkers for predicting prognosis of breast cancer.

The grainyhead-like (GRHL) transcription factors constitute a family whose first member, Grainyhead (GRH), was discovered in the fruit fly Drosophila melanogaster [4]. In mammals, the Grainyhead-like (GRHL) family of transcription factors has three members, called Grainyhead-like 1 (GRHL1), Grainyhead-

like 2 (GRHL2) and Grainyhead-like 3 (GRHL3). In some study, GRHL transcription factors were considered as tumor suppressors [5, 6]. However, in other condition, they show carcinogenic function. GRHL factors are involved in many biological processes, including tumor epithelial-mesenchymal transition (EMT), invasion and metastasis. Decreased expression of GRHL1 and GRHL3 genes increased the risk of skin cancer [7, 8]. GRHL2 is also a member of the GRHL family. The regulatory effect of GRHL2 in tumorigenesis and development is different in different types of cancer. For example, in breast cancer, overexpressed GRHL2 is reported to induce resistance to apoptosis by modulating death receptor ligands [9]. On the contrary, it is suggested that GRHL2 has a tumor suppressor effect in gastric cancer and colorectal cancer cells [10, 11]. However, the efficacy of GRHL2 as a potential cancer prognostic biomarker has not been fully elucidated.

The tumor microenvironment (TME) contains various cells. Among them, infiltrating immune cells including tumor-associated macrophages (TAM), B cells, CD8+ T cells, CD4+ T cells, neutrophils, natural killer (NK) cells and dendrites shaped cells (DC) account for a large proportion [12]. In recent years, immunotherapy targeting the interaction between immune cells and tumor cells has been developed into the clinical field, but only a limited number of cancer patients with certain molecular characters respond well to current immunotherapy [13]. In these biological processes, immune-related genes may influence the prognosis of cancer patients by affecting the abundance of infiltrating immune cells [14]. Therefore, exploring GRHL2 related immune cells could contribute to find new therapeutic targets.

In this study, we used multiple databases, including ONCOMINE, CCLC, TIMER and HPA, to report the expression of GRHL2 in a variety of cancers, and high levels of GRHL2 expression are related to tumor progression. Besides, we used PrognoScan, GEPIA, Kaplan-Meier plotter and ROC curve to visualize the prognostic and diagnostic value of GRHL2. Therefore, GRHL2 may prove to be a potential biomarker for the prognosis of breast cancer. Then, the relationship between GRHL2 expression and immune infiltration was explored through the TIMER and GEPIA databases. The findings of this study also analyzed the DNA methylation of GRHL2. In addition, the analysis of KEGG and GO showed the possible mechanism promoting breast cancer development mediated by GRHL2.

Results

GRHL2 mRNA expression in pan-cancer

To check the expression of GRHL2 in all cancer types, we analyzed the expression level of GRHL2 mRNA in Oncomine database. The results showed that the expression of GRHL2 was higher in bladder cancer, breast cancer, colorectal cancer, lung cancer and ovarain cancer tissues when compared with their corresponding GRHL2 mRNA expression in pan-cancer To check the expression of GRHL2 in all cancer types, we analyzed the expression level of GRHL2 mRNA in Oncomine database. The results showed that the expression of GRHL2 was higher in bladder cancer, breast cancer, colorectal cancer, lung cancer and ovarain cancer tissues when compared with their corresponding normal tissues (Figure 1A). We also performed a comprehensive analysis on 33 types of tumors from TCnormal tissues (Figure 1A). We also

performed a comprehensive analysis on 33 types of tumors from TCGA. Among them, 18 kinds of tumors overexpressed GRHL2 (Figure 1B). In addition, data from the Cancer Cell Line Encyclopedia (CCLE) database revealed that the high expression of GRHL2 mRNA was also detected in 28 kinds of cancer cell lines, especially in breast cancer cell lines (Figure 1C). Therefore, our results indicate that GRHL2 might play an important role in breast cancer.

Expression of GRHL2 in breast cancer

Further investigation by using HPA database, we found that GRHL2 was low expressed in normal breast tissues (Figure 2A) and over expressed in cancer tissues (Figure 2B). It was also confirmed from GEPIA database that GRHL2 was more expressed in cancer tissues (n = 1085) than in normal tissues (n = 291) (Figure 2C). Immunohistochemical staining obtained from HPA also confirmed GRHL2 protein expression was higher in tumor tissues than in normal tissues (Figure 2D).

Next, we further verified the correlation between GRHL2 mRNA levels and clinical data of breast cancer patients, including age, gender, and cancer stages. It can be seen that the expression of GRHL2 is not correlated with age, cancer stage and nodal metastasis status ($P > 0.05$), but significantly correlated with gender (Figure 3A-D) ($P < 0.05$).

The prognostic value of GRHL2

We used Kaplan-Meier plotter to assess the prognostic value of GRHL2. GRHL2 can predict the poorer overall survival (OS) of kidney renal clear cell carcinoma (KIRC) ($P < 0.05$), however, it could not predict relapsing free survival (RFS) ($P = 0.05$) (Figure 4A, B). For pancreatic ductal adenocarcinoma (PDA), GRHL2 has predictive effect on OS and RFS (Figure 4C, D) ($P < 0.05$). In a total of 1643 and 1089 BC patients, the higher GRHL2 was associated with poorer OS and RFS ($P < 0.05$) (Figure 4E, F).

In order to further verify the prognostic role of GRHL2, PrognoScan and GEPIA database were used. The data in PrognoScan mainly comes from the GEO database. Overexpression of GRHL2 in three breast cancer data sets and one bladder cancer data set is associated with poorer survival (DMFS, distant metastasis-free survival, and OS) (Figure 5A-D). GEPIA database also showed GRHL2 high expression was related to poorer OS in breast cancer (Figure 5E).

GRHL2 expression is a diagnostic biomarker for breast cancer

In order to evaluate the diagnostic value of GRHL2, the ROC curve was generated from the data of the TCGA database. The area under the ROC curve is 0.818, indicating a higher diagnostic value of GRHL2 for breast cancer (Figure 6).

Correlation between GRHL2 expression and immune cells infiltration in breast cancer

In order to evaluate the correlation between GRHL2 expression and immune cells infiltration in breast cancer, we used the TIMER database for analysis. GRHL2 expression level is significantly correlated with tumor purity, positively correlated with CD8+ cells, macrophages, and neutrophils infiltration, negatively correlated with DC infiltration, and has no significant correlation with B cells and CD4+ cells (Figure 7A). We further evaluated the relationship of several immune cell infiltration levels with GRHL2 gene copy number, and found that CD4+ cells and macrophages was related to GRHL2 gene copy number in breast cancer (Figure 7B).

Relationships between GRHL2 expression and immune markers

In order to further explore the potential relationship between GRHL2 and immune markers, we used TIMER and GEPIA to observe B cells, CD8+ T cells, M1/M2 macrophages, tumor-associated macrophages, monocytes, NK cells, neutrophils and DC markers in breast cancer. And we also analyzed different functional T cells, including Tfh, Th1, Th2, Th9, Th17, Th22, Treg and T cell exhaustion (Table 1 and Figure 8). In TIMER, after adjusting tumor purity, GRHL2 expression level was significantly correlated with 22 of the 45 immune cell markers in breast cancer (Table 1).

As shown in Figure 7A, CD8+ T cells and macrophages in breast cancer have the closer relationship with GRHL2 expression. Therefore, we further analyzed the immune cell markers in GEPIA (Table 2). Interestingly, B cell marker CD19, CD38 and MS4A1 were negatively related with GRHL2 in breast cancer, not in normal tissue (Table 2). These results indicate that the different immune cells related to GRHL2 might involve in breast cancer aggressiveness under different microenvironment.

Function enrichment analysis

To clarify the genes and signal transduction pathways related with GRHL2, we performed KEGG and GO analysis. We first used the LinkedOmic database to analyze the upstream and downstream genes co-expressed with GRHL2 in the volcano map (Figure 9A-C). KEGG and GO analysis identified 3 main group related to tumor aggressiveness (Figure 9D-E). The first group included lymphocyte activation and Th1, Th2 and Th17 cell differentiation. This further verified the analysis results of TIMER and GEPIA, which demonstrated that GRHL2 could regulate immune cells infiltration in tumor tissue. The second group included establishment or maintenance of cell polarity, regulation of actin filament length and polymerization, actin filament polymerization or depolymerization. This was consistent with previous research[15], which demonstrated GRHL2 could regulate EMT. Our result suggested that GRHL2 might regulate actin filament status to determine EMT phenotype of tumor cells. The third group included cell

cycle, DNA replication, nuclear division, mismatch repair, nucleotide excision repair, double-strand break repair, cell adhesion molecules, NF-kappa β signaling pathway, PI3K-Akt signaling pathway and positive regulation of angiogenesis. This suggested GRHL2 could involve in cell cycle control and have an effect on tumor cell proliferation. In addition, GRHL2 might promote tumor invasiveness by cooperation with NF-kappa β signaling pathway and PI3K-Akt signaling, affecting cell adhesion molecules expression and regulating angiogenesis.

Methylation could regulate GRHL2 expression

In order to further elucidate the mechanism of regulating GRHL2 expression in breast cancer, we explored the correlation between GRHL2 expression level and methylation. First, the analysis results of GRHL2 from the UALCAN database showed that promoter methylation level in normal tissues is higher than that in cancer tissues (Figure 10A). The analysis results in DiseaseMeth version 2.0 are the same as those in UALCAN (Figure 10B). In addition, we analyzed the relationship between GRHL2 mRNA expression and methylation level through the Cbioportal database, which was negatively correlated (Figure 10C). Then, the results of MEXPRESS analysis showed that in the DNA methylation sequences of GRHL2, there are 25 methylation sites that are negatively correlated with its expression level (Figure 10D). One of the probes, cg15679829, is related to promoter methylation of GRHL2 in MethSurv. And we analyzed this methylation site with survival in this database, which showed no significance (Figure 10E). However, the density and the methylation level of GRHL2 were different in different age groups of BC (Figure 10F-G). It can be seen from the density graph that the β-value is 0.844, which is significant (β-value>0.6). These results demonstrate that the promoter methylation of GRHL2 could regulate GRHL2 expression.

Discussion

Breast cancer is a very common female disease. Although early detection and treatment have reduced the mortality rate of breast cancer, patients with metastases will have a poor prognosis [16]. In recent years, breakthroughs in the field of immunotherapy have provided new approaches for breast cancer treatment. Therefore, exploring new biomarkers for predicting breast cancer recurrence and metastasis, survival outcome and immunotherapy response are valuable for breast cancer patients.

