

RESEARCH ARTICLE

Open Access

Identification of a four-gene panel predicting overall survival for lung adenocarcinoma



Chunyu Li, Qizhong Long, Danni Zhang, Jun Li and Xianming Zhang*

Abstract

Background: Lung cancer is the most frequently diagnosed carcinoma and the leading cause of cancer-related mortality. Although molecular targeted therapy and immunotherapy have made great progress, the overall survival (OS) is still poor due to a lack of accurate and available prognostic biomarkers. Therefore, in this study we aimed to establish a multiple-gene panel predicting OS for lung adenocarcinoma.

Methods: We obtained the mRNA expression and clinical data of lung adenocarcinoma (LUAD) from TCGA database for further integrated bioinformatic analysis. Lasso regression and Cox regression were performed to establish a prognosis model based on a multi-gene panel. A nomogram based on this model was constructed. The receiver operating characteristic (ROC) curve and the Kaplan–Meier curve were used to assess the predicted capacity of the model. The prognosis value of the multi-gene panel was further validated in TCGA-LUAD patients with EGFR, KRAS and TP53 mutation and a dataset from GEO. Gene set enrichment analysis (GSEA) was performed to explore potential biological mechanisms of a novel prognostic gene signature.

Results: A four-gene panel (including DKK1, GNG7, LDHA, MELTF) was established for LUAD prognostic indicator. The ROC curve revealed good predicted performance in both test cohort (AUC = 0.740) and validation cohort (AUC = 0.752). Each patient was calculated a risk score according to the model based on the four-gene panel. The results showed that the risk score was an independent prognostic factor, and the high-risk group had a worse OS compared with the low-risk group. The nomogram based on this model showed good prediction performance. The four-gene panel was still good predictors for OS in LUAD patients with TP53 and KRAS mutations. GSEA revealed that the four genes may be significantly related to the metabolism of genetic material, especially the regulation of cell cycle pathway.

Conclusion: Our study proposed a novel four-gene panel to predict the OS of LUAD, which may contribute to predicting prognosis accurately and making the clinical decisions of individual therapy for LUAD patients.

Keywords: Lung adenocarcinoma, Biomarkers, Four-gene panel, Prognosis, GNG7, DNA methylation

* Correspondence: zxm13078524367@163.com

Department of Respiratory and Critical Care Medicine, Affiliated Hospital of Guizhou Medical University, Guiyang 550004, China



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

Lung cancer is the most frequently diagnosed carcinoma and the leading cause of cancer-related mortality worldwide, with 2.1 million new lung cancer cases and 1.8 million deaths predicted in 2018 [1]. More than 80% of lung cancers are non-small cell lung cancer, mainly lung adenocarcinoma and lung squamous cell carcinoma [2]. Among them, lung adenocarcinoma is on the rise and occupies the main part gradually [3–5]. Traditional treatments for NSCLC included surgery, chemotherapy, and radiotherapy. Although molecular targeted therapy and immunotherapy for NSCLC (especially lung adenocarcinoma) have made great progress in recent years, the OS of NSCLC is still poor, with a 5-year OS of less than 18% [6]. Hence, the identification of accurate prognostic biomarkers and novel and effective therapeutic targets remains particularly urgent for improving the poor survival of NSCLC patients.

Recent advances in genome-wide technologies have promoted the development of tumor biomarkers studies. Large numbers of biomarkers related to diagnosis, prognosis, and drug resistance of cancers have been detected. However, many studies were confined to a single biomarker or a small sample cohort, which made the accuracy and availability of biomarkers insufficient. Therefore, the combination of multiple biomarkers and large sample analysis is more promising. For example, Liu et al. established a six-gene signature prognostic model

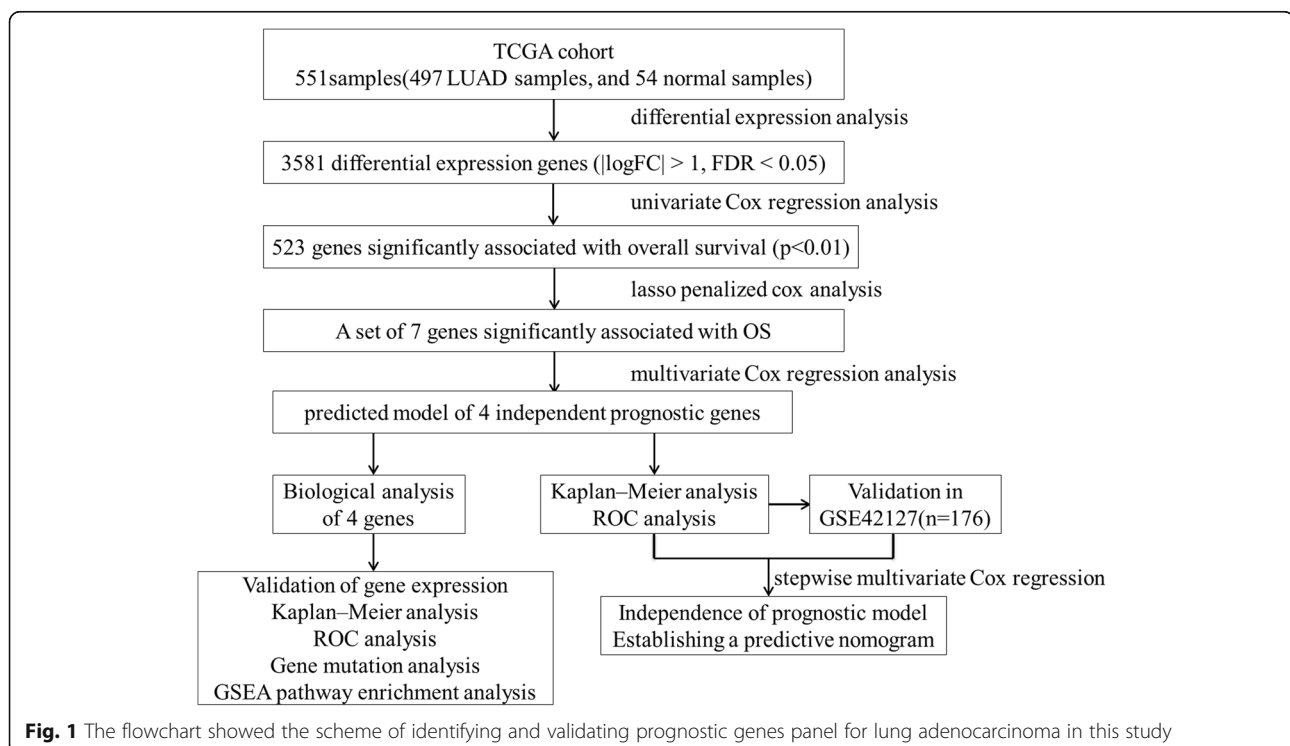
(including CSE1L, CSTB, MTHFR, DAGLA, MMP10, and GYS2) using data from The Cancer Genome Atlas–Liver Hepatocellular Carcinoma Dataset (TCGA-LIHC) [7]. Mining of novel and reliable gene prognostic markers contribute to the prognosis risk stratification and precision therapy of cancer patients.

