

GREB1L overexpression correlates with prognosis and immune cell infiltration in lung adenocarcinoma.

Yilin Yu

Fujian Medical University

Zhiping Wang

Fujian Medical University

Qunhao Zheng

Fujian Medical University

Jiancheng Li (✉ jianchengli_jack@126.com)

Fujian Medical University

Research Article

Keywords: GREB1L, Lung adenocarcinoma, TCGA, Prognosis, Immune Methylation.

Posted Date: January 28th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-154298/v1>

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

[Read Full License](#)

Abstract

GREB1L is a protein-coding gene that is an important paralog of GREB1. However, its effects in lung adenocarcinoma (LUAD) have not been determined. Thus, we evaluated the prognostic value of GREB1L in LUAD using bioinformatics approaches. In particular, we evaluated the relationship between GREB1L and LUAD using a wide range of databases and analysis tools, including TCGA, GEO, HPA, TIMER, cBioPortal, and MethSurv. Compared with its expression in normal lung tissues, GREB1L expression was significantly increased in LUAD tissues. A univariate Cox analysis showed that high GREB1L expression levels were correlated with a poor OS in LUAD. Additionally, GREB1L expression was independently associated with OS through a multivariate Cox analysis. GO and GSEA revealed enrichment in cell cycle and methylation. Moreover, high GREB1L expression was associated with poor survival. We also found that the methylation and genetic alteration level was associated with prognosis in patients with LUAD. Finally, an analysis of immune infiltration showed that GREB1L is correlated with immune cell infiltration, PD-1 and PD-L1. In summary, these results indicate that GREB1L is a potential molecular marker for poor prognosis in LUAD and provide additional insight for the development of therapies and prognostic markers.

1 Introduction

Lung cancer is the leading cause of cancer death worldwide. It is reported to have the second-highest incidence of cancer in both men and women in the United States, second only to prostate cancer and breast cancer, respectively. It is also the most common cause of human cancer deaths, accounting for more than 25% of all cancer deaths^{1,2}. Lung adenocarcinoma (LUAD) is the most common type of lung cancer. In recent years, the number of LUAD patients is rising due to smoking and air pollution³. Some studies have shown that the overall 5-year survival rate of patients with advanced LUAD is less than 15%, but if patients are treated with targeted gene-targeting therapy, their survival rate will be improved⁴. However, due to the lack of specific biomarkers of LUAD, the metastasis rate and mortality rate of LUAD patients is very high⁵. Although molecular targeted therapy has made progress in recent years, but more targets need to be identified. Nowadays, immunotherapy has been paid more and more attention from oncologists. T cell is an important medium of tumor immunity. In most tumors, T cell infiltration is a useful prognostic marker^{6,7}. Therefore, the development of lung cancer treatment needs further research, especially to find potential prognostic molecular biomarkers and new targets of immune-related therapy.

GREB1L (GREB1 Like Protein) is a protein-coding gene. It is an important paralog of GREB1 (Growth Regulating Estrogen Receptor Binding 1). It plays a major role in early metanephros and genital development. Previous studies have shown that GREB1L is associated with bilateral renal hypoplasia, inner ear malformation, and deafness⁸⁻¹⁰. In addition, a study confirmed that the methylation level of GREB1L is related to immune response and cytolysis in gastric adenocarcinoma, suggesting that it may be a new prediction and prognostic biomarker that aids in the therapy and predicts the overall survival possibility in patients with gastric adenocarcinoma¹¹. However, research is lacking on the potential

function of GREB1L in LUAD. It is the first study to demonstrate the functional impact of GREB1L in lung adenocarcinoma.

In this study, we synthetically evaluated the prognostic value of GREB1L expression in patients with LUAD from the Cancer Genome Atlas (TCGA) database. Furthermore, the prognostic value of GREB1L expression in LUAD was validated using data from the Gene Expression Omnibus (GEO) databases. Additionally, we performed GSEA GO and KEGG analysis to gain further insights into the biological mechanism of GREB1L in LUAD pathogenesis. We also examined the correlation between GREB1L expression and genetic alteration and methylation. Finally, ssGSEA (single-sample Gene Set Enrichment Analysis) and TIMER were used to explore the relative proportions of different kinds of immune cell infiltration levels in tumor microenvironments to study the relationship of GREB1L, comprehensively analyze and discuss the possible molecular mechanism between GREB1L and tumorigenesis, and tumor-immune interactions.

2 Results

2.1 Clinical characteristics

Our data were obtained from TCGA, including 497 LUAD patients with clinical data and gene expression data. The clinical characteristics, including gender, age, number pack years smoked, T stage, N stage, M stage, pathological stage, vital status, and gene expression data were collected (**Table1**).

2.2 GREB1L is upregulated in LUAD

The results showed that GREB1L was overexpressed in LUAD tissues than in normal tissues ($p < 0.001$) (**Figure 1A**). In paired samples, the expression of GREB1L in the LUAD groups was significantly higher than the adjacent normal groups ($p < 0.001$) (**Figure 1B**). After exploring the gene expression of GREB1L in LUAD, we tried to examine the protein expression patterns by the Human Protein Atlas. As shown in **Figure 1C-D**, the expressions of GREB1L were low in normal lung tissues, while median protein expression of GREB1L was observed in LUAD tissues. Besides, the ROC showed that the expression of GREB1L in LUAD was 0.826 (CI: 0.788–0.864) (**Figure1E**).

2.3 High GREB1L expression is associated with adverse outcomes in LUAD

In order to explore the correlation between GREB1L expression and prognosis, the expression level of GREB1L from TCGA was classified as low- and high- expression according to the median expression value. The results showed that the overall survival (OS) of high GREB1L expression was significantly poorer than low GREB1L expression ($p < 0.01$) in LUAD (**Figure 2A**). To further validate the correlation between GREB1L expression and overall survival, we examined the GSE13213, GSE30219, and GSE 50081 datasets. The results also showed that high GREB1L expression had a worse OS than low GREB1L expression (**Figure 2B-D**). We next used the univariate and multivariate Cox regression model to explore the prognostic factors in LUAD. It showed that high GREB1L expression levels were correlated with

inferior OS. Due to the missing data of the M stage over 20% in TCGA, it was not included in the multivariate analysis. Multivariate analysis showed that GREB1L expression was an independent prognostic factor for LUAD in both TCGA and GSE13213 datasets (**Table 2**).

2.4 Development of a prognostic model based on GREB1L and clinical factors

The multivariate analysis result indicated that GREB1L is an independent prognostic factor in LUAD. We then constructed a prediction model for overall survival by fitting the expression of GREB1L and other clinical parameters. We established a nomogram to integrate GREB1L as a LUAD biomarker (**Figure 3A**). A higher point on the nomogram represented a worse prognostic factor. The Calibration curve evaluated the nomogram's performance of GREB1L. The C-index was 0.6424 of GREB1L with 1000 bootstrap resamples for the nomogram (**Figure 3B-D**). To sum up, this nomogram may be a model for predicting survival in LUAD with GREB1L than an individual prognostic factor.

2.5 Functional enrichment analysis by GSEA GO and GSEA KEGG

To elucidate the biological functions of GREB1L, we analyzed the differentially expressed genes (DEGs) between the DNA amplification and non-amplification groups. We divided patients into low- and high-expression groups based on the median GREB1L expression value. GSEA GO analysis showed that 18 biological processes (BP) and 2 cellular components (CC) were enriched (**Figure 4A-G**). The 20 GSEA GO analyses were all upregulated. Among the 18 BP terms, eight were associated with cell cycle, including "organelle fission", "nuclear division", "DNA recombination", "positive regulation of cell cycle process", "DNA replication", "mitotic nuclear division", "meiotic cell cycle", and "chromosome segregation", two were associated with methylation, including "regulation of gene expression, epigenetic" and "gene silencing". The remaining BP terms were "detection of stimulus", "DNA conformation change", "response to virus", "protein-DNA complex subunit organization", "double-strand break repair", "protein-DNA complex assembly", "detection of stimulus involved in sensory perception" and "DNA packaging". The 2 CC terms were "nuclear chromatin" and "chromosomal region". To further explore the potential biological pathways of GREB1L that promote tumor progression, we also performed a GSEA KEGG analysis (**Figure 5A**). The results showed that high GREB1L expression was upregulated in the Cell cycle, MicroRNAs in cancer, and JAK-STAT signaling pathway (**Figure 5B-D**). However, the Wnt signaling pathway, Ras signaling pathway, and cGMP-PKG signaling pathway were down-regulated (**Figure 5E-G**).