In mammals, the structure and regeneration of various epithelial cells depend on the three members of the GRHL family of transcription factors, GRHL1, GRHL2, and GRHL3. From a recent review, it was found that all GRHLs are associated with various types of cancer [17]. GRHL2 has been shown to be a key determinant of keratinocyte differentiation and lung epithelial morphogenesis, and is considered to be a lineage determinant of breast cancer epithelial cells [18]. However, its prognostic effects in other aspects have not been fully studied. New evidence shows that GRHL2 is a novel oncogene [19], but on the contrary, it has a tumor suppressor effect in gastric cancer, cervical cancer, clear cell renal cell carcinoma, and sarcoma [20, 21]. Therefore, GRHL2 has different regulatory effects in different cancers, and it has not been studied in depth in breast cancer.

In our study, the Oncomine and TIMER databases were used to assess the correlation between GRHL2 expression and the prognosis of 33 different types of cancer, indicating that there are significant differences between normal tissues and cancer tissues. In Oncomine, we found that GRHL2 was highly expressed in bladder cancer, breast cancer, colorectal cancer, lung cancer and ovarian cancer compared with the expression level in normal tissues. Meantime, in the TIMER database, GRHL2 expression is higher in bladder urothelial carcinoma, breast invasive carcinoma, cervical squamous cell carcinoma and endocervical adenocarcinoma, cholangio carcinoma, esophageal carcinoma, head and neck squamous cell carcinoma, lung adenocarcinomaand squamous cell carcinoma, pancreatic adenocarcinoma, prostate adenocarcinoma, rectum, colon and stomach adenocarcinoma, and uterine corpus endometrial carcinoma. In different databases, these different GRHL2 expression levels in cancer are due to different data collection methods and biological potential analysis methods. Interestingly, the results obtained for breast cancer through these two databases are consistent. The expression of GRHL2 is high in breast cancer tissues and low in normal tissues. Therefore, we have reason to speculate that GRHL2 may be involved in the occurrence and development of breast cancer.

Then, we analyzed GRHL2 expression levels through HPA, GEPIA and UALCAN databases and conducted research on different ages, genders, and pathological data. Through HPA database and immunohistochemical staining, it can be found that GRHL2 protein expression is consistent with mRNA expression and is also highly expressed in breast cancer tissues. Therefore, we infer that the high expression of GRHL2 may play critical role in breast cancer occurrence and development as a carcinogenic factor. Then we used Kaplan-Meier plotter, PrognoScan and GEPIA and found that high expression of GRHL2 could induce shorter survival time of breast cancer patients. Therefore the high expression level of GRHL2 can be used as an independent risk factor for poor prognosis of breast cancer. We further investigated whether GRHL2 can be used as a diagnostic marker for breast cancer. ROC curve shows that the expression of GRHL2 has high diagnostic value in breast cancer.

In our study, the expression of GRHL2 is significantly positively correlated with tumor purity in breast cancer tissue, indicating that its expression is different in tumor cells and tumor microenvironment (TME), demonstrating that it is relatively enriched in tumor cells. TME refers to the cellular environment in which tumor or cancer stem cells exist. TME includes surrounding immune cells, blood vessels, extracellular matrix, fibroblasts, bone marrow-derived inflammatory cells and signal molecules [22, 23]. Previous studies have reported that immune infiltration in TME can affect immune therapy responses and prognosis of patients [24-26]. In our research, we found that GRHL2 is associated with multiple types of immune cells infiltration in breast cancer. TIMER analyses indicate that GRHL2 gene copy number is related to CD4+ T cell and macrophages, and GRHL2 expression level is significantly related to CD8+ T cell and macrophages. KEGG and GO analyses also show that GRHL2 and its related genes involve in lymphocyte activation and T helper cells differentiation, demonstrating GRHL2 expression of tumor cells is associated with immune cell infiltration in TME. A common type of T lymphocytes in TME is CD8+ T cells, which can kill tumor cells by their immune killing effect. However, tumours progress despite the presence of CD8+ T cells in TME, which suggests CD8+ T cell differentiation to dysfunctional states fail to achieve responses to immunotherapy [27]. Therefore, our results indicate that GRHL2 expression

might function in breast cancer aggressiveness and resist immunotherapy by inducing the dysfuntion of CD8+ T cells. After adjusting tumor purity, GRHL2 expression is positively correlated with M2 macrophages. Macrophages are the most prominent immune cell type of TME [28, 29]. Macrophages in TME can promote tumor reoccurrence and metastases. They can facilitate the escape of tumor cells into the circulatory system, and can inhibit the anti-tumor immune mechanism and response [29]. CD68 is a pan-macrophage marker that points to activated M1 and M2 in TAM, while CD163 only refers to the antigen associated with M2 macrophages [30]. In our study, GRHL2 has a significant correlation with CD163, which suggests that GRHL2 expressing tumor cells may recruit M2 macrophages into tumor tissue to regulate breast cancer development. In addition, our results also indicate that GRHL2 may be related to Treg gene markers. Tregs highly enriched in the tumor microenvironment are widely known for their immunosuppressive effects in tumors [31]. We also studied the relationship between GRHL expression and TAM molecules (CCL2, CD68 and IL10). The results show that GRHL2 expression is negatively correlated with CCL2. Recent research [32] supported a favourable prognostic value of tumour-infiltrating CD20+ B lymphocytes in colorectal cancer. In our study, B cell marker CD19, CD38 and MS4A1 were negatively related with GRHL2 in breast cancer, not in normal tissue. This phenomenon not only brings important clues to the prognosis of breast cancer, but also helps to explore new therapeutic targets.

In order to further explore the biological functions of GRHL2, we performed KEGG and GO analysis on GRHL2. The enrichment analysis showed that GRHL2 and its related factors involve in mutiple tumor related signaling pathways, which may be related to breast cancer cell proliferation, invasion and metastasis.

According to research, DNA methylation plays an important role in regulating gene expression [33-36]. In mammals, DNA methylation occurs at the CPG site[37]. DNA methylation is achieved by linking the methyl group to the C5 position of cytosine (5mC) through CpG context. Hypermethylation of CpG islands located in the promoter region of tumor related genes is considered to be one of the earliest, most frequent and powerful changes in cancer development [38, 39]. Therefore, we use methylation databases to explore the effect of DNA methylation on GRHL2 expression that is abnormally expressed in breast cancer. We found that GRHL2 methylation levels are lower in breast cancer tissues compared with normal tissues. In addition, we found that some methylation sites are negatively correlated with the prognosis of breast cancer patients. By analyzing the association between GRHL2 and genome-wide methylation in MEXPRESS, the results show that more methylation sites are closer to the open sea, suggesting GRHL2 expression could be regulated by methylation. In our study, a visual analysis of cg15679829 was performed through the MethSurv database. We found that one of the cg15679829 probe sites had higher methylation in S-shelf. These results demonstrate that the promoter of GRHL2 can be methylated and this abnormal methylation may be related to the development of breast cancer.

Conclusions

In summary, GRHL2 is overexpressed in breast cancer and might promote tumor invasion and metastasis. GRHL2 is significantly related to poor prognosis and immune cells infiltration. Our study demonstrates an important role of GRHL2 in the regulation of breast cancer immune cells infiltration and thus will have a significant impact on breast cancer immune therapy responses. In addition, GRHL2 expression is regulated by DNA methylation. Therefore, our results indicate that GRHL2 is a valuable biomarker for breast cancer prognosis and provide new insight for immune therapy.

Materials And Methods

Oncomine database

Oncomine database (www.oncomine.org)[40] can be used for clinical practice and drug development. In order to understand the expression of GRHL2 in cancer, the oncomine database was used.

TIMER database

TIMER (http://cistrome.org/TIMER/)[41] is a website used to evaluate various types of cancers and their clinical effects on immune cells. A new algorithm is used to estimate the abundance of six types of cells in the TME, including B cell, CD4+ T cell, CD8+ T cell, neutrphil, macrophage and dendritic cell. The current version of TIMER contains 10,009 samples from 23 cancers in the TCGA database. We used TIMER to analyze the expression of GRHL2 in various cancers and analyzed the expression of GRHL2 with the abundance of six immune infiltrating cells. In addition, a scatter plot was used to analyze the relationship between GRHL2 and gene marker.

Cancer Cell Line Encyclopedia (CCLE) database

CCLC (https://portals.broadinstitute.org/ccle/about)[42] is a compilation of gene expression, chromosome copy number and massively parallel sequencing data from 947 human cancer cell lines. We used the CCLC database to confirm the expression of GRHL2 in breast cancer.

GEPIA database

GEPIA (http://gepia.cancer-pku.cn/)[43] is a Web tool that provides fast and customizable functions based on TCGA and GTEx data. GEPIA provides key interactive and customizable functions. Just click GEPIA to perform comprehensive expression analysis, including differential expression analysis, correlation analysis, patient survival rate analysis, etc. We explored the expression of GRHL2 in cancer and normal tissues (total = 1376) and its effect on OS (total = 1069).

The human protein altas (HPA) database

The human protein altas (HPA https://www.proteinatlas.org/) database represents protein expression in 44 major human tissues and certain cancer tissues through immunohistochemical methods[44]. GRHL2 antibodies used in the database is HPA004820, which has detected its expression in human normal tissues and cancer tissues, and verified the previous research by immunohistochemical staining.

UALCAN database

UALCAN (http://ualcan.path.uab.edu/) [45] uses TCGA level 3 RNA-seq and clinical data from 31 cancer types. The UALCAN database was used to analyze the relative expression of GRHL2 in each cancer stage, tumor grade, age, and gender. In addition, we also evaluated the methylation of GRHL2 expression between breast cancer and normal adjacent tissues.

Kaplan-Meier plotter database

Kaplan-Meier Plotter (https://kmplot.com/analysis/)[46] is an online database that can assess the association of genes in four types of cancer samples. The correlation between GRHL2 mRNA expression and survival in breast cancer and pan-cancer can be analyzed online, and expressed by hazard ratio, 95% confidence interval and calculated log-rank p value.

PrognoScan database

PrognoScan (http://dna00.bio.kyutech.ac.jp/PrognoScan/index.html)[47] database has a large number of publicly available cancer microarray data sets with clinical annotations and tools to assess the biological relationship between gene expression and prognosis . PrognoScan uses the smallest P value method to group patients for survival analysis. We used the PrognoScan database to determine the correlation between GRHL2 mRNA expression and survival in breast cancer, and its cox p value was adjusted to <0.05.