In the present study, we performed lasso regression, univariate Cox regression, and multivariate Cox regression analysis to screen novel prognostic biomarkers and established a multi-gene panel as a prognostic indicator using data from TCGA-LUAD. ROC curve and Kaplan–Meier curve were used to estimate the prognostic performance of the multi-gene panel. Then, prognosis value of the multi-gene panel was further validated using a dataset from GEO database. Furthermore, we further investigated the clinical significance and possible biological functions of one of the key gene signatures. Overall, our results indicated that the four-gene panel might contribute to predicting OS of LUAD patients effectively and might become a novel target for precision therapy.

Methods

Identification of differentially expressed mRNA in LUAD

The mRNA expression and clinical data were downloaded from the TCGA Database (LUAD mRNA expression (IlluminaHiSeq), containing 497 LUAD samples and 54 normal samples). Raw expression data underwent a log₂ transformation. Differential expression genes



(DEGs) were screened via using limma package in R version 3.5.3 [8]. DEGs were defined according to the criterion: $|\logFC| > 1$, $FDR < 0.05$.

Establishment of the prognostic gene panel

The genes associated with the OS for LUAD patients were identified using Univariate Cox regression analysis, with a cut-off of $P < 0.001$ being considered significant. Lasso penalized regression analysis was utilized to further narrow the range of prognostic genes [9]. Then a prognostic risk model of gene panel was set up based on a linear combination of the multivariate Cox regression model coefficients (β) multiplied with its mRNA expression value. The risk score = (β_{mRNA1} * expression value of mRNA1) + (β_{mRNA2} * expression value of mRNA2) + (β_{mRNA3} * expression value of mRNA3) + ... + (β_{mRNA_n} * expression value of mRNA_n). Each patient was calculated a risk score according to this model. Then we divided these patients into a high-risk group and a low-risk group according to a cut-off value calculated via the R package “survminer” and “survival” and two-sided log-rank test. The predictive performance of the gene panel for OS was

estimated using a time-dependent ROC curve by the “survivalROC” package in R software [10]. The Kaplan–Meier survival curve was executed to compare the survival difference in the high- and low-risk cohort by the “survival” package in R software.

Validation of the prognostic gene panel

To further validate the prognostic value of the gene panel, GSE42127 data from the GEO database were downloaded [11]. The gene expression of GSE42127 data and TCGA-LUAD data were uniformly corrected using the R package “sva” to make them comparable. The risk score was computed with the gene-panel model for each included patient. The Kaplan–Meier curve and ROC curve were performed to validate the predictive capacity of the prognostic gene panel.

The four-gene panel is an independent prognostic factor for LUAD

Univariate and multivariate Cox regression analyses with forwarding stepwise procedure were performed to investigate whether the four-gene panel could be an independent

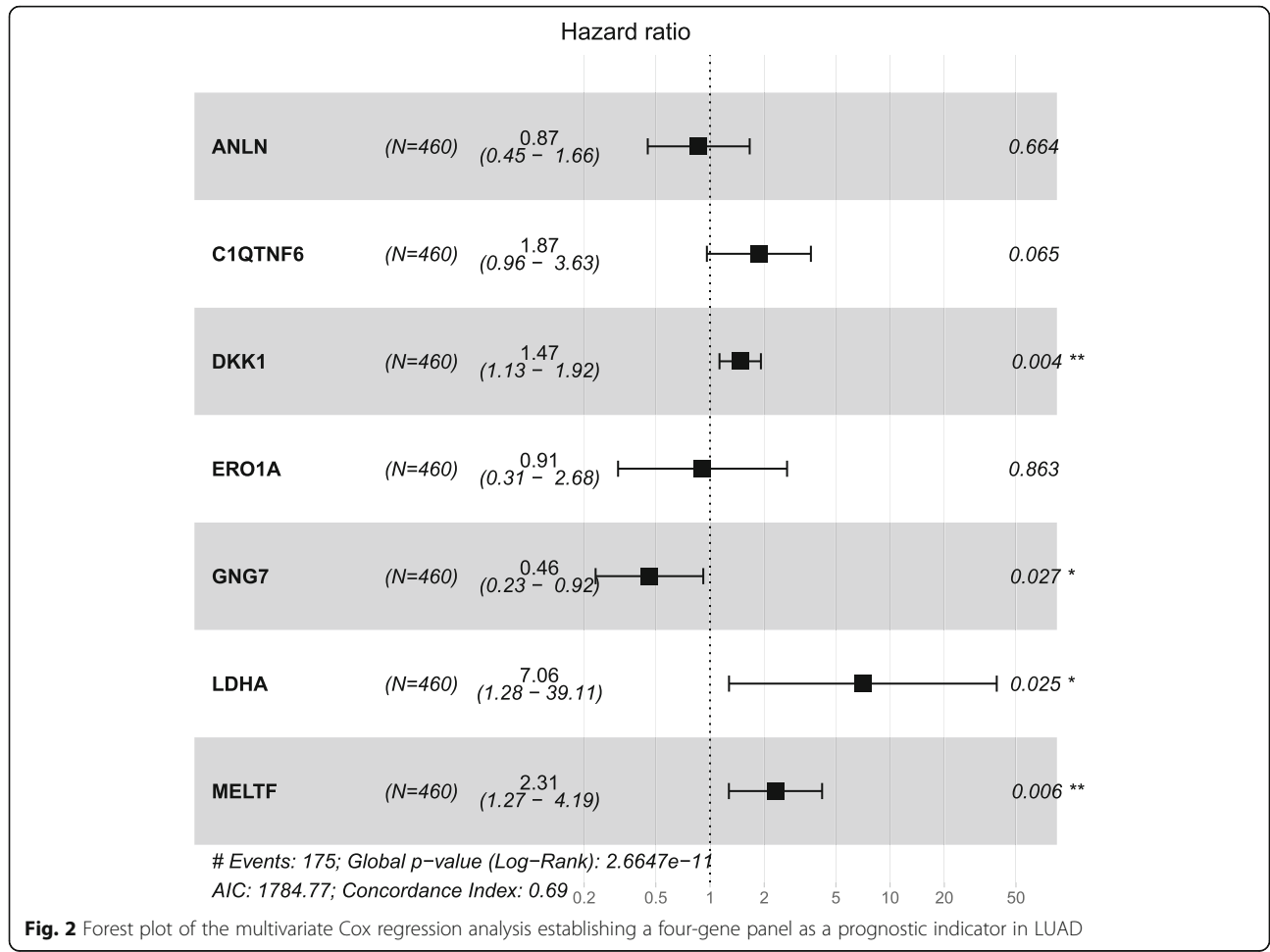


Fig. 2 Forest plot of the multivariate Cox regression analysis establishing a four-gene panel as a prognostic indicator in LUAD

prognostic factor for LUAD patients. Clinical parameters included including gender, age, TNM stage.

Establishment of a predictive nomogram

Nomogram, a simple data evaluation model for the probability of an event, is often used to predict tumor prognosis [12]. Clinical parameters and risk scores from TCGA-LUAD patients were used to build a nomogram in the R package “rms” to detect the predictive probability of 1-year, 3-year, and 5-year OS for LUAD. The discrimination of the nomogram was assessed by using the concordance index (C-index) with a bootstrap method. The calibration curve of the nomogram was plotted by calibrating function of R software to compare predicted OS against observed OS.

MethHC database

MethHC (A database of DNA Methylation and gene expression in Human Cancer) is an online analysis web based on TCGA database resource focused on the DNA methylation of human diseases. We explored DNA methylation level and mRNA expression of GNG7 using MethHC (<http://methhc.mbc.nctu.edu.tw/>) [13].