2.6 Correlation between GREB1L mutation, hypomethylation, and prognosis in LUAD

After validating the prognosis of GREB1L, we used cBioPortal to analyze the GREB1L expression and its mutation in LUAD. We analyzed genetic alteration in GREB1L and its associations with the prognosis of LUAD patients. As was shown in **Figure 6A**, a high mutation rate of GREB1L was observed in LUAD patients. In the 501 sequenced LUAD patients, the genetic alteration was found in 52 LUAD patients and the mutation rate was 10%. Besides, the result showed that genetic alteration in GREB1L was associated with inferior OS of LUAD patients (**Figure 6B**). These results implied that the genetic mutation of GREB1L could also affect LUAD patients' prognosis. Considering that GSEA GO enrichment analysis found that

the GREB1L may participate in the methylation process, we then analyzed GREB1L methylation and GREB1L expression using the cBioPortal dataset. The results showed that GREB1L expression was highly negatively correlated with methylation ($R = -0.42$, $P < 0.001$) in LUAD (**Figure 6C**). In addition, the MethSurv analysis showed that patients with low GREB1L methylation had a worse overall survival than patients with high GREB1L methylation ($p < 0.05$). We discovered that 4 CpG sites located on the CpG island indicated a poor prognosis, including cg03735496, cg05109245, cg12473406, and cg06711831 (**Figure 7A-D**). Finally, the methylation level of GREB1L is low in LUAD by MethSurv software (**Figure.7E**).

2.7 The correlation between GREB1L expression and the infiltration of the immune cells

We further performed ssGSEA to examine the correlation between GREB1L expression and immune cell infiltration in LUAD. As is shown in **Table 3**, the activated CD4 T cell was positively correlated with GREB1L expression with Spearman up to 0.349 ($P < 0.001$). We observed that GREB1L expression was also positively correlated with central memory CD4 T cell, effector memory CD4 T cell, central memory CD8 T cell, effector memory CD8 T cell, gamma delta T cell, immature B cell, memory B cell, regulatory T cell, T follicular helper cell, Type 1 T helper cell, Type 2 T helper cell, MDSC, Natural killer cell, Natural killer T cell and neutrophil infiltration (all $p < 0.05$). Additionally, PD-1 (PDCD1) and PD-L1 (CD274) expression play an important role in tumor immune escape. They are also a predictive marker for the therapeutic efficacy of immune checkpoint inhibitors (ICIs). The result showed that PD-1 and PD-L1 expression was positively correlated with GREB1L expression ($R = 0.257$ and 0.396 , $P < 0.001$, respectively) in the LUAD-TCGA datasets (**Figure 8A-B**). Finally, the analyses by TIMER software showed that the expression level of GREB1L was also positively correlated with the infiltration of CD8+ T cell ($R = 0.128$, $p < 0.01$), CD4+ T cell ($R = 0.135$, $p < 0.01$), macrophage ($R = 0.177$, $p < 0.001$), neutrophil ($R = 0.272$, $p < 0.001$) and dendritic cell ($R = 0.187$, $p < 0.001$), but not with the infiltration of B cell ($R = 0.082$, $p = 0.0698$) (**Figure 8C**).

3 Discussion

Lung adenocarcinoma is a malignant tumor characterized by uncontrolled growth of cells in the lung and bronchus². The clinical LUAD outcomes are far from satisfactory using current treatments. Therefore, it is crucial to find stable potential biomarkers to predict prognosis and guide individualized therapies. By analyzing TCGA datasets, we found GREB1L was overexpressed in LUAD and correlated with poor OS. Our result showed that it is highly expressed in LUAD by HPA. Meanwhile, the prognostic value of GREB1L was validated using three GEO datasets (GSE13213, GSE30219, and GSE50081). Furthermore, we confirmed that GREB1L is an independent prognostic factor for LUAD using TCGA and GSE13213 datasets. Functional enrichment analyses found that GREB1L was associated with cell cycle and methylation. We also indicated that GREB1L genetic alteration and hypomethylation were associated with inferior OS in LUAD. Further analyses showed that GREB1L overexpression was positively associated with immune cell infiltration. GREB1L expression positively correlated with the expression of PD-1 and PD-L1. Thus, our study provides new insights into understanding the potential roles of GREB1L in tumor immunology and its potential use as cancer therapeutic and prognostic biomarker.

GREB1L is an important paralog of GREB1 (Growth Regulating Estrogen Receptor Binding 1). At present, the biological mechanism of GREB1L in tumors is only little understood. Our GSEA GO enrichment analysis found that GREB1L may participate in cell cycle and methylation in LUAD. Other GO enrichment results are related to the DNA conformation change, protein-DNA complex subunit organization, double-strand break repair, protein-DNA complex assembly, DNA packaging, nuclear chromatin, and chromosomal region, which need to be confirmed by further study. Interestingly, GSEA KEGG analysis found that GREB1L is related to the cell cycle, microRNAs in cancer, JAK-STAT signaling pathway, Wnt signaling pathway, Ras signaling pathway, and cGMP-PKG signaling pathway. These results suggested that GREB1L may play an important role in the tumorigenesis and methylation process. Therefore, targeting GREB1L seems to be an alternative strategy for tumor therapy.

DNA methylation is a common epigenetic mechanism present in all forms of cancer¹². Promoter methylation accompanies gene silencing¹³. In this study, the GSEA GO analysis showed that GREB1L participates in the epigenetic regulation of gene expression and gene silencing. A previous study also confirmed that the methylation level of GREB1L is related to immune response and cytotoxicity in gastric adenocarcinoma, suggesting that it may be a new prediction and prognostic biomarker that aids in the therapy and predicts the overall survival possibility in patients with gastric adenocarcinoma¹¹. In the study, we further explored the mechanism of GREB1L overexpression in LUAD, and our results showed that GREB1L overexpression might be related to GREB1L hypomethylation ($R = -0.42$, $p < 0.001$). Interestingly, GREB1L methylation was associated with the prognosis of LUAD, and hypomethylated patients have worse overall survival, which is consistent with the prognostic value of the mRNA expression of this gene. Although many mechanisms can give rise to elevated gene expression, hypomethylation is one of the main regulatory mechanisms of gene expression.

In recent years, immunotherapy to boost T cell functionality in tumors is rapidly becoming standard treatment. The immunotherapy focus has been on recruiting tumor-infiltrating T cells^{14 15}. CD4+ T cells secrete a variety of cytokines that have direct effector functions and activate other immune cells (such as B cells and CD8 T Cells)^{16,17}. In lung cancer, tumor-infiltrating CD4+ T cells play an essential role in the immune response¹⁸. CD4+ T cells affect tumors by allowing CD8+ T cells entry to tumor sites or mucosa^{19,20}. Furthermore, they are also required for the inhibition of angiogenesis at tumor sites²¹. Our results showed that the GREB1L expression was positively correlated with the immune infiltration level of T cells, neutrophils, NK cells, MDSC, and dendritic cells, especially in activated CD4 T cell, central memory CD4 T cell, effector memory CD4 T cell, central memory CD8 T cell, and effector memory CD8 T cell. Similarly, TIMER analysis showed that GREB1L expression was positively correlated with the infiltration of CD4+ T cell, CD8+ T cell, macrophage, neutrophil, and dendritic cell. These results suggest that the expression level of GREB1L may indicate the level of immune infiltration levels of tumor cells, providing a reference for the immunotherapy of LUAD. Interestingly, our results showed that GREB1L expression was also positively correlated with PD-1 and PD-L1 expression in LUAD. High expression of PD-L1 has been detected in many tumors, including NSCLC, and is associated with poor prognosis²². Previous studies indicated that tumor patients with overexpression of PD-L1 have better clinical outcomes in anti-PD-L1

therapy, while some patients with low expression of PD-L1 have an inferior effect²³. Therefore, our result demonstrated that GREB1L might affect immune cell infiltration and immunotherapy efficacy, which makes them a predictive biomarker for immunotherapy in LUAD patients.

Although this study improved our understanding of GREB1L in LUAD, there were some limitations. Firstly, the correlation between GREB1L mRNA and protein expression should be verified using cellular and clinical experiments. Secondly, we also cannot clearly estimate the direct mechanisms of GREB1L involved in the development of LUAD. Thirdly, the specific role of GREB1L in the development of LUAD should be comprehensively elucidated. Further experimental researches are needed to confirm it. In order to further investigate the mechanism of GREB1L in LUAD, we plan to conduct cellular experiments in the near future.