LinkedOmics database

The LinkedOmics (http://www.linkedomics.org/)[48] database contains multi-omics data and clinical data from 32 cancers and 11 158 patients from the Cancer Genome Atlas (TCGA) project. It integrates multiple sets of databases of global proteomics data, from which we analyzed the volcano map of the upstream and downstream genes of GRHL2 and the gene correlation heat map. On this basis, perform KEGG and GO analysis, and download related data sets.

Metascape database

Metascape (http://metascape.org/)[49] is a web-based database that can be used to gain insight into the biological functions of these co-expressed genes. Therefore, we downloaded the GRHL2 co-expressed genes using the LinkedOmics database, sorted out the top 50 genes, and performed go analysis using the Metascape database.

Cbioportal database

The cBioPortal online database (http://www.cbioportal.org/)[50] is a multi-dimensional cancer genomics data set that can query the frequency of core gene changes and interactively explore more than 5,000 tumor samples in 20 cancer studies. Through this database, we visualized the copy number and DNA methylation of GRHL2.

DiseaseMeth database

The number of samples in the DiseaseMeth database is 32,701, the number of diseases is 88, and the relationship between diseases and genes is 679602. DiseaseMeth version 2.0 (http://bio-bigdata.hrbmu.edu.cn/diseasemeth/)[51] provides search engines and visualization tools, enhanced difference analysis tools, and can now automatically identify humans online through case-control or disease-disease methods abnormal DNA methylation in disease. We use the Human Disease Methylation Database DiseaseMeth Version 2.0 to evaluate the expression of GRHL2 between breast cancer and normal tissues.

MEXPRESS database

MEXPRESS (https://mexpress.be)[52] is a database that visualizes the relationship between patient clinical information and methylation expression in the TCGA database. You can see the change information of the methylation site of genes from the analysis map, such as the genome The length, the different transcripts, the location of the cg site, and the location of the cpG island. We used this database to study the relationship between GRHL2 gene expression and its DNA methylation status.

MethSurv database

MethSurv (https://biit.cs.ut.ee/methsurv/)[53] is a network tool for survival analysis based on CpG methylation patterns. It uses 7358 nails from 25 different human cancers from "TCGA". Based on the data, the Cox proportional hazard model was used to develop an interactive network tool for survival analysis. The GRHL2 gene in the TCGA HM450K cancer methylation data set was used to perform a multivariate survival analysis of a single CpG site to evaluate the scattering of different CpG islands; the density map was used to visualize and highlight the critical point of the dichotomy of the patient's methylation level. All cutoff points evaluated in MethSurv; violin chart to visually query the methylation

distribution of CpG sites, the median and interquartile range, and the methylation level between patient characteristics.

Statistical analysis

Use SPSS25.0 to perform statistical analysis on the obtained data. A receiver operating characteristic (ROC) curve is generated to evaluate the diagnostic value expressed by GRHL2, and the area under the curve represents the diagnostic value. P <0.05 was considered statistically significant.

Abbreviations

Breast cancer (BC)

Grainyhead-like (GRHL)

Grainyhead-like 1 (GRHL1)

Grainyhead-like 2 (GRHL2)

Grainyhead-like 3 (GRHL3)

Human Protein Atlas (HPA)

Cancer Cell Line Encyclopedia

Receiver operating characteristic (ROC)

Epithelial-mesenchymal transition (EMT)

The tumor microenvironment (TME)

Tumor-associated macrophages

Natural killer (NK)

Dendrites shaped cells (DC)

Overall survival (OS)

Relapsing free survival (RFS)

Pancreatic ductal adenocarcinoma (PDA)

Distant metastasis-free survival(DMFS)

Declarations

Funding

This study was supported by the project of National Nature Science Foundation of China (No. 81572872).

Availability of data and materials

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding authors.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

The authors consent for publication

Authors' contributions

XB and YaL contributed to conception and design of the study. XB, YuL and FL organized the database. XB and CN performed the statistical analysis. XB wrote the first draft of the manuscript. XB , CN and NZ summarized resources. XB and TL reviewed and edited the writing. XZ acquired the funding.All authors contributed to manuscript revision, read, and approved the submitted version.

References

1. Huang Y, Tong Z, Chen K, Wang Y, Liu P, Gu L, et al. Interpretation of breast cancer screening guideline for Chinese women. CANCER BIOL MED (2019) 16: 825-835. 10.20892/j.issn.2095-3941.2019.0322
2. Cao LandNiu Y. Triple negative breast cancer: special histological types and emerging therapeutic methods. CANCER BIOL MED(2020)17: 293-306. 10.20892/j.issn.2095-3941.2019.0465
3. Pace LE and KeatingNL.A systematic assessment of benefits and risks to guide breast cancer screening decisions. JAMA(2014)311(13):1327-35.10.1001/jama.2014.1398
4. Nusslein-VolhardC, WieschausEand KludingH.Mutations affecting the pattern of the larval cuticle inDrosophila melanogaster : I. Zygotic loci on the second chromosome. Wilehm Roux Arch Dev

Biol(1984)193(5): 267-282.10.1007/BF00848156

5. Frisch SM, FarrisJC and PiferPM. Roles of Grainyhead-like transcription factors in cancer. Oncogene(2017)36(44): 6067-6073.10.1038/onc.2017.178
6. Mlacki M, Kikulska A, Krzywinska E, Pawlak M, Wilanowski T.Recent discoveries concerning the involvement of transcription factors from the Grainyhead-like family in cancer. Exp Biol Med (Maywood)(2015)240(11): 1396-401.10.1177/1535370215588924
7. Mlacki M, Darido C, Jane SM, Wilanowski T.Loss of Grainy head-like 1 is associated with disruption of the epidermal barrier and squamous cell carcinoma of the skin. PLoS One(2014)9(2): e89247.10.1371/journal.pone.0089247
8. Darido C, Georgy SR, Wilanowski T, Dworkin S, Auden A, Zhao Q,et al. Targeting of the tumor suppressor GRHL3 by a miR-21-dependent proto-oncogenic network results in PTEN loss and tumorigenesis. Cancer Cell(2011)20(5): 635-48.10.1016/j.ccr.2011.10.014
9. Dompe N, Rivers CS, Li L, Cordes S, Schwickart M, Punnoose EA,et al. A whole-genome RNAi screen identifies an 8q22 gene cluster that inhibits death receptor-mediated apoptosis. Proc Natl Acad Sci U S A(2011)108(43): E943-51.10.1073/pnas.1100132108
10. Xiang J, Fu X, Ran W, Chen X, Hang Z, Mao H,et al. Expression and role of grainyhead-like 2 in gastric cancer. Med Oncol(2013)30(4): 714.10.1007/s12032-013-0714-5
11. Quan Y, Jin R, Huang A, Zhao H, Feng B, Zang L,et al. Downregulation of GRHL2 inhibits the proliferation of colorectal cancer cells by targeting ZEB1. Cancer Biol Ther(2014)15(7): 878-87.10.4161/cbt.28877
12. Bindea G, Mlecnik B, Tosolini M, Kirilovsky A, Waldner M, Obenauf AC,et al. Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. Immunity(2013)39(4): 782-95.10.1016/j.immuni.2013.10.003
13. Topalian SL, Drake CG, Pardoll DM.Immune checkpoint blockade: a common denominator approach to cancer therapy. Cancer Cell(2015)27(4): 450-61.10.1016/j.ccell.2015.03.001
14. Lin P, Guo YN, Shi L, Li XJ, Yang H, He Y,et al. Development of a prognostic index based on an immunogenomic landscape analysis of papillary thyroid cancer. Aging (Albany NY)(2019)11(2): 480-500.10.18632/aging.101754
15. He J, Feng C, Zhu H, Wu S, Jin P, Xu T. Grainyhead-like 2 as a double-edged sword in development and cancer. AM J TRANSL RES(2020)12: 310-331.
16. Emens LA.Breast Cancer Immunotherapy: Facts and Hopes. Clin Cancer Res(2018)24(3): 511-520.0.1158/1078-0432.CCR-16-3001
17. Mlacki M, Kikulska A, Krzywinska E, Pawlak M, Wilanowski T.Recent discoveries concerning the involvement of transcription factors from the Grainyhead-like family in cancer. Exp Biol Med (Maywood)(2015)240(11): 1396-401.10.1177/1535370215588924
18. Xiang X, Deng Z, Zhuang X, Ju S, Mu J, Jiang H,et al. Grhl2 determines the epithelial phenotype of breast cancers and promotes tumor progression. PLoS One(2012)7(12): e50781.