GSEA

To explore potential biological mechanisms of prognostic gene signature expression on LUAD prognosis, GSEA was used to investigate the enrichment of a priori defined set of genes between the high- and low-expression groups [14]. Gene sets enriched significantly were screened according to the criterion: a normal P -value < 0.05.

Statistical analysis

Univariate Cox regression, lasso regression, multivariate Cox regression analysis, Kaplan–Meier curve, the ROC curve, and log-rank test were used in the present study. All statistical analyses and the generation of relevant figures were operated by R software version 3.5.3. The statistical significance was established at $P < 0.05$.

Results

Identification of DEGs in LUAD

A flowchart for our study was presented in Fig. 1. The lung adenocarcinoma mRNA sequencing dataset was downloaded from the TCGA database. A total of 3581 DEGs were obtained according to the criterion: $|\logFC| > 1$, $FDR < 0.05$, including 2386 up-regulated genes and 1195 down-regulated genes. List, Heatmap, and volcano plot of the DEGs were shown in the supplementary document: Additional file 1, Additional file 2, Additional file 3.

Establishment of a four-gene panel as a prognostic indicator

Univariate Cox regression analysis was performed for identifying the DEGs associated with OS using the “survival” package of R language. Of the 3581 DEGs, 523 genes were identified as being associated with OS for LUAD patients ($p < 0.01$, Additional file 4). Then, lasso regression analysis was implemented to further obtain a stable set of genes (Additional file 5). Seven genes significantly associated with OS were screened out via this analysis (ANLN, C1QTNF6, DKK1, ERO1A, GNG7, LDHA, MELTF). At last, a four-gene

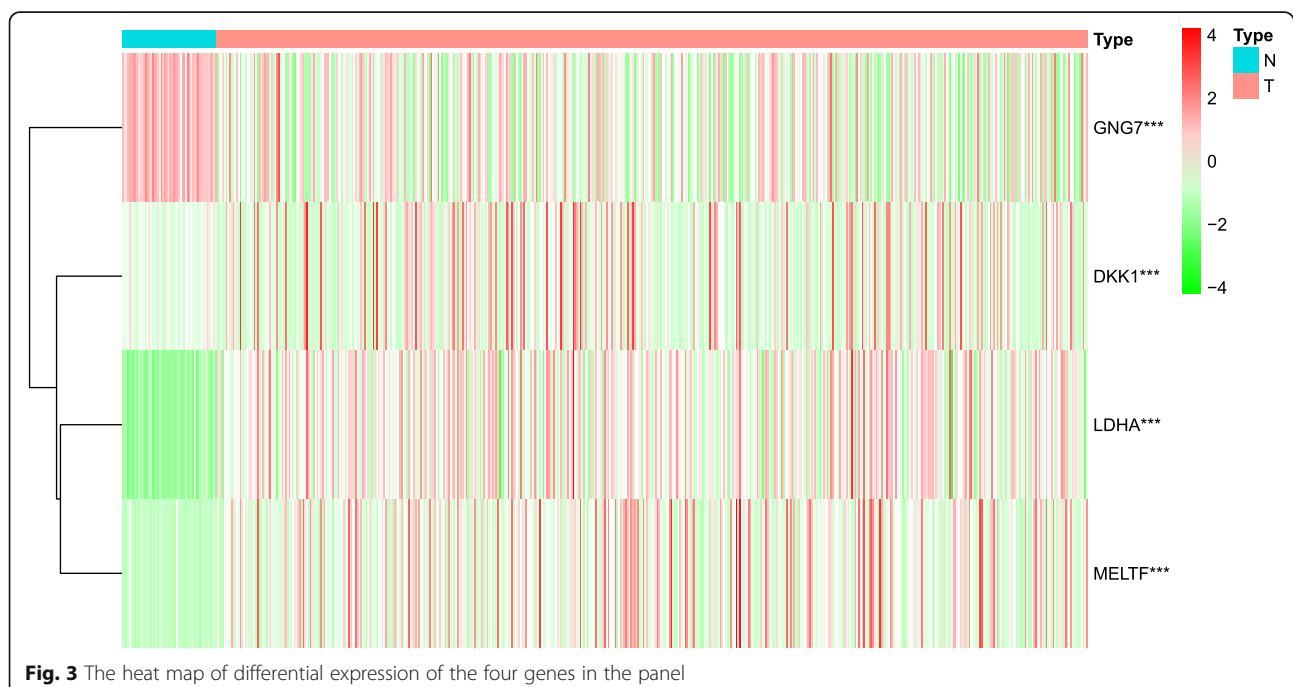


Fig. 3 The heatmap of differential expression of the four genes in the panel

panel as a prognostic indicator was obtained via multivariate Cox regression analysis. The forest map of Cox regression analysis was shown in Fig. 2. The four genes screened were dickkopf WNT signaling pathway inhibitor 1(DDK1), G protein subunit gamma 7(GNG7), lactate dehydrogenase A (LDHA), melano-transferrin (MELTF, also known as MTF1(metal regulatory transcription factor 1)). Among them, DKK1, LDHA and MELTF are high-expressed in tumor tissues compared with tissue adjacent to carcinoma, but GNG7 is low-expressed. The heat map of differential expression was shown in Fig. 3. The risk score = $(0.38606 * \text{Expression}_{\text{DKK1}}) + (-0.77458 * \text{Expression}_{\text{GNG7}}) + (1.95469 * \text{Expression}_{\text{LDHA}}) + (0.83740 * \text{Expression}_{\text{MELTF}})$. Each patient from the TCGA-LUAD database was awarded a risk score based on the Cox regression model composed of the four genes. The results indicated that high-risk group had a worse prognosis compared with the low-risk group. The area under the ROC curve (AUC) of this four-gene panel as a prognostic indicator was

0.740 and was superior to other clinical indicators used for prognostic classification (Fig. 4a).

Validation of the four-gene panel as a prognostic indicator

To further validate the accuracy of the four-gene panel as a prognostic indicator, we computed the risk score of each patient in GSE42127 using the same model. Consistent with previous results, a significantly worse OS was observed in the high-risk group compared with the low-risk group. ROC curve showed that the AUC for OS was 0.752, indicating a better predictive performance compared with other clinical indicators used for prognostic classification (Fig. 4b).