4 Conclusion

In conclusion, our findings suggested that GREB1L overexpression is an independent adverse prognostic factor in LUAD. GREB1L DNA amplification and promoter demethylation might contribute to GREB1L upregulation, and GREB1L DNA amplification, genetic alteration, and hypomethylation are associated with poor outcome. Besides, GREB1L mediates increased immune cell infiltration in the tumor microenvironment. This study demonstrates GREB1L as a prognostic biomarker for LUAD, highlighting its potential as a predictive biomarker and an immunotherapy target.

5 Materials And Methods

5.1 Data acquisition

Datasets from the TCGA database were included: gene expression data (HTSeq-Counts and HTSeq-FPKM) and the corresponding detailed clinical data from LUAD samples. We downloaded these data from the UCSC Xena browser (version 07-20-2019, <https://xenabrowser.net/datapages/>). Cases with insufficient or missing data were removed from subsequent data processing. Finally, there are 497 LUAD cases and 57 normal cases in our study. LUAD patients were classified into low- and high- GREB1L expression groups according to the median GREB1L expression value. GREB1L expression data and clinical information from datasets GSE13213, GSE30219, and GSE50081 were downloaded from the GEO database and validated for survival analyses. Our datasets used in the study were consistent with the publication guidelines provided by the online database. The ethics approval and informed consent were not required.

5.2 Over-expression of GREB1L in LUAD patients

We explored the expression differences of GREB1L in LUAD patients between non-pair and pair and tissue. The Human Protein Atlas (<https://www.proteinatlas.org>) analyzed mRNA expression and protein expression to examine the expression of GREB1L in LUAD. Antibody HPA 044218 was used for immunohistochemistry staining of GREB1L. Besides, the ROC curve was used to evaluate the diagnostic

value of GREB1L using the pROC R package (version:1.16.2, <https://cran.r-project.org/web/packages/pROC/index.html>) and ggplot2 R package (version:3.3.2, <https://cran.r-project.org/web/packages/ggplot2/index.html>).

5.3 Role of GREB1L in LUAD patient survival

The survival curve is generated by the survminer R package (version 0.4.8, <https://cran.r-project.org/web/packages/survminer/index.html>) and the survival R package (version 0.1.3, <https://cran.r-project.org/web/packages/survivalAnalysis/index.html>).

5.4 Construction and evaluation of the nomogram

The rms R package (Version:6.0-1, <https://cran.r-project.org/web/packages/rms/index.html>) was performed to generate a nomogram. C-index and calibration curve was performed by the Hmisc R package (Version:4.4-1, <https://cran.r-project.org/web/packages/Hmisc/index.html>). In our study, C-index was performed to determine the discrimination of nomogram and used a bootstrap method with 1000 resamples to calculated C-index.

5.5 GO and KEGG analysis by gene set enrichment analysis (GSEA)

Expression datasets (HTSeq-Counts) were compared between low- and high- GREB1L expression groups to identify the differentially expressed genes (DEGs) using the DESeq2 R package (version 1.28.1, <http://www.bioconductor.org/packages/release/bioc/html/DESeq2.html>). Gene set enrichment analysis (GSEA) is a calculation method that determines whether a set of prior defined genes show statistically significant and consistent differences between two biological states²⁴. In this study, GSEA was performed using the clusterProfiler R package (version 3.18.0, <http://bioconductor.org/packages/release/bioc/html/clusterProfiler.html>) to demonstrate the significant functions and pathways of Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) between the low- and high- GREB1L groups. The expression level of GREB1L was used as a phenotype label. An adjusted *P*-value < 0.05, normalized enrichment score (|NES|) > 1, and False discovery rate (FDR) < 0.25 were considered as significant difference.

5.6 Analysis of GREB1L mutation, methylation, and prognosis

The mutation data of GREB1L was obtained from the cBioPortal (<https://www.cbioportal.org/>) web platform. In the study, we explored the genomic profiles of GREB1L with a z-score threshold ± 1.5 . Genetic mutations in GREB1L and their association with overall survival was carried out to identify its prognostic value. GREB1L methylation data were also downloaded from the cBioPortal (<https://www.cbioportal.org/>) web platform. The correlation between the GREB1L gene expression and the GREB1L methylation level (Spearman correlation) was conducted. Moreover, we analyzed the prognostic value of the GREB1L methylation level in LUAD by MethSurv online tool (<https://biit.cs.ut.ee/methsurv/>). It is a web tool to provide survival analysis based on DNA methylation biomarkers using TCGA data.

5.7 Analysis of immune infiltration and its correlation with GREB1L expression

To elucidate the correlation between GREB1L and the immune cells infiltration levels, ssGSEA (single-sample Gene Set Enrichment Analysis) from the GSVA package (version 1.36.3, <http://www.bioconductor.org/packages/release/bioc/html/GSVA.html>) was used to examine the relative proportions of different kinds of immune cells infiltration level in tumor microenvironments to study the relationship of GREB1L and immune infiltration. The Immune reference set was downloaded from the web platform (<http://dx.doi.org/10.1016/j.celrep.2016.12.019>)²⁵. The relative levels of 28 types of tumor-infiltrating immune cells in the immunocyte signatures, including 782 genes for predicting in individual tissue samples. Based on the 28 types of immunocytes' signature genes in the literature, every immunocyte's relative enrichment score was quantified from each tumor sample's gene expression profile²⁶. The correlation between GREB1L and the immune cells infiltration levels was performed by Spearman correlation. Finally, TIMER software was performed to validate the correlation between the different GREB1L expression levels and the infiltration of the immune cells in LUAD samples from the TCGA databas²⁷.

5.8 Statistical analysis

The statistical analysis was analyzed using the SPSS (version 26.0) and R (version 4.0.2). The Wilcoxon signed-rank test and Wilcoxon rank-sum test were performed to investigate the expression of GREB1L in paired and non-paired samples, respectively. The Wilcoxon signed-rank test was used to analyze the relations between GREB1L expression and clinical features. The univariate and multivariate analyses using the Cox regression model were carried out to evaluate death risk, including gender, age, number pack years smoked, T stage, N stage, M stage, pathological stage, and GREB1L expression. All the tests were two-sided, and P -values < 0.05 were considered statistically significant.

Declarations

Acknowledgments

We sincerely thank the public databases, including TCGA, GEO, HPA, TIMER, cBioPortal, and MethSurv, for providing open access. This project was supported by the Innovation of Science and Technology, Fujian Province (Grant number:2018Y9111) and the Financial Foundation of Fujian Province (Grant number: (2020)729). All authors contributed to the drafting of this manuscript, and have read and approved the final version of the manuscript.

Author Contributions

YL Yu, QH Zheng, and JC Li designed this study. YL Yu and ZP Wang contributed to the data collection. YL Yu and ZP Wang analyzed the data. JC Li supervised the study. YL Yu and ZP Wang wrote the manuscript. All authors read and revised the manuscript.

Competing interests

The authors declare no competing interests

Data Availability Statement

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

References

- 1 Global Burden of Disease Cancer, C. *et al.* Global, Regional, and National Cancer Incidence, Mortality, Years of Life Lost, Years Lived With Disability, and Disability-Adjusted Life-Years for 29 Cancer Groups, 1990 to 2017: A Systematic Analysis for the Global Burden of Disease Study. *JAMA Oncol* **5**, 1749-1768, doi:10.1001/jamaoncol.2019.2996 (2019).
- 2 Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2020. *CA Cancer J Clin* **70**, 7-30, doi:10.3322/caac.21590 (2020).
- 3 Zheng, M. Classification and Pathology of Lung Cancer. *Surg Oncol Clin N Am* **25**, 447-468, doi:10.1016/j.soc.2016.02.003 (2016).
- 4 Kris, M. G. *et al.* Using multiplexed assays of oncogenic drivers in lung cancers to select targeted drugs. *JAMA* **311**, 1998-2006, doi:10.1001/jama.2014.3741 (2014).
- 5 Aris, V. M. *et al.* Noise filtering and nonparametric analysis of microarray data underscores discriminating markers of oral, prostate, lung, ovarian and breast cancer. *BMC Bioinformatics* **5**, 185, doi:10.1186/1471-2105-5-185 (2004).
- 6 Fridman, W. H., Pages, F., Sautès-Fridman, C. & Galon, J. The immune contexture in human tumours: impact on clinical outcome. *Nat Rev Cancer* **12**, 298-306, doi:10.1038/nrc3245 (2012).
- 7 Becht, E., Giraldo, N. A., Dieu-Nosjean, M.-C., Sautès-Fridman, C. & Fridman, W. H. Cancer immune contexture and immunotherapy. *Current Opinion in Immunology* **39**, 7-13, doi:10.1016/j.coi.2015.11.009 (2016).
- 8 Boissel, S. *et al.* Genomic study of severe fetal anomalies and discovery of GREB1L mutations in renal agenesis. *Genet Med* **20**, 745-753, doi:10.1038/gim.2017.173 (2018).
- 9 De Tomasi, L. *et al.* Mutations in GREB1L Cause Bilateral Kidney Agenesis in Humans and Mice. *Am J Hum Genet* **101**, 803-814, doi:10.1016/j.ajhg.2017.09.026 (2017).
- 10 Schrauwen, I. *et al.* De novo variants in GREB1L are associated with non-syndromic inner ear malformations and deafness. *Hum Genet* **137**, 459-470, doi:10.1007/s00439-018-1898-8 (2018).