19. Cieply B, Farris J, Denvir J, Ford HL, Frisch SM.Epithelial-mesenchymal transition and tumor suppression are controlled by a reciprocal feedback loop between ZEB1 and Grainyhead-like-2. Cancer Res(2013)73(20): 6299-309.10.1158/0008-5472.CAN-12-4082
20. Xiang J, Fu X, Ran W, Chen X, Hang Z, Mao H,et al. Expression and role of grainyhead-like 2 in gastric cancer. Med Oncol(2013)30(4): 714.10.1007/s12032-013-0714-5
21. Torres-Reyes LA, Alvarado-Ruiz L, Pina-Sanchez P, Martinez-Silva MG, Ramos-Solano M, Olimon-Andalon V,et al. Expression of transcription factor grainyhead-like 2 is diminished in cervical cancer. Int J Clin Exp Pathol(2014)7(11): 7409-18.
22. Spill F, Reynolds DS, Kamm RD, Zaman MH.Impact of the physical microenvironment on tumor progression and metastasis. Curr Opin Biotechnol(2016)40: 41-48.10.1016/j.copbio.2016.02.007
23. Del PA, Schioppa T, Tiberio L, Stabile H, Sozzani S.Leukocyte trafficking in tumor microenvironment. Curr Opin Pharmacol(2017)35: 40-47.10.1016/j.coph.2017.05.004
24. Jin Y, Chen DL, Wang F, Yang CP, Chen XX, You JQ,et al. The predicting role of circulating tumor DNA landscape in gastric cancer patients treated with immune checkpoint inhibitors. Mol Cancer(2020)19(1): 154.10.1186/s12943-020-01274-7
25. Yang CY, Fan MH, Miao CH, Liao YJ, Yuan RH, Liu CL.Engineering Chimeric Antigen Receptor T Cells against Immune Checkpoint Inhibitors PD-1/PD-L1 for Treating Pancreatic Cancer. Mol Ther Oncolytics(2020)17: 571-585.10.1016/j.omto.2020.05.009
26. Zhang L, Wang W, Wang R, Zhang N, Shang H, Bi Y,et al. Reshaping the Immune Microenvironment by Oncolytic Herpes Simplex Virus in Murine Pancreatic Ductal Adenocarcinoma. Mol Ther(2021)29(2): 744-761.10.1016/j.ymthe.2020.10.027
27. Hanahan DandWeinberg RA.Hallmarks of cancer: the next generation. Cell(2011)144(5): 646-74.10.1016/j.cell.2011.02.013
28. Mantovani A, Allavena P, Sica A, Balkwill F.Cancer-related inflammation. Nature(2008)454(7203): 436-44.10.1038/nature07205
29. Mary Philip, Andrea Schietinger. CD8 + T cell differentiation and dysfunction in cancer. Nat Rev Immunol(2021)12.10.1038/s41577-021-00574-3.
30. Jeong H, Hwang I, Kang SH, Shin HC, Kwon SY.Tumor-Associated Macrophages as Potential Prognostic Biomarkers of Invasive Breast Cancer. J Breast Cancer(2019)22(1): 38-51.10.4048/jbc.2019.22.e5
31. Josefowicz SZ, Lu LFandRudensky AY.Regulatory T cells: mechanisms of differentiation and function. Annu Rev Immunol(2012)30: 531-64.10.1146/annurev.immunol.25.022106.141623
32. Edin S, Kaprio T, Hagstrom J, Larsson P, Mustonen H, Bockelman C,et al.The Prognostic Importance of CD20(+) B lymphocytes in Colorectal Cancer and the Relation to Other Immune Cell subsets. Sci Rep(2019)9: 19997. 10.1038/s41598-019-56441-8
33. Chen Y, Ren B, Yang J, Wang H, Yang G, Xu R,et al. The role of histone methylation in the development of digestive cancers: a potential direction for cancer management. Signal Transduct Target Ther(2020)5(1): 143.10.1038/s41392-020-00252-1

34. Khalaj-Kondori M, Hosseinnejad M, Hosseinzadeh A, Behroz SS, Hashemzadeh S.Aberrant hypermethylation of OGDHL gene promoter in sporadic colorectal cancer. Curr Probl Cancer(2020)44(1): 100471.10.1016/j.currproblcancer.2019.03.001
35. Yan JB, Lai CC, Jhu JW, Gongol B, Marin TL, Lin SC,et al. Insulin and Metformin Control Cell Proliferation by Regulating TDG-Mediated DNA Demethylation in Liver and Breast Cancer Cells. Mol Ther Oncolytics(2020)18: 282-294.10.1016/j.omto.2020.06.010
36. Shu WJandDu HN.The methyltransferase SETD3-mediated histidine methylation: Biological functions and potential implications in cancers. Biochim Biophys Acta Rev Cancer(2021)1875(1): 188465.10.1016/j.bbcan.2020.188465
37. Li Y, Li F, Bai X, Li Y, Ni C, Zhao X,et al.ITGA3 Is Associated With Immune Cell Infiltration and Serves as a Favorable Prognostic Biomarker for Breast Cancer. FRONT ONCOL(2021)11: 658547. 10.3389/fonc.2021.658547
38. Baylin SB.DNA methylation and gene silencing in cancer. Nat Clin Pract Oncol(2005)2 Suppl 1: S4-11.10.1038/ncponc0354
39. Wittenberger T, Sleigh S, Reisel D, Zikan M, Wahl B, Alunni-Fabbroni M,et al. DNA methylation markers for early detection of women's cancer: promise and challenges. Epigenomics(2014)6(3): 311-27.10.2217/epi.14.20
40. Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, Ghosh D,et al. ONCOMINE: a cancer microarray database and integrated data-mining platform. Neoplasia(2004)6(1): 1-6.10.1016/s1476-5586(04)80047-2
41. Li B, Severson E, Pignon JC, Zhao H, Li T, Novak J,et al. Comprehensive analyses of tumor immunity: implications for cancer immunotherapy. Genome Biol(2016)17(1): 174.10.1186/s13059-016-1028-7
42. Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S,et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature(2012)483(7391): 603-7.10.1038/nature11003
43. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z.GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. Nucleic Acids Res(2017)45(W1): W98-W102.10.1093/nar/gkx247
44. Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A,et al. Proteomics. Tissue-based map of the human proteome. Science(2015)347(6220): 1260419.10.1126/science.1260419
45. Chandrashekar DS, Bashel B, Balasubramanya S, Creighton CJ, Ponce-Rodriguez I, Chakravarthi B,et al. UALCAN: A Portal for Facilitating Tumor Subgroup Gene Expression and Survival Analyses. Neoplasia(2017)19(8): 649-658.10.1016/j.neo.2017.05.002
46. Gyorffy B, Lanczky A, Eklund AC, Denkert C, Budczies J, Li Q,et al. An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. Breast Cancer Res Treat(2010)123(3): 725-31.10.1007/s10549-009-0674-9
47. Mizuno H, Kitada K, Nakai K, Sarai A.PrognoScan: a new database for meta-analysis of the prognostic value of genes. BMC Med Genomics(2009)2: 18.10.1186/1755-8794-2-18

48. Vasaikar SV, Straub P, Wang J, Zhang B.LinkedOmics: analyzing multi-omics data within and across 32 cancer types. Nucleic Acids Res(2018)46(D1): D956-D963.10.1093/nar/gkx1090
49. Yadav RandSrivastava P.Clustering, Pathway Enrichment, and Protein-Protein Interaction Analysis of Gene Expression in Neurodevelopmental Disorders. Adv Pharmacol Sci(2018)2018: 3632159.10.1155/2018/3632159
50. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA,et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov(2012)2(5): 401-4.10.1158/2159-8290.CD-12-0095
51. Xiong Y, Wei Y, Gu Y, Zhang S, Lyu J, Zhang B,et al. DiseaseMeth version 2.0: a major expansion and update of the human disease methylation database. Nucleic Acids Res(2017)45(D1): D888-D895.10.1093/nar/gkw1123
52. Koch A, De Meyer T, Jeschke J, Van Criekinge W.MEXPRESS: visualizing expression, DNA methylation and clinical TCGA data. BMC Genomics(2015)16: 636.10.1186/s12864-015-1847-z
53. Modhukur V, Iljasenko T, Metsalu T, Lokk K, Laisk-Podar T, Vilo J. MethSurv: a web tool to perform multivariable survival analysis using DNA methylation data. Epigenomics (2018) 10(3): 277-288. 10.2217/epi-2017-0118

Tables

Table 1

Correlations between GRHL2 and Gene Markers of Immune Cells in TIMER

Cell type	Gene marker	Breast cancer			
		None		Purity	
		Cor	*P*	Cor	*P*
B cell	CD19	-0.159	***	-0.02	0.484
	CD38	-0.048	0.11	0.088	*
	MS4A1	-0.109	**	0.057	0.0729
CD8+ T cell	CD8A	-0.126	0.684	0.04	0.211
	CD8B	-0.2	***	-0.054	0.0901
Tfh	CXCR5	-0.14	***	0.015	0.629
	ICOS	-0.045	0.133	0.103	**
Th1	IL12RB2	-0.002	0.947	0.081	*
	TBX21	-0.2	***	-0.059	0.0979
Th2	CCR3	-0.056	0.0631	0.014	0.653
	STAT6	0.111	**	0.158	***
	GATA3	0.292	***	0.024	***
Th9	TGFBR2	-0.029	0.336	0.145	***
	IRF4	-0.057	0.0605	0.122	**
	SPI1	-0.331	***	-0.209	***
TH17	IL21R	-0.116	**	0.034	0.286
	IL23R	0.015	0.631	0.105	**
	STAT3	0.312	***	0.358	***
Th22	CCR10	-0.253	***	-0.205	***
	AHR	0.163	***	0.244	***
Treg	FOXP3	-0.033	0.267	0.116	**
	CCR8	0.14	**	0.252	***
T cell exhaustion	PDCD1	-0.234	***	-0.107	**
	CTLA4	-0.135	***	0.002	0.954
Macrophage	CD68	-0.079	*	0.029	0.359
	ITGAM	-0.07	0.02	0.028	0.370

M1	NOS2	-0.017	0.569	0.003	0.925
	ROS1	0.022	0.466	0.047	0.140
M2	ARG1	0.036	0.236	0.087	**
	MRC1	-0.059	0.0486	0.088	**
TAM	HLA-G	-0.192	***	-0.144	***
	CD80	0.069	0.022	0.156	***
	CD86	-0.085	**	0.036	0.254
Monocyte	CD14	-0.307	***	-0.234	***
	FCGR3A	0.042	0.168	0.13	***
NK	XCL1	-0.144	***	0.002	0.939
	KIR3DL1	-0.099	*	-0.017	0.588
	CD7	-0.312	***	-0.197	***
Neutrophil	FUT4	-0.086	*	0.036	0.254
	MPO	-0.11	**	-0.009	0.774
DC	CDIC	-0.21	***	-0.075	0.0184
	THBD	-0.109	**	-0.026	0.410

Tfh, follicular helper T cell; Th, T helper cell; Treg, regulatory T cell; TAM, tumor-associated-macrophage; NK, natural killer cell; DC, dendritic cell; None, correlation without adjustment; Purity, correlation adjusted for tumor purity; Cor, R value of Spearman's correlation.

*P < 0.01;

**P < 0.001;

****P < 0.0001*.

Table 2

Correlations between GRHL2 and genes markers of $CD8^+$ T cells, B cells, macrophages, and monocytes in GEPIA.