Independent prognostic value of the four-gene panel

Three hundred forty-four patients from the TCGA-LUAD database with complete clinical information including age, gender, and TNM stage were included for further analysis. Univariate and Multivariate Cox regression analysis

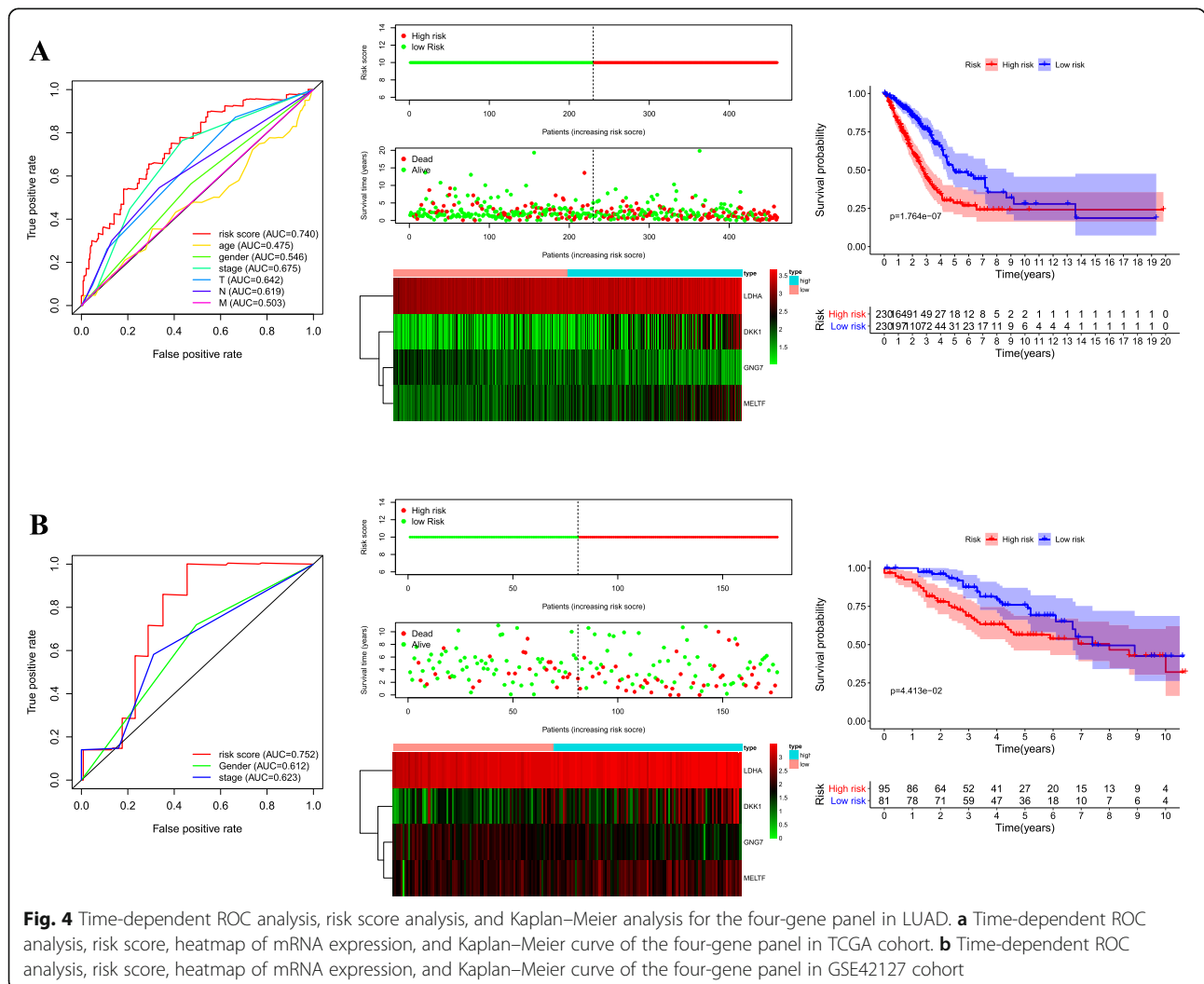


Fig. 4 Time-dependent ROC analysis, risk score analysis, and Kaplan–Meier analysis for the four-gene panel in LUAD. **a** Time-dependent ROC analysis, risk score, heatmap of mRNA expression, and Kaplan–Meier curve of the four-gene panel in TCGA cohort. **b** Time-dependent ROC analysis, risk score, heatmap of mRNA expression, and Kaplan–Meier curve of the four-gene panel in GSE42127 cohort

suggested that only the risk score calculated from the four-gene panel was independent prognostic factors for OS (HR = 1.270, $p < 0.001$). The consistent result was obtained in the patients from GSE42127 (HR = 1.204, $p = 0.018$). The results were presented in Fig. 5.