- 11 Hu, S., Yin, X., Zhang, G. & Meng, F. Identification of DNA methylation signature to predict prognosis in gastric adenocarcinoma. *J Cell Biochem*, doi:10.1002/jcb.28450 (2019).
- 12 Klutstein, M., Nejman, D., Greenfield, R. & Cedar, H. DNA Methylation in Cancer and Aging. *Cancer Res* **76**, 3446-3450, doi:10.1158/0008-5472.CAN-15-3278 (2016).
- 13 Yamashita, K., Hosoda, K., Nishizawa, N., Katoh, H. & Watanabe, M. Epigenetic biomarkers of promoter DNA methylation in the new era of cancer treatment. *Cancer Sci* **109**, 3695-3706, doi:10.1111/cas.13812 (2018).
- 14 Oja, A. E. *et al.* Functional Heterogeneity of CD4+ Tumor-Infiltrating Lymphocytes With a Resident Memory Phenotype in NSCLC. *Frontiers in Immunology* **9**, doi:10.3389/fimmu.2018.02654 (2018).
- 15 Bruno, T. C. *et al.* Antigen-Presenting Intratumoral B Cells Affect CD4(+) TIL Phenotypes in Non-Small Cell Lung Cancer Patients. *Cancer Immunol Res* **5**, 898-907, doi:10.1158/2326-6066.CIR-17-0075 (2017).
- 16 Kamphorst, A. O. & Ahmed, R. CD4 T-cell immunotherapy for chronic viral infections and cancer. *Immunotherapy* **5**, 975-987, doi:10.2217/imt.13.91 (2013).
- 17 Swain, S. L., McKinstry, K. K. & Strutt, T. M. Expanding roles for CD4(+) T cells in immunity to viruses. *Nat Rev Immunol* **12**, 136-148, doi:10.1038/nri3152 (2012).
- 18 Hiraoka, K. *et al.* Concurrent infiltration by CD8+ T cells and CD4+ T cells is a favourable prognostic factor in non-small-cell lung carcinoma. *Br J Cancer* **94**, 275-280, doi:10.1038/sj.bjc.6602934 (2006).
- 19 Bos, R. & Sherman, L. A. CD4+ T-cell help in the tumor milieu is required for recruitment and cytolytic function of CD8+ T lymphocytes. *Cancer Res* **70**, 8368-8377, doi:10.1158/0008-5472.CAN-10-1322 (2010).
- 20 Nakanishi, Y., Lu, B., Gerard, C. & Iwasaki, A. CD8(+) T lymphocyte mobilization to virus-infected tissue requires CD4(+) T-cell help. *Nature* **462**, 510-513, doi:10.1038/nature08511 (2009).
- 21 Rakhra, K. *et al.* CD4(+) T cells contribute to the remodeling of the microenvironment required for sustained tumor regression upon oncogene inactivation. *Cancer Cell* **18**, 485-498, doi:10.1016/j.ccr.2010.10.002 (2010).
- 22 Ohaegbulam, K. C., Assal, A., Lazar-Molnar, E., Yao, Y. & Zang, X. Human cancer immunotherapy with antibodies to the PD-1 and PD-L1 pathway. *Trends Mol Med* **21**, 24-33, doi:10.1016/j.molmed.2014.10.009 (2015).
- 23 Patel, S. P. & Kurzrock, R. PD-L1 Expression as a Predictive Biomarker in Cancer Immunotherapy. *Mol Cancer Ther* **14**, 847-856, doi:10.1158/1535-7163.MCT-14-0983 (2015).

- 24 Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550, doi:10.1073/pnas.0506580102 (2005).
- 25 Charoentong, P. *et al.* Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade. *Cell Rep* **18**, 248-262, doi:10.1016/j.celrep.2016.12.019 (2017).
- 26 Bindea, G. *et al.* Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. *Immunity* **39**, 782-795, doi:10.1016/j.immuni.2013.10.003 (2013).
- 27 Li, T. *et al.* TIMER: A Web Server for Comprehensive Analysis of Tumor-Infiltrating Immune Cells. *Cancer Res* **77**, e108-e110, doi:10.1158/0008-5472.CAN-17-0307 (2017).

Tables

Table 1

Clinical characteristics of LUAD patients in TCGA.

Clinical characteristics	Total (497)	Percentage (%)
Gender		
male	228	45.9
female	269	54.1
Age		
<=65 years old	236	48.5
> 65 years old	251	51.5
Number pack years smoked		
<40	167	33.6
>=40	174	35
T stage		
T1	166	33.4
T2	267	53.7
T3	43	8.7
T4	18	3.6
N stage		
N0	321	64.6
N1	94	18.9
N2	69	13.9
N3	2	0.4
M stage		
M0	331	66.6
M1	24	4.8
TNM stage		
stage1	267	53.7
stage2	118	23.7
stage3	80	16.1
stage4	25	5
Vital status		
dead	180	36.2
alive	317	63.8
GREB1L expression		
low	248	49.9
high	249	50.1

Table 2

The univariate and multivariate analyses of overall survival according to GREB1L expression, after adjusting for other potential predictors in TCGA and GSE13213.

Clinicopathologic variable	Total(N)	HR (95% CI)	p-value
TCGA-LUAD			
a.			
Gender (Male vs. Female)	497	0.954(0.711~1.279)	0.752
Age (>65 vs. <=65)	487	1.221(0.907~1.643)	0.188
number pack years smoked (>40 vs. <=40)	341	1.026(0.714~1.475)	0.888
T stage (T2/T3/T4 vs. T1)	494	1.678(1.187~2.373)	0.003
N stage (N2/N3 vs. N0/N1)	486	2.274(1.589~3.255)	<0.001
M stage (M1 vs. M0)	355	2.129(1.243~3.648)	0.006
Pathologic stage (Stage III/Stage IV vs. Stage I/Stage II)	490	2.629(1.924~3.591)	<0.001
GREB1L (High vs. Low)	497	1.554(1.156~2.088)	0.003
b.			
T stage (T2/T3/T4 vs. T1)		1.557(1.014~2.392)	0.043
N stage (N2/N3 vs. N0/N1)		1.013(0.576~1.781)	0.964
Pathologic stage (Stage III/Stage IV vs. Stage I/Stage II)		2.141(1.258~3.646)	0.005
GREB1L (High vs. Low)		1.441(1.0645~1.952)	0.018
GSE13213			
a.			
Gender (Male vs. Female)	117	0.736(0.419~1.293)	0.286
Age (>65 vs. <=65)	117	1.394(0.784~2.480)	0.258
Smoking (Yes vs. No)	117	1.361(0.774~2.391)	0.284
T stage (T2/T3/T4 vs. T1)	117	1.539(0.866~2.735)	0.142
N stage (N2 vs. N0/N1)	117	2.810(1.543~5.116)	0.001
Pathologic stage (Stage III vs. Stage I/Stage II)	117	2.974(1.659~5.331)	<0.001
EGFR status (Mut vs. Wt)	117	1.021(0.574~1.815)	0.945
K-ras status (Mut vs. Wt)	117	1.530(0.717~3.267)	0.272
p53 status (Mut vs. Wt)	117	1.357(0.764~2.412)	0.298
GREB1L (High vs. Low)	117	2.564(1.408~4.666)	0.002
b.			
N stage (N2 vs. N0/N1)		1.147(0.261~9.515)	0.856
Pathologic stage (Stage III vs. Stage I/Stage II)		2.217(0.5174~9.515)	0.284
GREB1L (High vs. Low)		2.207(1.195~4.077)	0.011

Table 3

The association between the expression level of GREB1L and the immune infiltration in the tumor microenvironment.