Cell type	Gene marker	Breast cancer			
		Tumor		Normal	
		R	*P*	R	*P*
CD8+ T cell	CD8A	-0.17	***	0.51	***
	CD8B	-0.18	***	0.53	***
B cell	CD19	-0.17	***	0.029	0.76
	CD38	-0.01	**	-0.081	0.39
	MS4A1	-0.11	**	0.032	0.74
Monocyte	CD14	-0.15	***	-0.32	**
	FCGR3A	0.021	0.5	-0.017	0.86
TAM	CCL2	-O.15	***	-0.23	0.013
	CD68	-0.031	0.31	-0.38	***
	IL10	-0.023	0.45	-0.5	***
M2	CD163	-0.12	***	-0.43	***
	VSIG4	-0.01	**	-0.48	***
	MSA4A	-0.1	**	-0.54	***
M1	NOS2	-0.012	0.69	0.27	*
	ROS1	-0.017	0.57	-0.18	0.054

*P < 0.01;

**P < 0.001;

****P < 0.0001*.

Figures

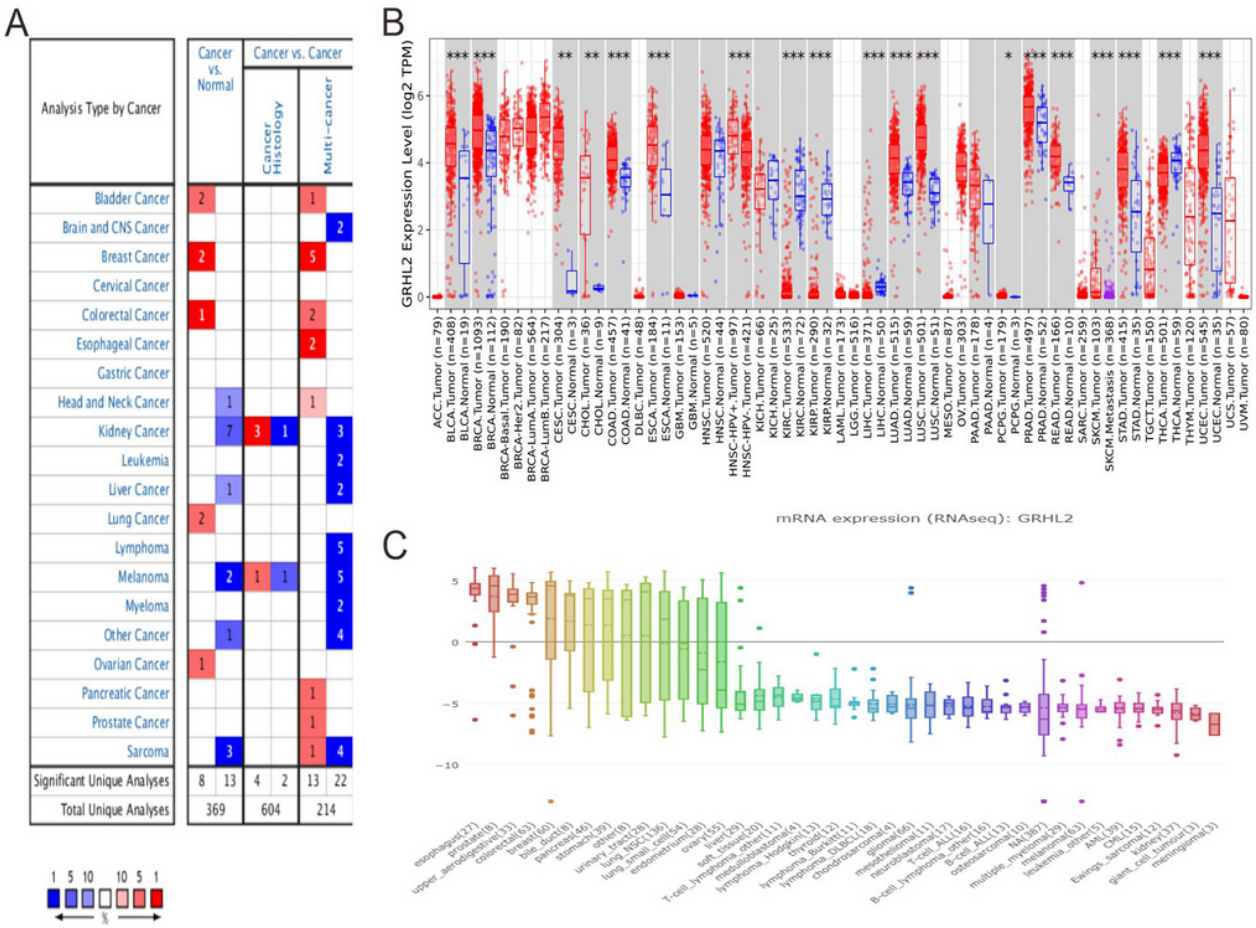

Figure 1

GRHL2 expression levels in different types of human cancers.(A) Increased or decreased GRHL2 in different cancers compared with normal tissues in the Oncomain database.(B) Human GRHL2 expression levels in different tumor types from TCGA database were determined by TIMER(*P<0.05,**P<0.01,***P<0.001).(C) The expression of GRHL2 in different types of cancer cells was obtained from the CCLE database.

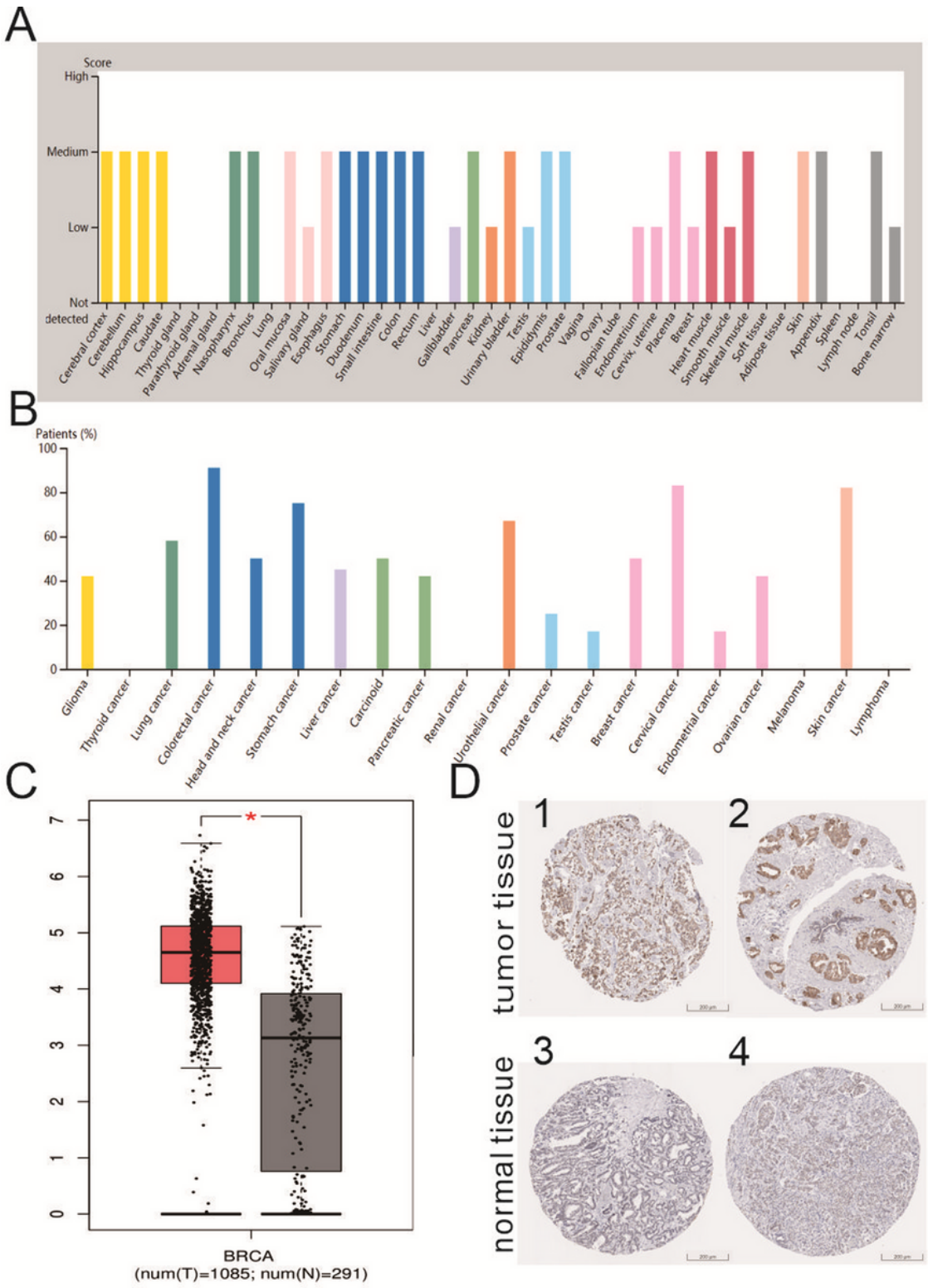

Figure 2

Expression of GRHL2 in Expression of GRHL2 in breast cancer.(A) Expression of GRHL2 in normal tissues. (B) Expression of GRHL2 in pan-cancer tissues. (C) GRHL2 expression was compared between normal tissues (n=291) and breast cancer tissues (n=1086) in TCGA database. (D)The expression of GRHL2 was verified by immunohistochemistry of HPA database.