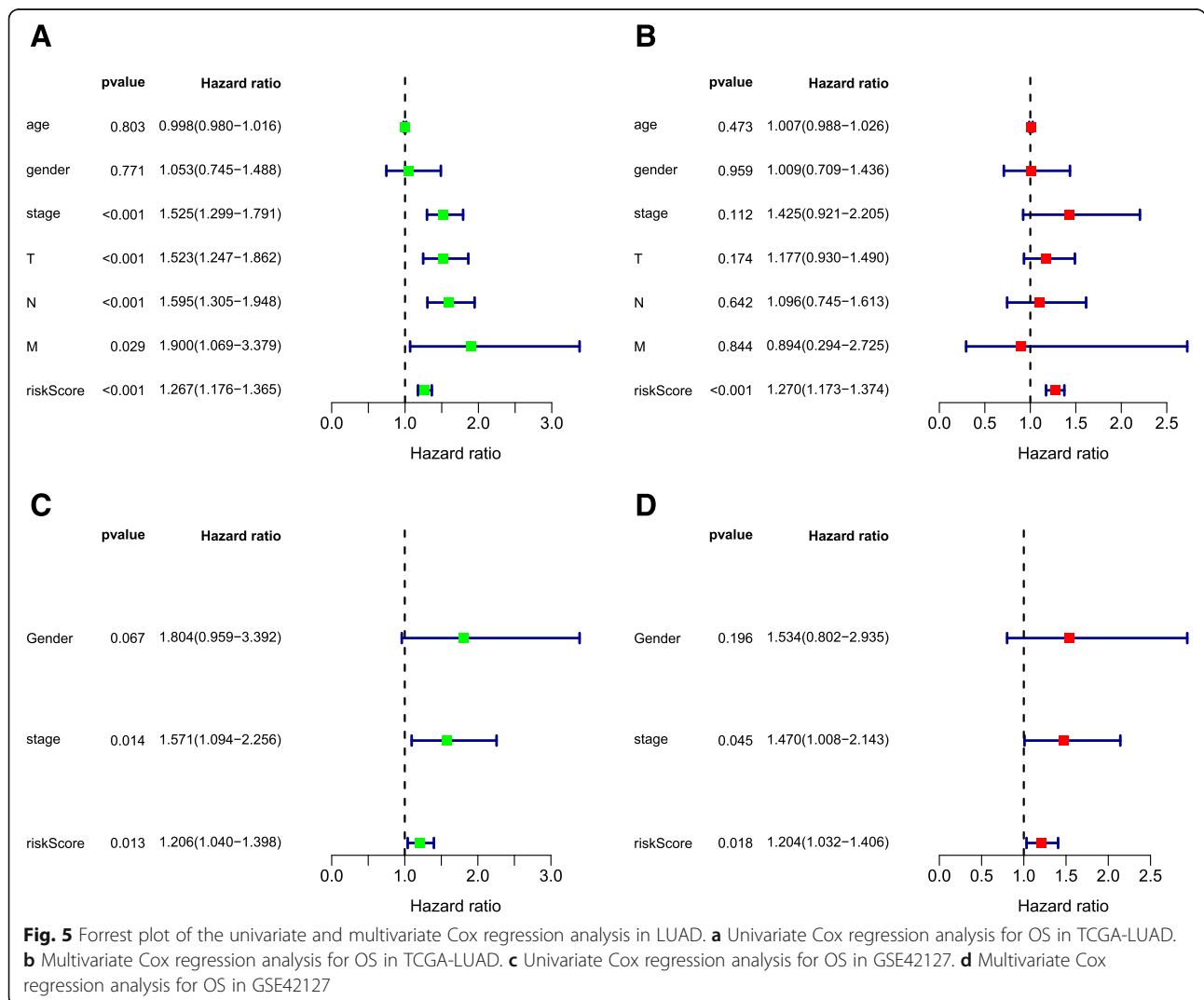
Establishment of predictive nomogram

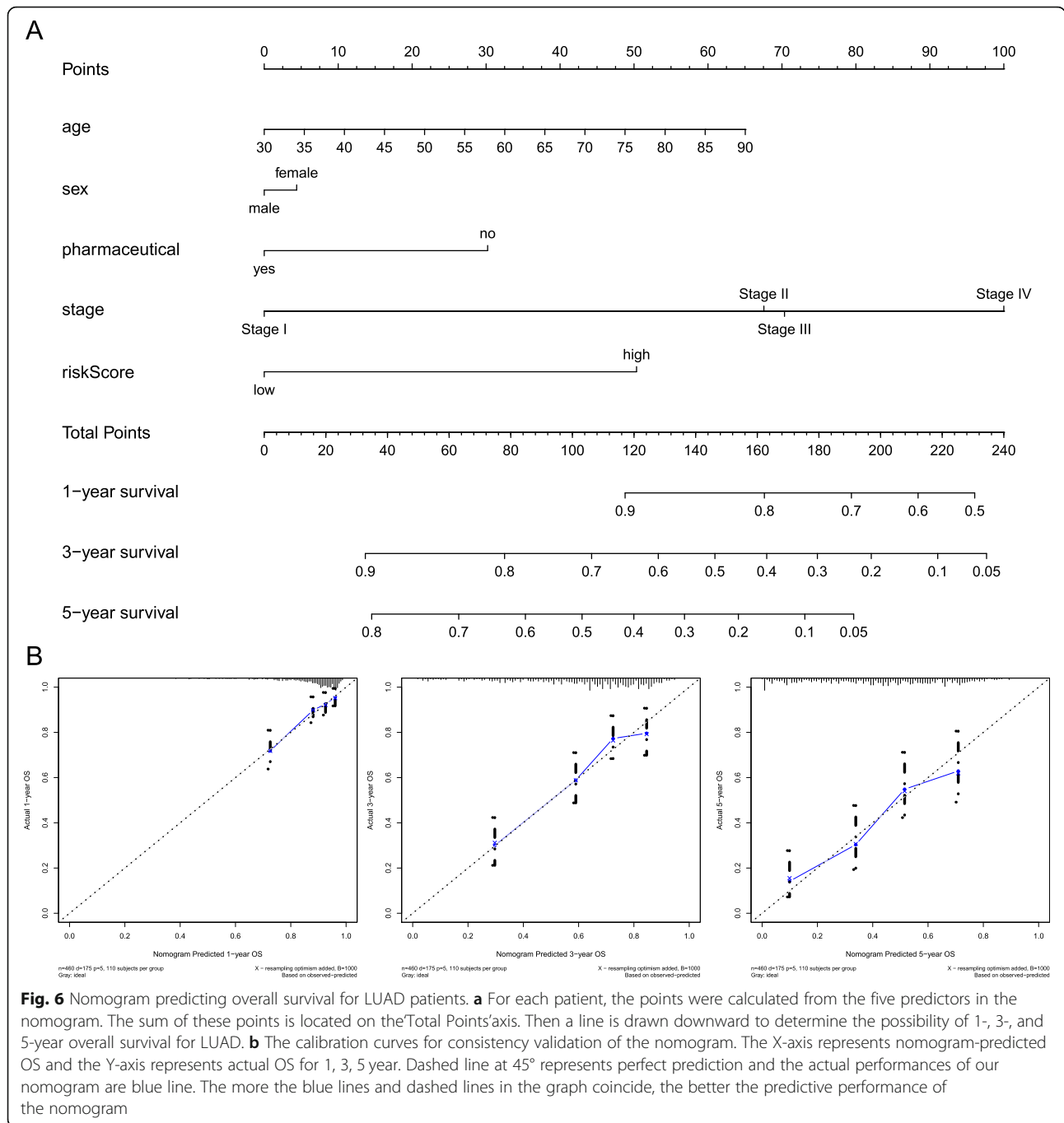
We established a nomogram to predict 1-year, 3-year, and 5-year OS in 460 patients with complete clinical information from the TCGA-LUAD database using five factors including risk score, age, sex, pharmaceutical, and pathologic stage (Fig. 6a). The C-index for the nomogram model was 0.710 (95% CI 0.624–0.796). Calibration curve showed that the nomogram had the superior prediction efficiency (Fig. 6b). These results indicated that the nomogram might be to

serve as a prognostic model used for clinical management of LUAD patients.

The genetic alteration, expression and survival analysis of the four genes

We explored the genetic alteration of the four genes by using the mutation data obtained from the cBioPortal for Cancer Genomics (<https://www.cbioportal.org/>) [15, 16]. In this database, 9 % (21/230) of patients showed genetic alterations in the four genes. Missense mutation, amplification and deep deletion were common genetic alteration (Fig. 7a). We further validated the expression of the four genes using the lung adenocarcinoma dataset (GSE75037 [17]) from the GEO database, and the results were consistent with the analysis of the TCGA-LUAD dataset (Fig. 7b). Kaplan-Meier survival curves indicated that high expression of DKK1, LDHA, and MELTF and