Immune cell	Spearman correlation	P-value
Activated CD4 T cell	0.349	<0.001
Central memory CD4 T cell	0.103	0.018
Central memory CD8 T cell	0.207	<0.001
Effector memory CD4 T cell	0.19	<0.001
Effector memory CD8 T cell	0.232	<0.001
Gamma delta T cell	0.209	<0.001
Immature B cell	0.121	0.005
Memory B cell	0.245	<0.001
Regulatory T cell	0.222	<0.001
T follicular helper cell	0.089	0.042
Type 1 T helper cell	0.159	<0.001
Type 17 T helper cell	-0.112	0.010
Type 2 T helper cell	0.162	<0.001
MDSC	0.124	0.004
Monocyte	-0.099	0.024
Natural killer cell	0.156	<0.001
Natural killer T cell	0.245	<0.001
Neutrophil	0.094	0.031

Figures

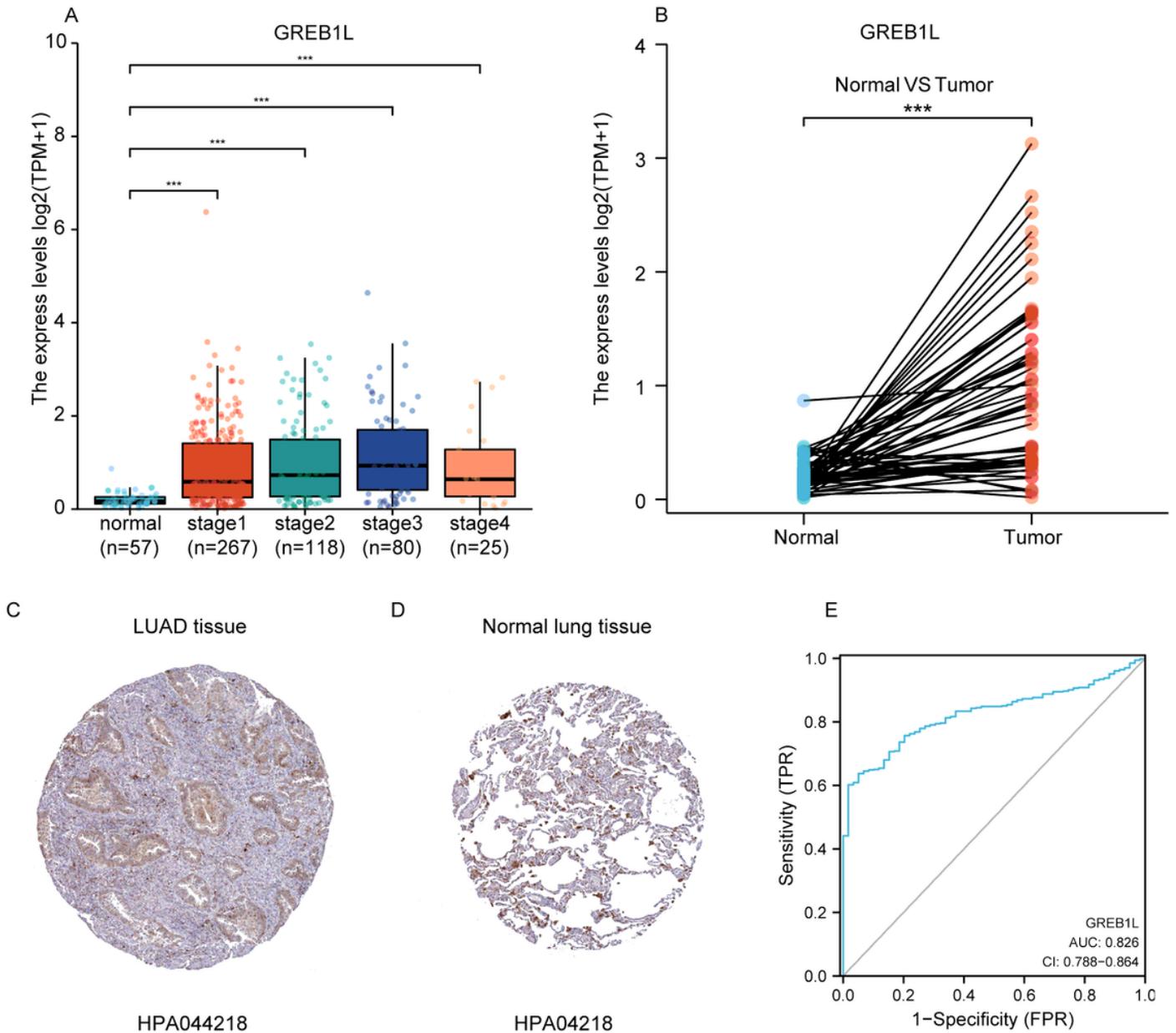


Figure 1

GREB1L expression levels in LUAD from TCGA data. (A) Expression levels of GREB1L in LUAD and normal tissue; (B) The expression of GREB1L in LUAD and its paired adjacent tissues; (C-D) The expression of GREB1L in LUAD was obtained from the Human Protein Atlas; (E) Receiver operating characteristic analysis (ROC) of GREB1L in LUAD. (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$)

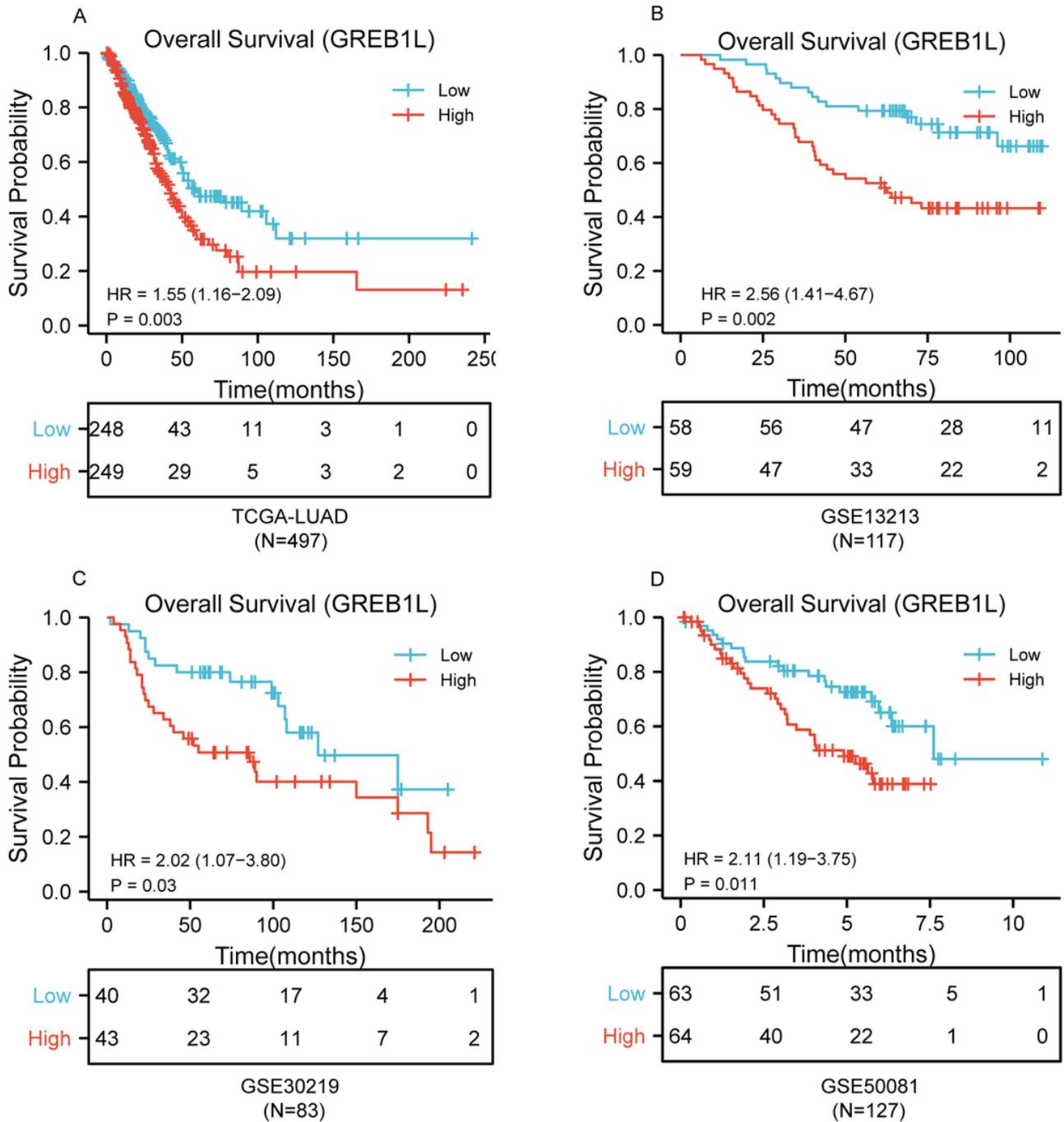


Figure 2

The prognostic value of GREB1L expression in LUAD. (A) Survival curves of OS from TCGA data (n = 497); (B) Survival curves of OS from GSE 13213 (n = 117); (C) Survival curves of OS from GSE30219 (n = 83); (D) Survival curves of OS from GSE50081 data (n = 127).