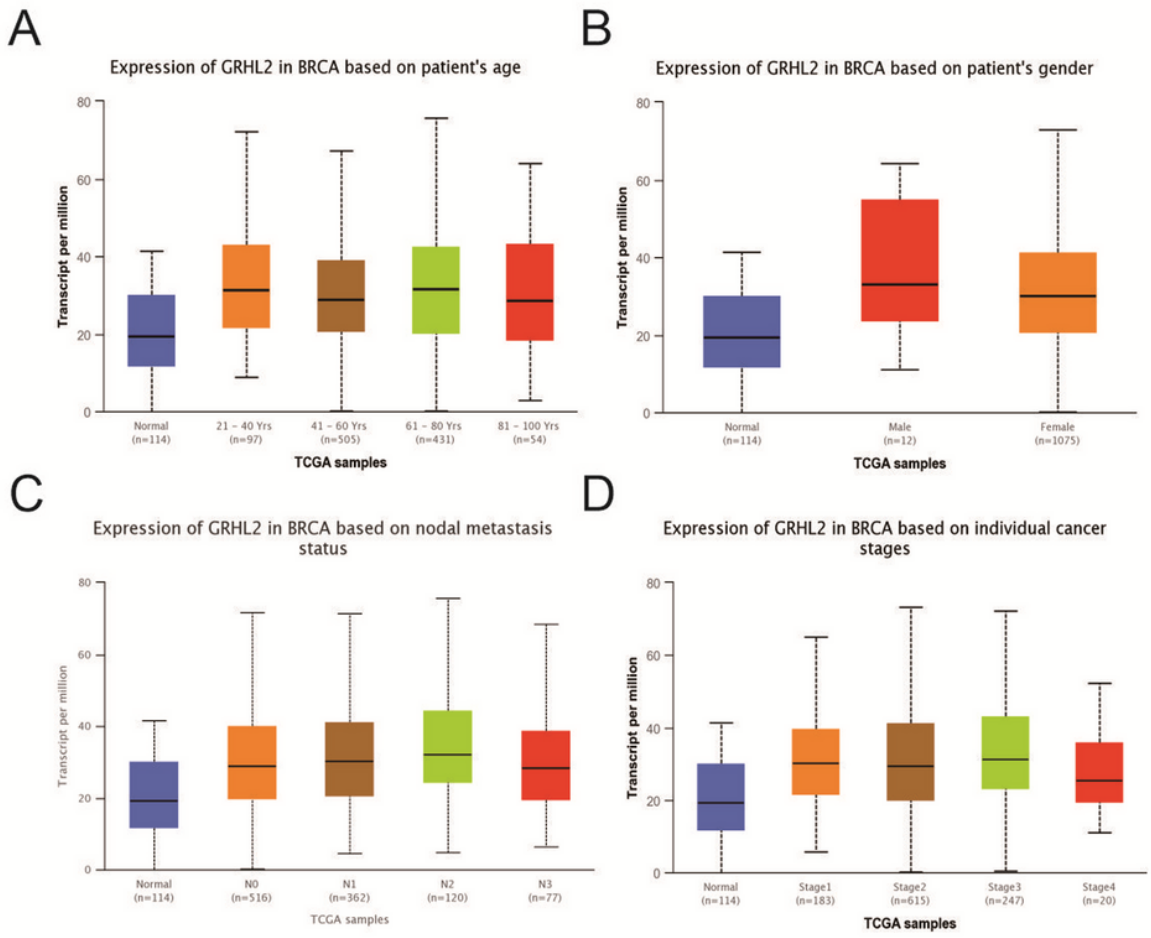

Figure 3

GRHL2 expression is correlated with clinicopathological characteristics. (A) Age (21-40 n=97, 41-60 n=505, 61-80 n=431, 81-100 n=54). (B) Gender (male n=12, female n=1075). (C) Clinical stage (stage 1 n=183, stage 2 n=615, stage 3 n=247, stage 4 n=20). (D) Nodal matastasis status (N0 n=516, N1 n=362, N2 n=120, N3 n=77).

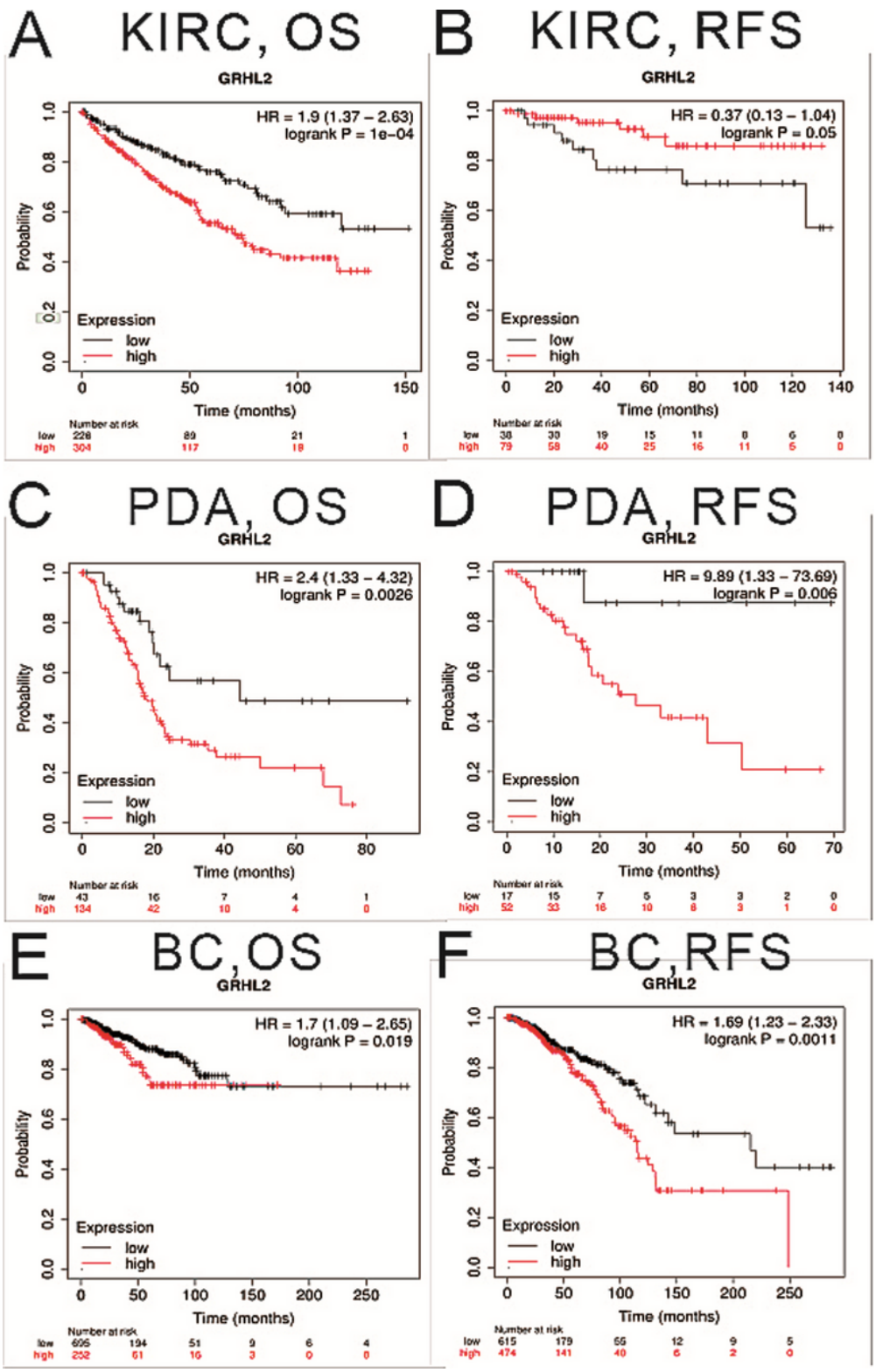

Figure 4

Kaplan-Meier survival curves.OS and RFS of higher and lower expression of GRHL2 in Kaplan-Meier Plotter. (A, B) Kidney renal clear cell carcinoma. (C, D) Pancreatic ductal adenocarcinoma. (E, F) Breast cancer.

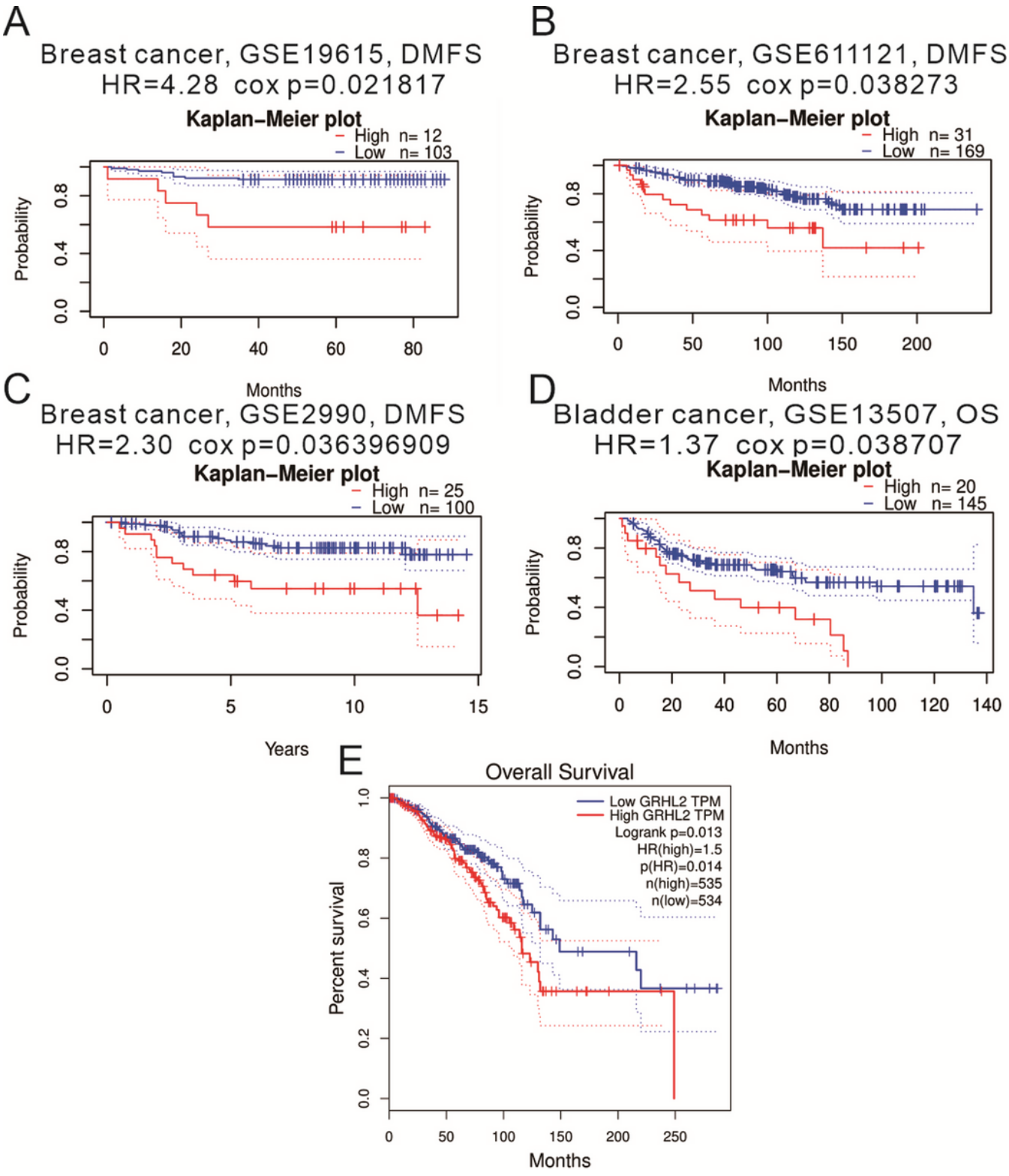

Figure 5

OS and DMFS of higher and lower expression of GRHL2 in PrognoScan and GEPIA. (A) DMFS (n=115) in breast cancer cohort GSE19615. (B) DMFS (n=200) in breast cancer cohort GSE611121. (C) DMFS (n=125) in breast cancer cohort GSE2990. (D) DMFS (n=165) in bladder cancer cohort GSE13507. (E) OS (n=1069) in breast cancer in GEPIA.