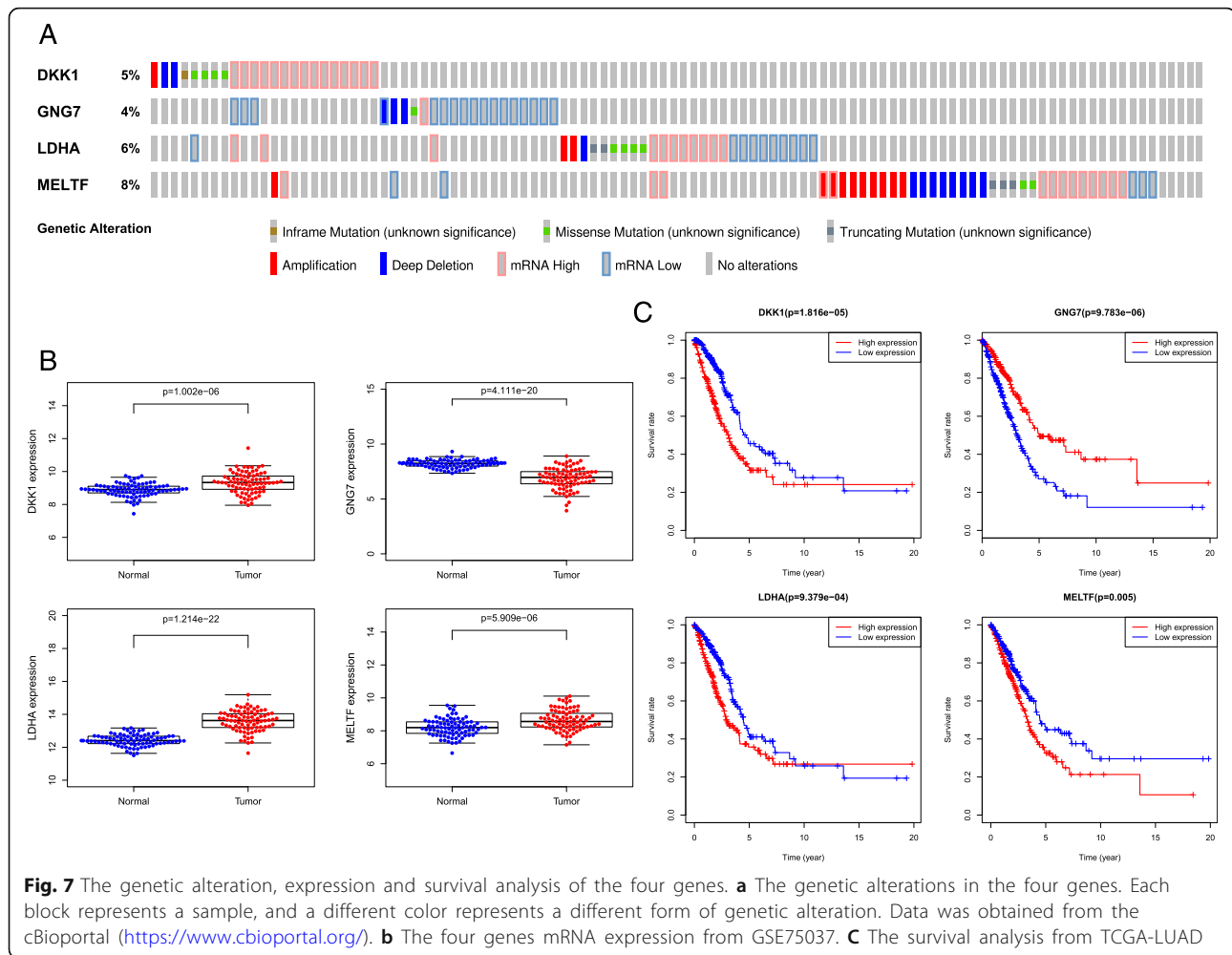


low expression of GNG7 were associated with a poor OS for LUAD (Fig. 7c).

Predictive value of the four-gene panel for patients with EGFR, KRAS and TP53 mutation

To explore the predictive value of the four-gene panel for patients with EGFR, KRAS and TP53 mutation, we performed a combined analysis of gene mutation and transcription data. A heatmap of mutations in TP53, KRAS, EGFR and the four genes was shown in Fig. 8.

The results showed that mRNA expression differences of GNG7 and MTF1 (MELTF) were only observed in TP53 mutant and wild-type patients (Fig. 9). We further analyzed the predictive value of the four-gene panel for OS in LUAD patients with EGFR, KRAS and TP53 mutations, respectively. The results showed that high-risk group had a worse OS in LUAD patients with TP53 mutations and KRAS mutations ($p < 0.05$). The AUC of this four-gene panel as a prognostic indicator was 0.718 and 0.793 in LUAD patients with TP53



and KRAS mutation, respectively. However, in patients with EGFR mutations, there was no significant difference in OS between the high-risk and low-risk groups (Fig. 10).

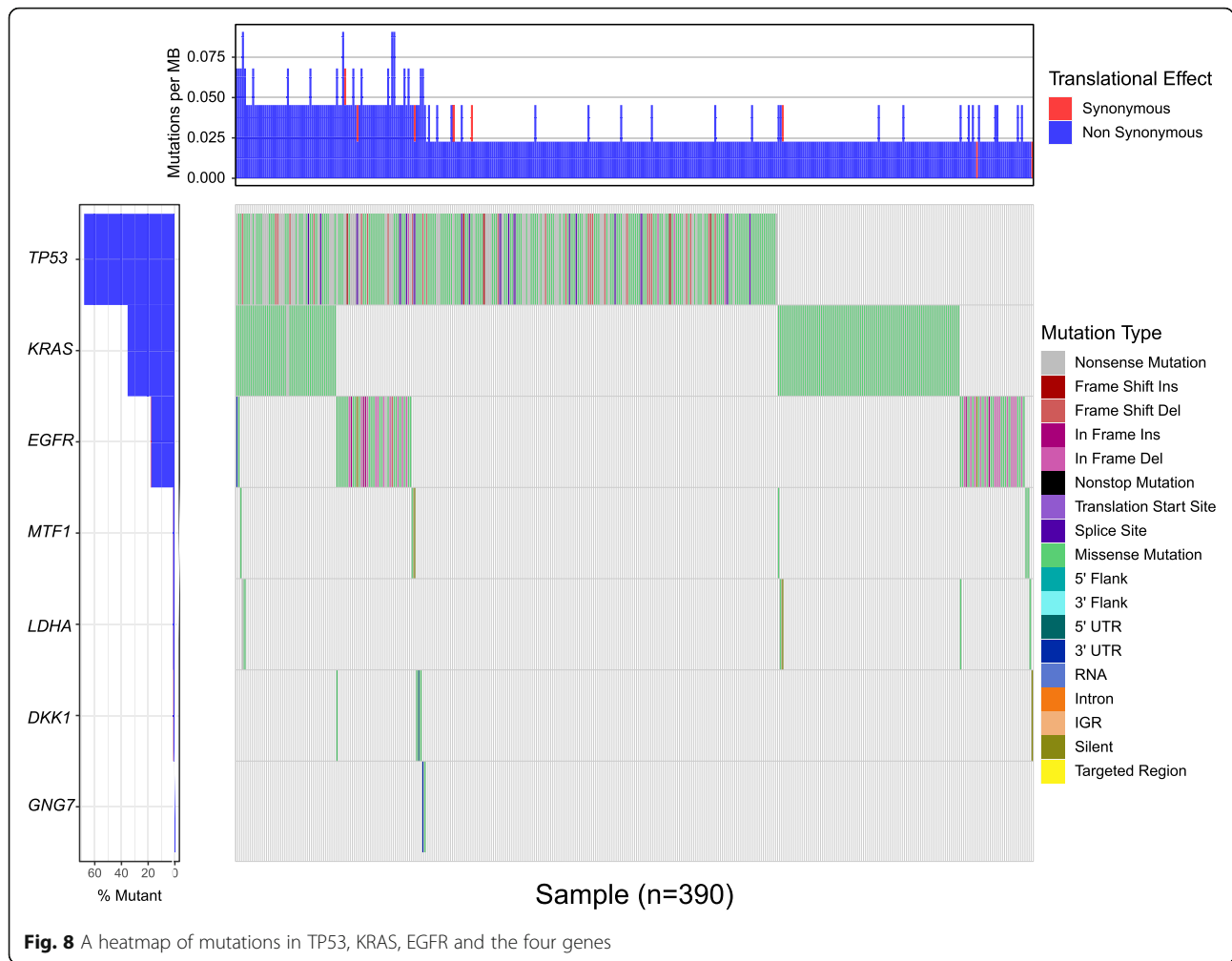
Gene set enrichment analyses

GSEA analysis was used to identify signaling pathways enriched in low and high expression of the four genes, respectively. The results revealed that genes involved in cell cycle, ubiquitin mediated proteolysis, RNA degradation, aminoacyl tRNA biosynthesis, DNA replication, proteasome, small cell lung cancer, and P53 signaling pathway were enriched in GNG7 low expression group. In the high-expressed group of DKK1, KEGG pathways including cell cycle, RNA degradation, spliceosome, proteasome, DNA replication, P53 signaling pathway and so on were enriched. In the high-expressed group of LDHA, KEGG pathways including proteasome, RNA degradation, spliceosome, DNA replication, RNA polymerase, cell cycle, P53 signaling pathway and so on were enriched. In

the high-expressed group of MELTF, the enriched KEGG pathways were mainly focused on bladder cancer, proteasome, DNA replication, base excision repair, pyrimidine metabolism. These results suggested that the absence of GNG7 expression and the increase of DKK1, LDHA and MELTF expression may be significantly related to the metabolism of genetic material, especially in the regulation of cell cycle pathway (Fig. 11).