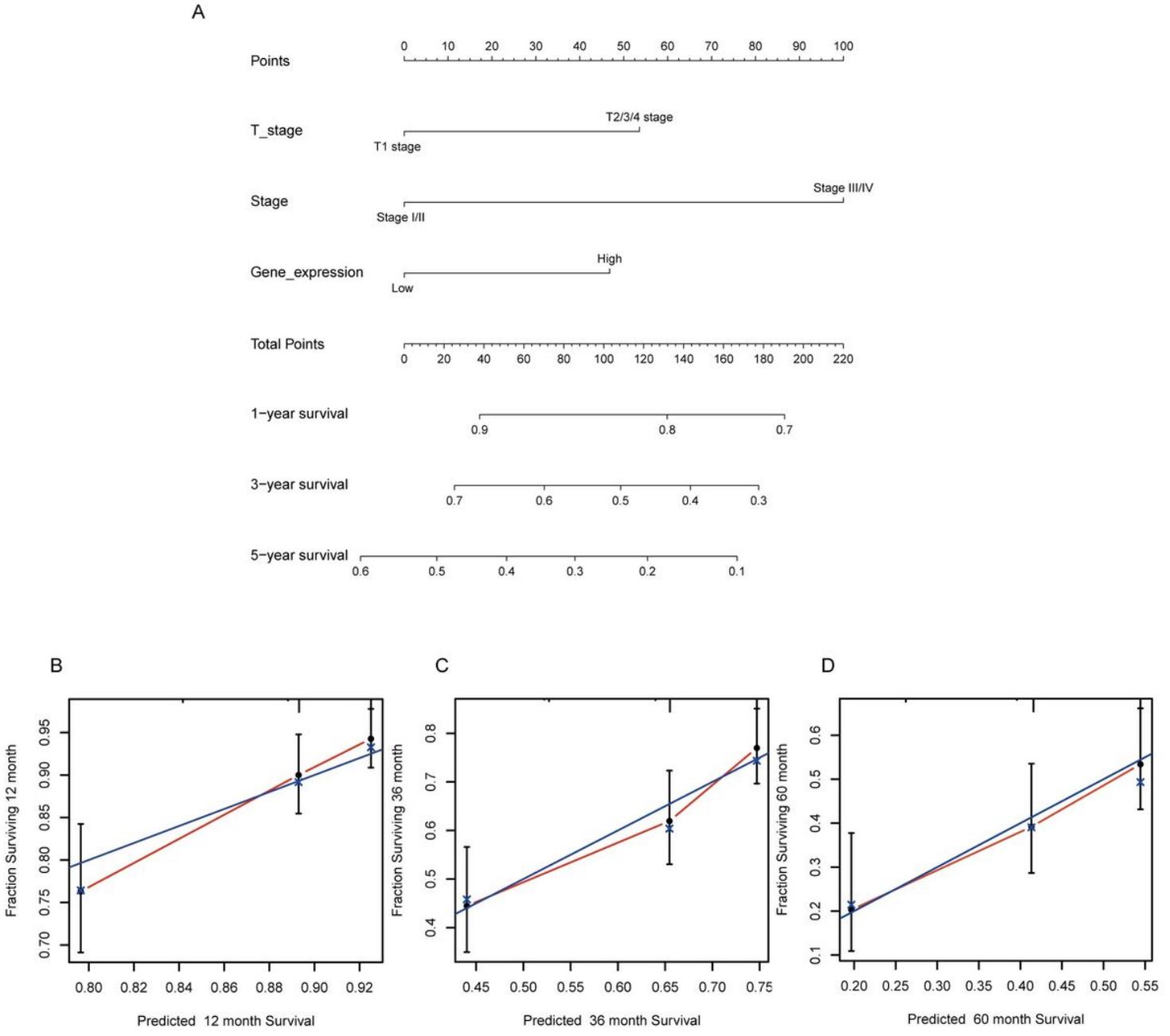


Figure 3

Nomogram and calibration curve for predicting the probability of 1-, 3- and 5-year OS for LUAD patients. (A) A nomogram integrates GREB1L and other prognostic factors in LUAD from TCGA data; (B-D) The calibration curve of the nomogram.

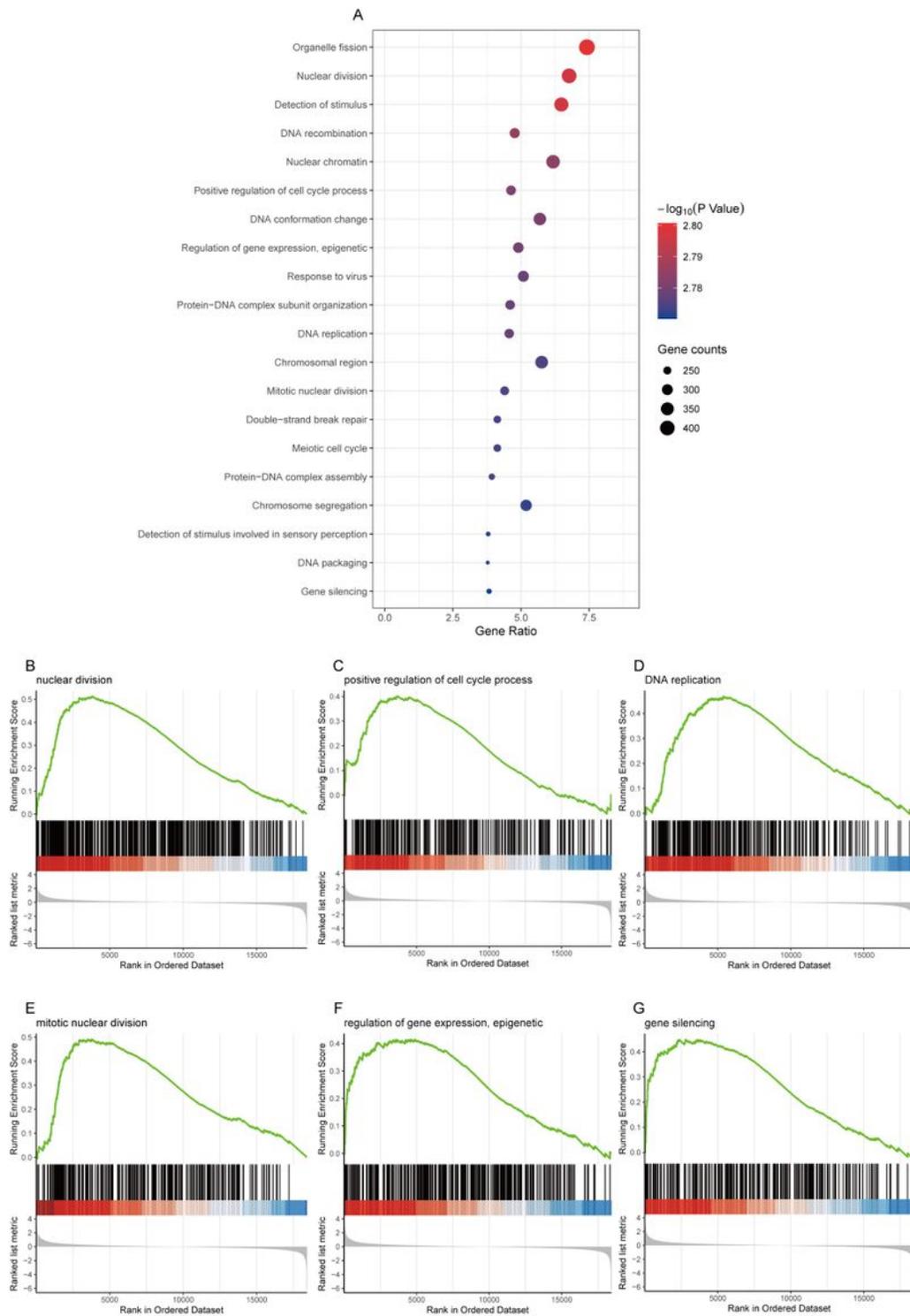


Figure 4

Functional enrichment of GREB1L in LUAD. (A) GSEA GO enrichment analysis of differentially expressed genes (DEGs) in low- and high- GREB1L expression samples; (B-G) Enrichment of genes in the nuclear division, positive regulation of cell cycle process, DNA replication, mitotic nuclear division, epigenetic regulation of gene expression and gene silencing by GSEA GO analysis.