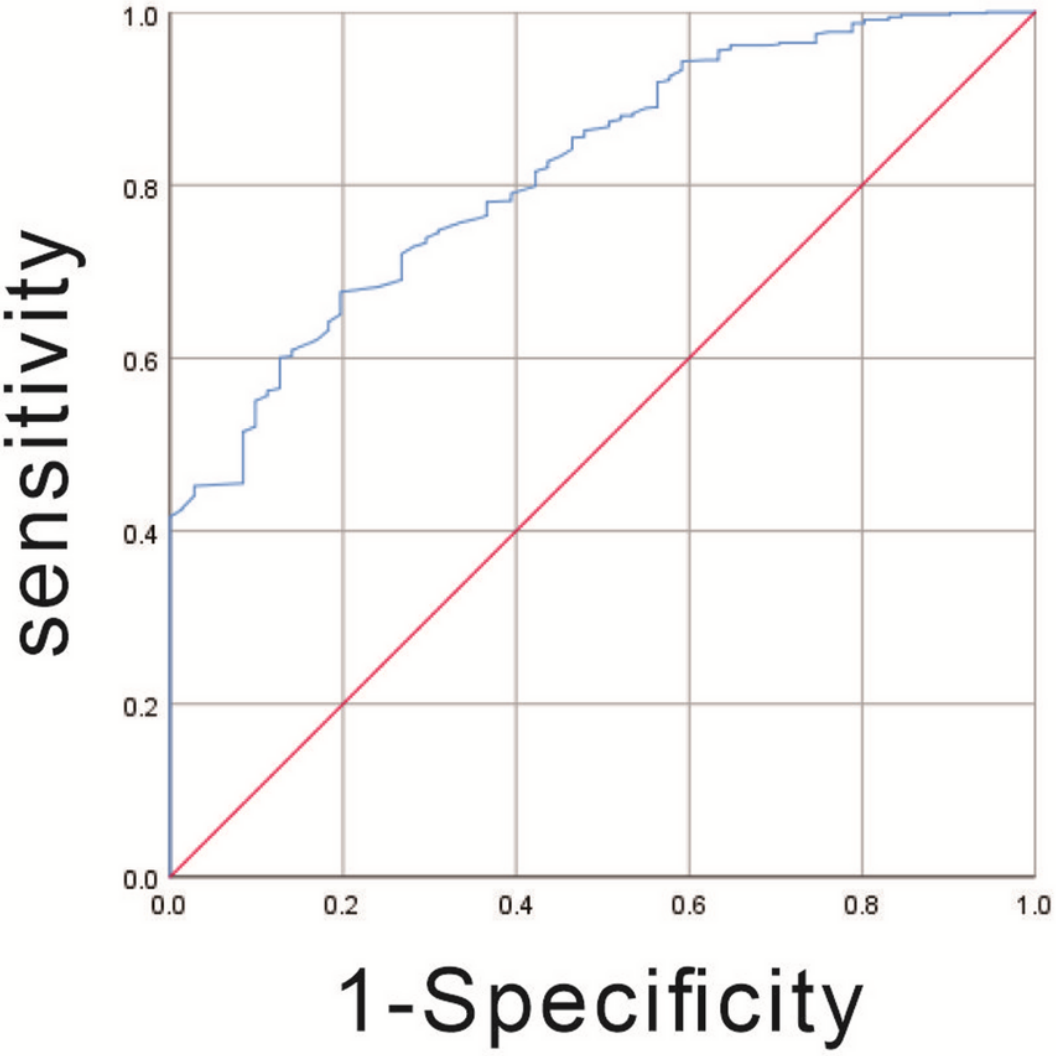

Figure 6

The diagnostic value of GRHL2 expression in breast cancer. ROC curve for GRHL2 expression in normal tissues (n=71) and breast cancer tissues (n=701) in TCGA.

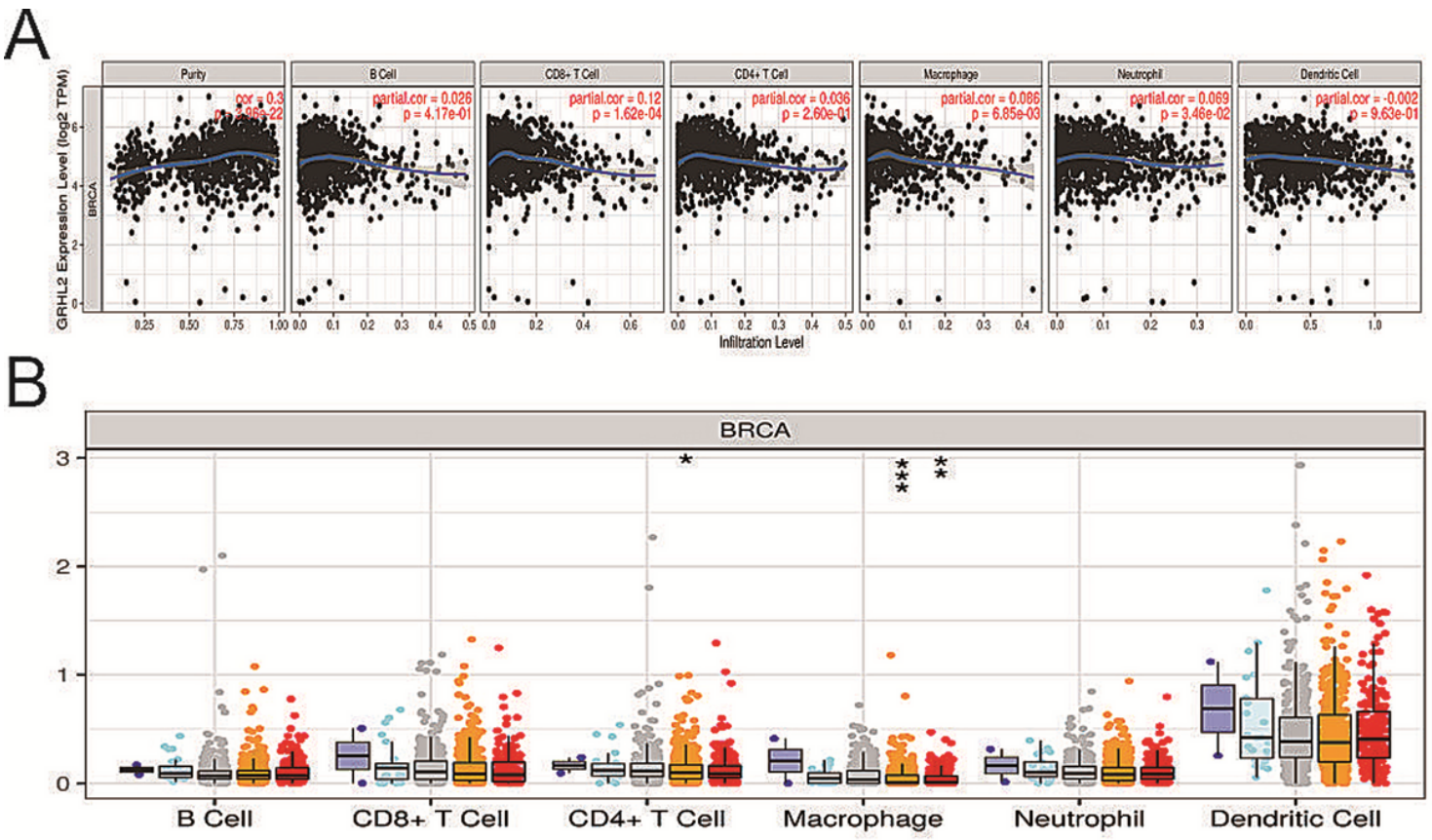

Figure 7

Correlation between GRHL2 expression and immune infiltration in breast cancer. (A) Correlation of GRHL2 expression level with immune cell infiltration levels in breast cancer. (B) Correlation between GRHL2 gene copy number and immune cell infiltration levels in breast cancer.