DNA methylation level and mRNA expression of GNG7

We further explored the relationship between DNA methylation level and mRNA expression of GNG7 using MethHC database. The results showed that the DNA methylation levels of GNG7 were significantly higher in 18 kinds of cancerous tissues than adjacent noncancerous tissues (Fig. 12). Furthermore, methylation level of the promoter and CpG Island region was negatively correlated with mRNA expression of GNG7 (Fig. 13).



Discussion

LUAD remains a serious threat to human health worldwide. Despite the fact that molecular targeted therapy and immunotherapy have made great progress, the OS of LUAD is still poor as the lack of accurate early diagnosis and prognosis markers. Owing to tumor heterogeneity, traditional clinical parameters such as TNM stage cannot meet the requirements of accuracy and individuation for prognostic prediction. Identification of accurate prognostic biomarkers and novel and effective therapeutic targets remains particularly urgent. And the combination of multiple prognostic genes seems to be more valuable and promising. Prognostic prediction models based on multiple genes combination have been established and validated in various cancers [7, 18, 19].

In the present study, we established a four-gene panel (including DKK1, GNG7, LDHA, and MELTF) as a prognostic prediction model for LUAD. Each patient from TCGA-LUAD obtained a risk score based on this model, and the risk score was an independent prognostic

indicator of LUAD. Besides, the patients in high-risk score group shown poorer OS compared with patients in the low-risk score group. The consistent result was achieved in another independent cohort from the GEO database (GSE42127). The ROC curve demonstrated that the predictive performance of the risk score model as a prognostic indicator was superior both in the TCGA-LUAD cohort and in the GSE42127 cohort, compared with other clinical parameters. Nomogram combining risk score with other clinical parameters may be to serve as a prediction model used for clinical monitoring for OS in LUAD patients. All these results suggested that the prediction model based on the four-gene panel could be an effective and promising prognostic indicator for OS in LUAD patients.

DKK1, also named as DKK-1(dickkopf WNT signaling pathway inhibitor 1), has been proved to be differential expression in various tumors and participate in the regulation of growth, invasion, angiogenesis and metastasis of tumor [20–23]. In NSCLC, DKK1 be thought to be involved in tumor cell migration,

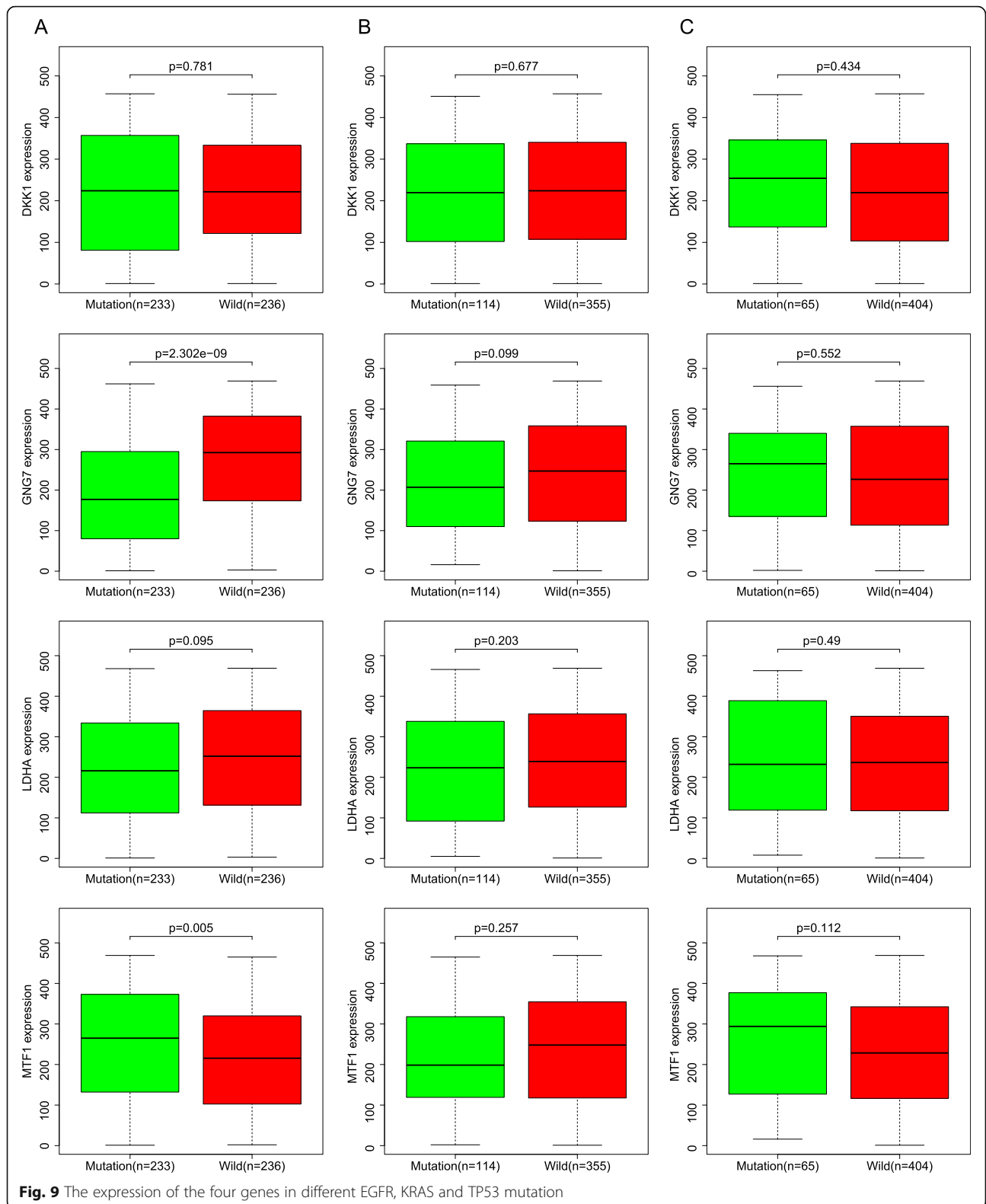


Fig. 9 The expression of the four genes in different EGFR, KRAS and TP53 mutation

invasion, and EMT processes, and could be used as an effective diagnostic and prognostic indicator and a potential therapeutic target [24–26]. LDHA (lactate

dehydrogenase A), a crucial enzyme of energy metabolism, is elevated in various cancers compared with normal tissues. Previous studies showed that LDHA could promote