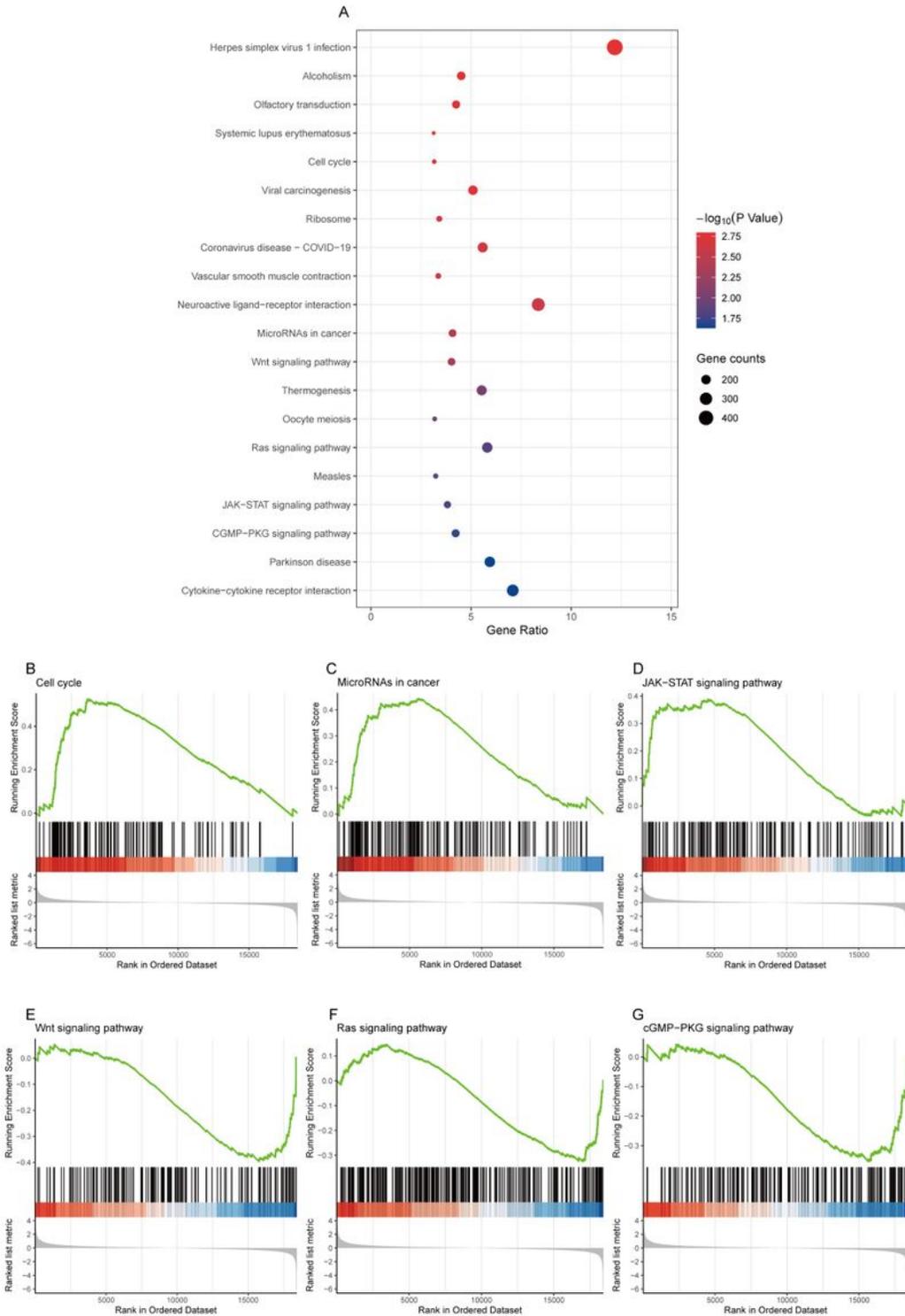


Figure 5

Functional enrichment of GREB1L in LUAD. (A) GSEA KEGG enrichment analysis of differentially expressed genes (DEGs) in low- and high- GREB1L expression samples; (B-G) Enrichment of genes in the Cell cycle, MicroRNAs in cancer, JAK-STAT signaling pathway, Wnt signaling pathway, Ras signaling pathway, and cGMP-PKG signaling pathway by GSEA KEGG analysis.

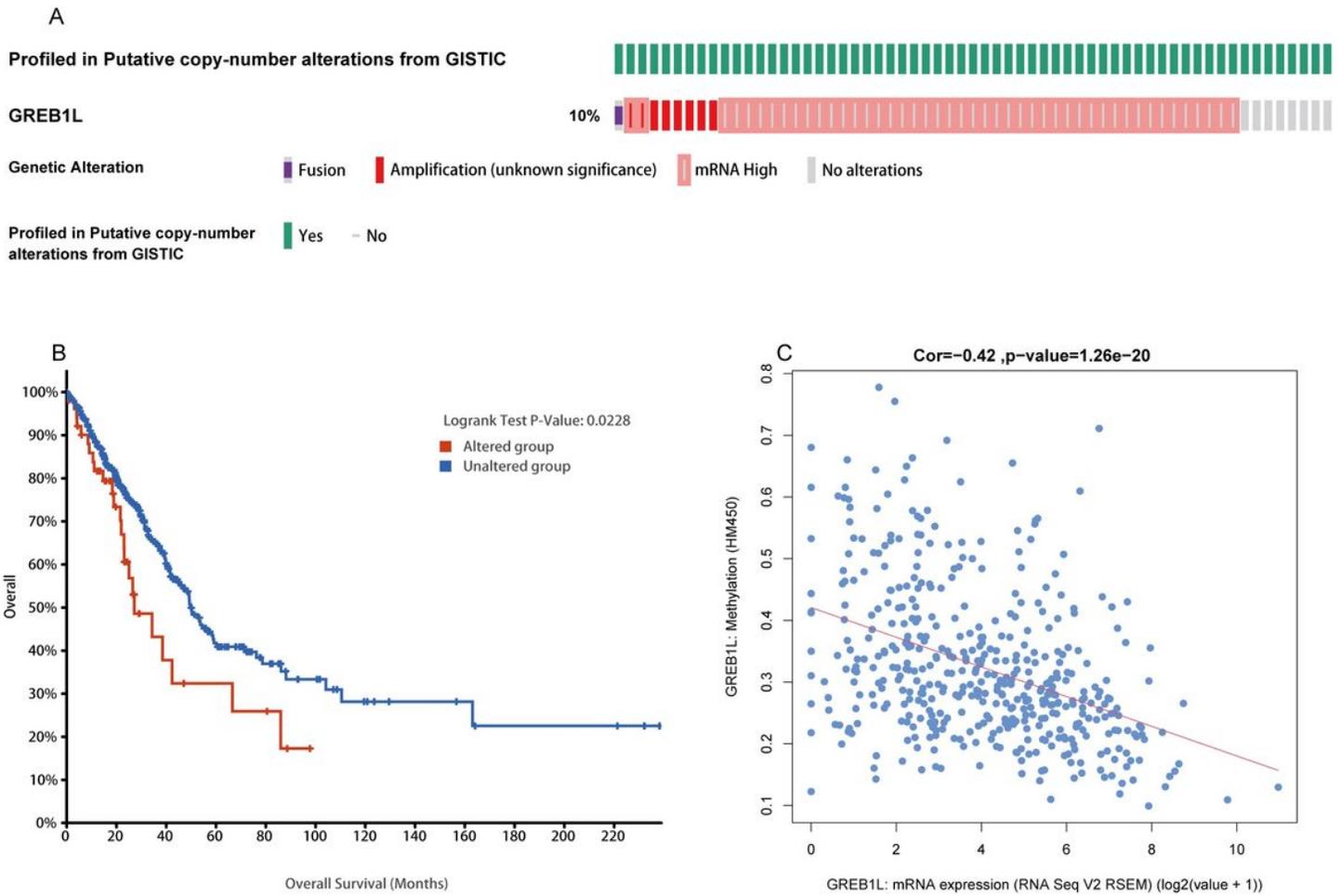


Figure 6

The gene alteration and methylation of GREB1L in LUAD. (A-B) Genetic alteration in GREB1L and its association with OS of LUAD patients; (C) The correlation between GREB1L methylation and its expression level.