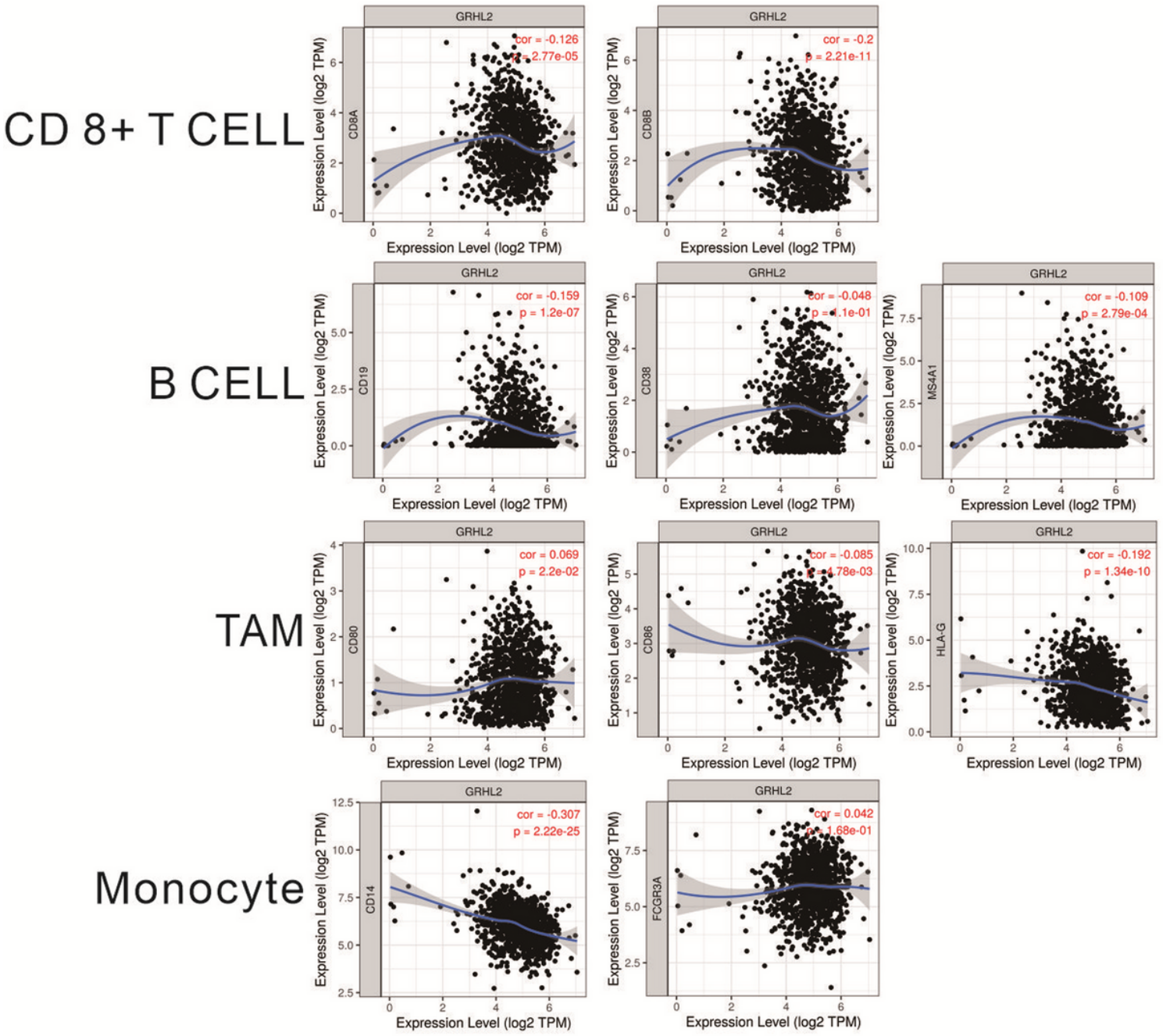

Figure 8

Markers include CD8A, CD8B of CD8+ T cell; CD19, CD38, MS4A1 of B cell; CD80, CD86, HLA-G of TAM; CD14, FCG13 of monocyte.

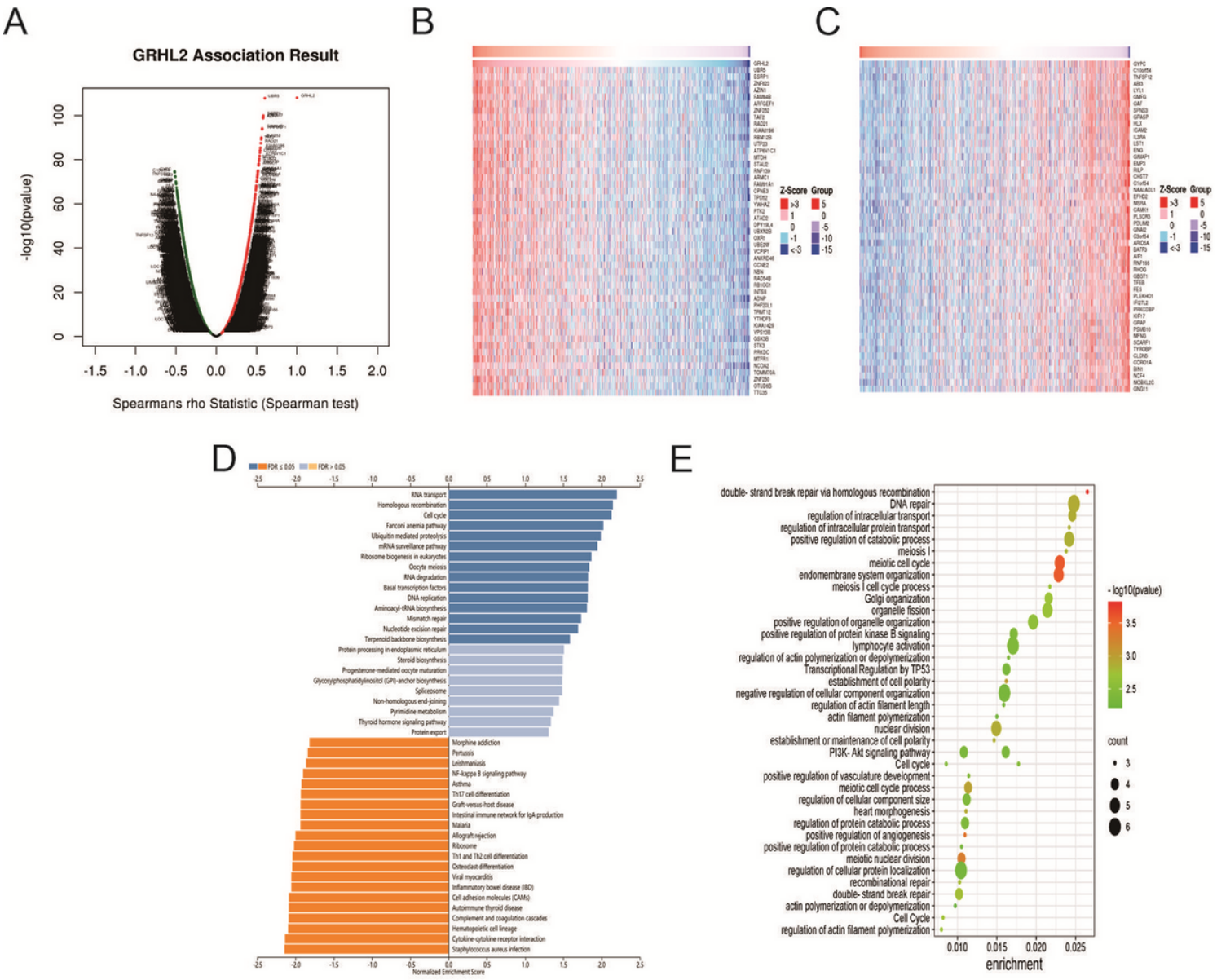

Figure 9

Function enrichment analysis. (A) GRHL2 upstream and downstream genetic volcano map. (B) Heat map of GRHL2 co-expression upstream genes. (C) Heat map of GRHL2 co-expression downstream genes. (D) KEGG signaling pathwaay enrichment analysis. (E) Gene Ontology enrichment analysis.

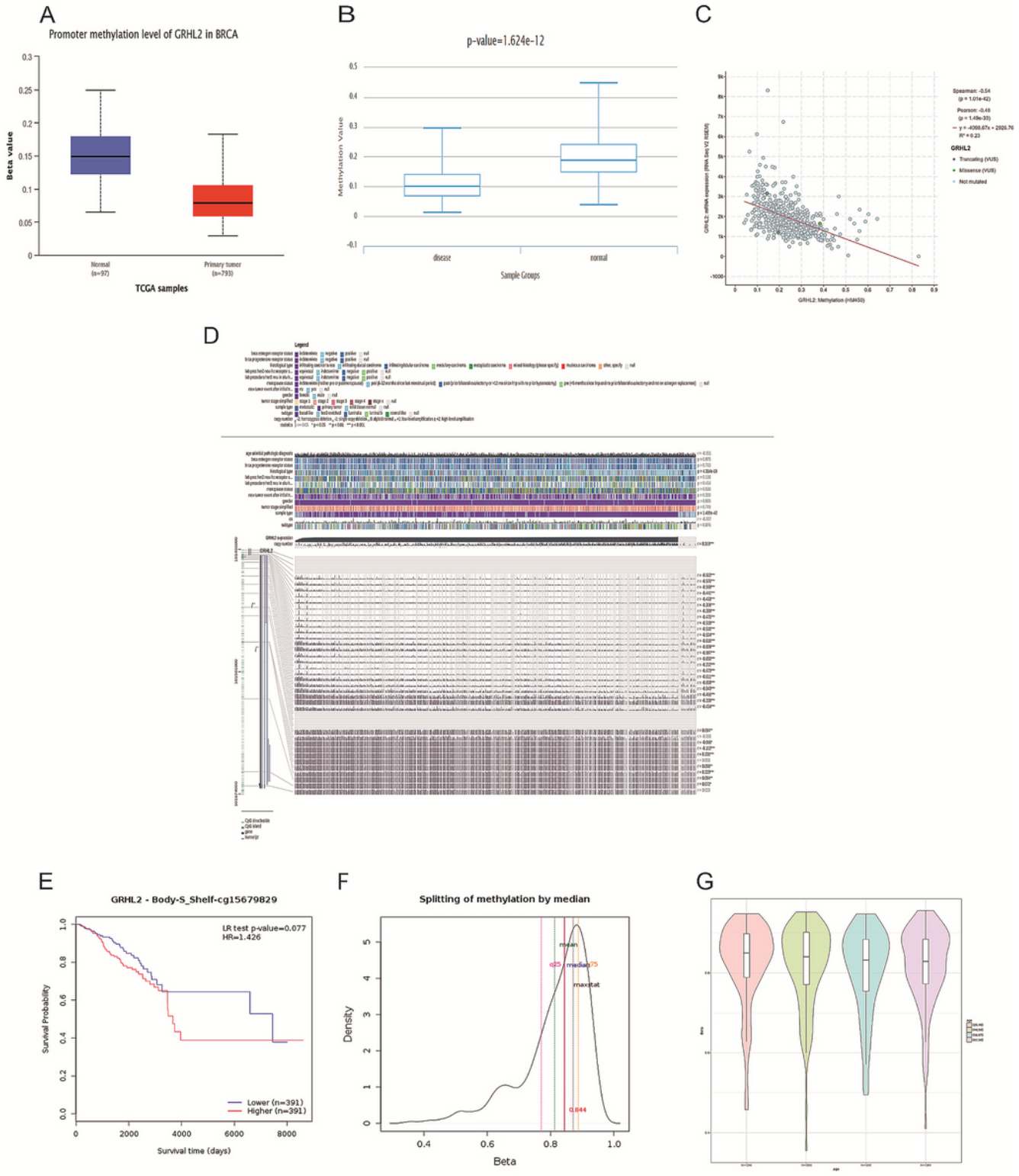

Figure 10

GRHL2 methylation analysis. (A) Using UALCAN analyzed methylation. (B) Using DiseaseMeth version 2.0 analyzed methylation. (C) GRHL2 and methylation expression was shown on Cbioportal. (D) The methylation site of GRHL2 DNA sequence association with gene expression was visualized using MEXPRESS. (E) Survival analysis of cg1567982. (F) Denisity of cg1567982. (G) The violin chart shows

the methylation levels between different age groups. Table 1. Correlations between GRHL2 and Gene Markers of Immune Cells in TIMER.