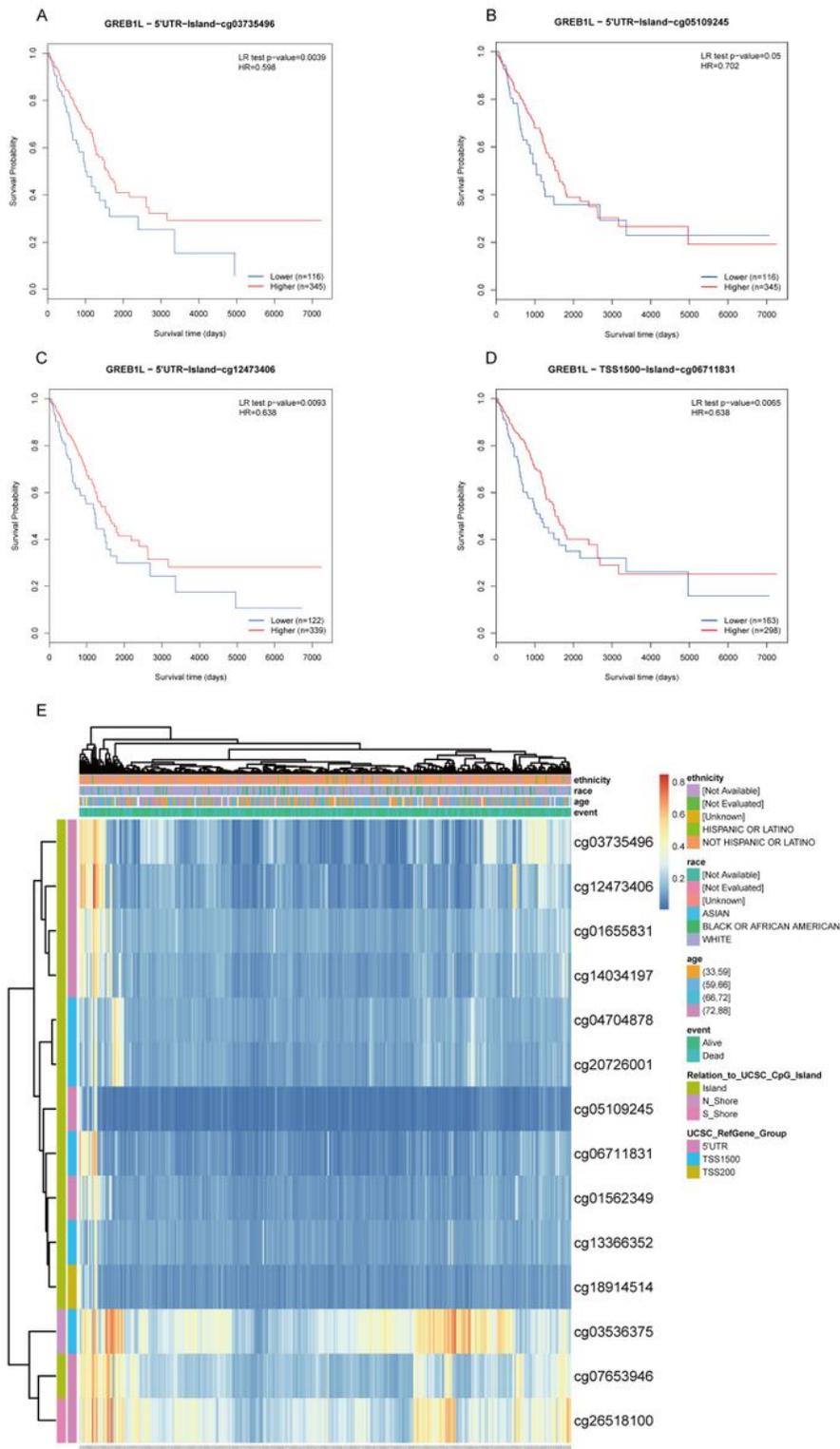


Figure 7

The MethSurv obtained the effect of hypomethylation level and GREB1L expression on prognosis in LUAD. (A-D) the Kaplan-Meier survival of the promoter methylation of GREB1L; (E) The visualization between the methylation level and the GREB1L expression.

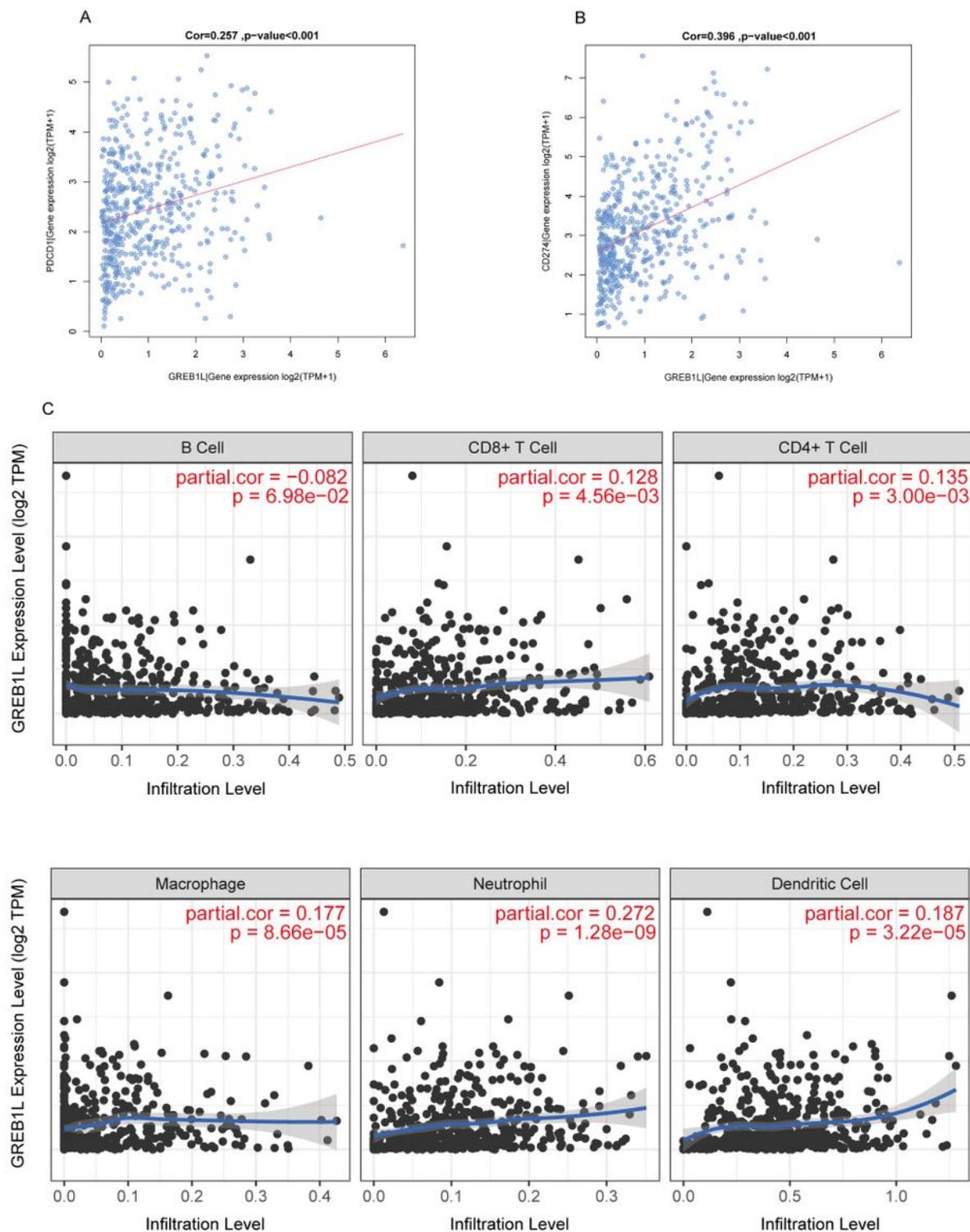


Figure 8

Correlation between PDCD1 (PD1), CD274 (PD-L1), immune cells infiltration, and GREB1L in LUAD. (A) The correlation of GREB1L expression and PDCD1 expression; (B) The correlation of GREB1L expression and CD274 expression; (C) GREB1L expression significantly positively correlated with the infiltration of CD8+ T cell, CD4+ T cell, macrophage, neutrophil, and dendritic cell, but not with the infiltration of B cell.