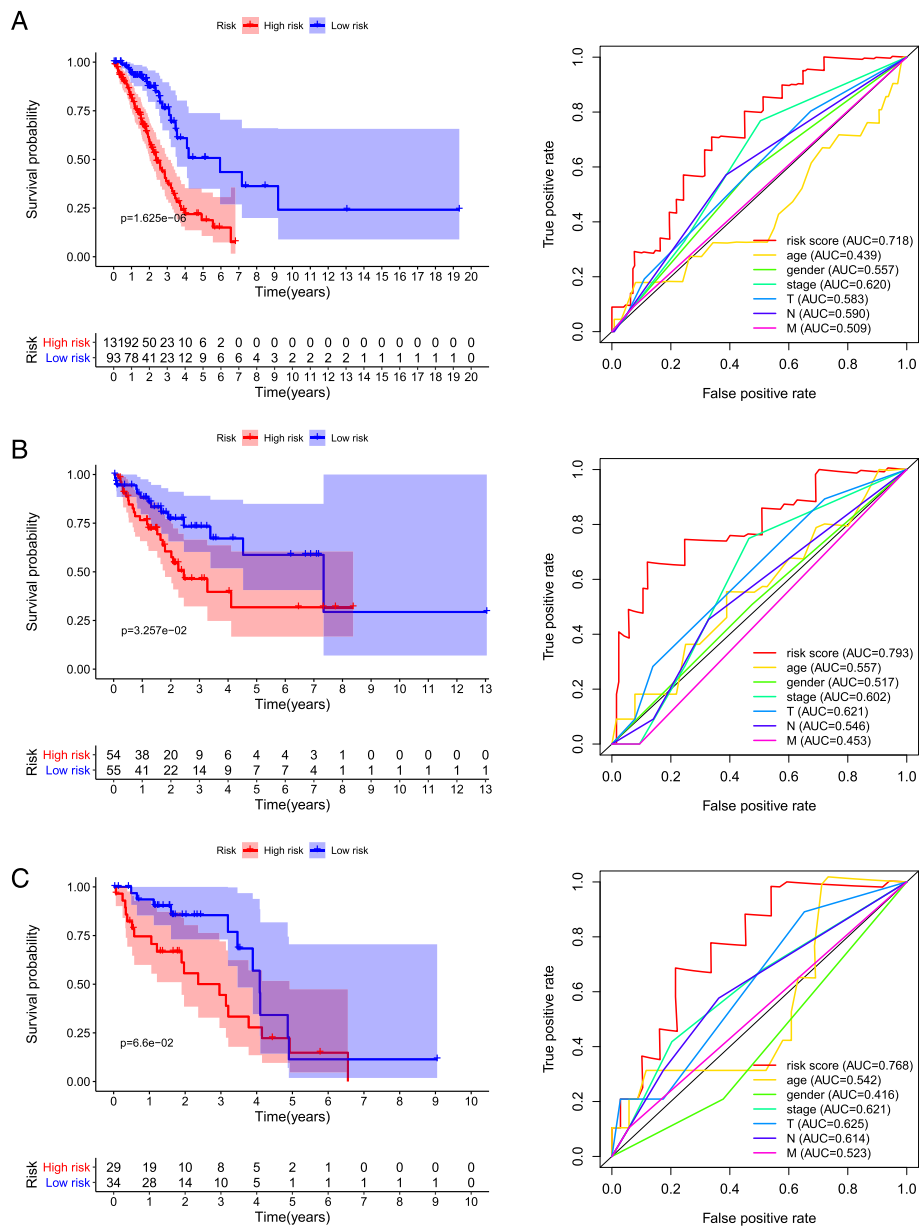
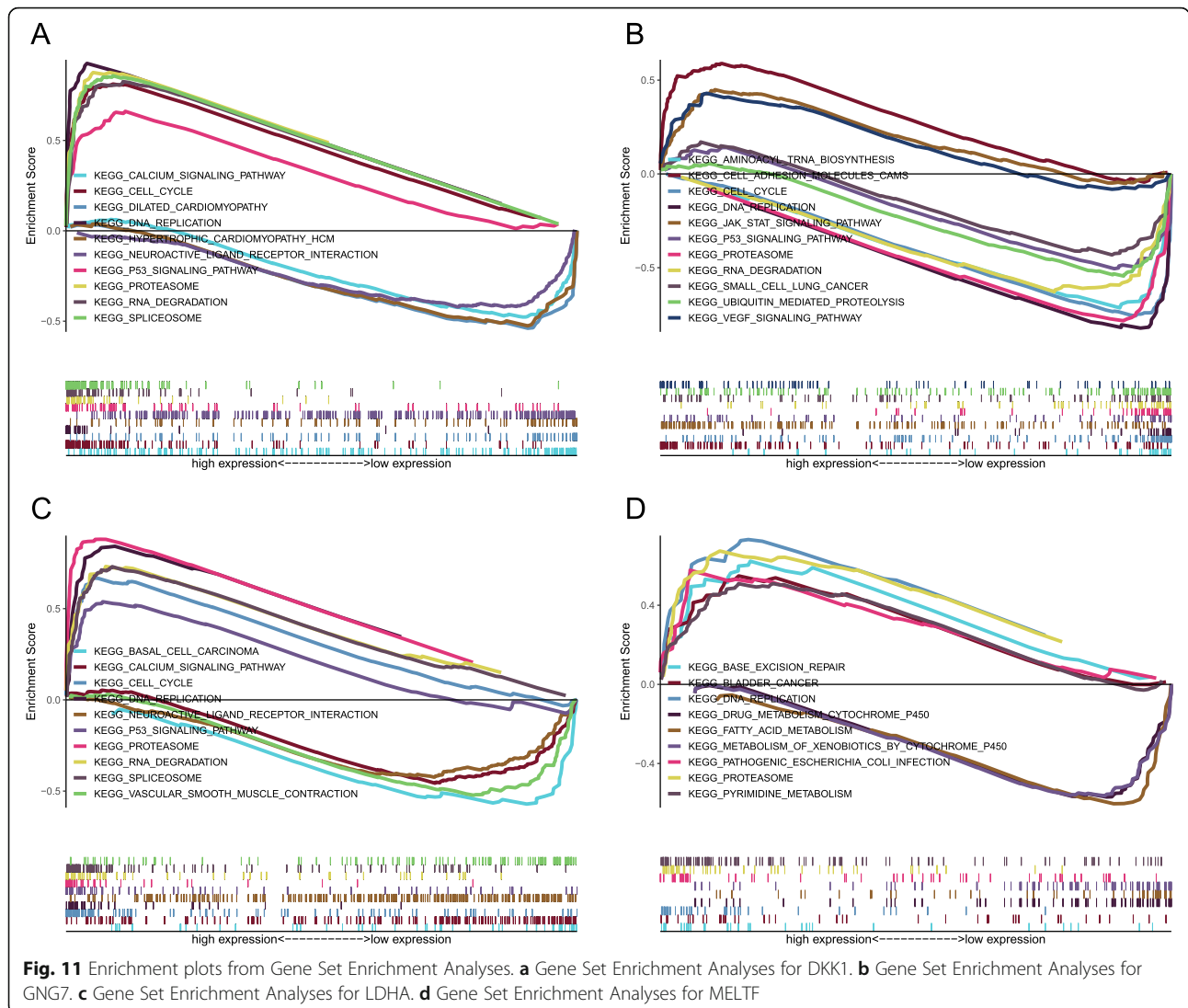


Fig. 10 Predictive value of the four-gene panel for patients with EGFR, KRAS and TP53 mutation. **a** Kaplan–Meier curve and Time-dependent ROC curve for patients with TP53 mutation. **b** Kaplan–Meier curve and Time-dependent ROC curve for patients with KRAS mutation. **c** Kaplan–Meier curve and Time-dependent ROC curve for patients with EGFR mutation

tumor cells proliferation, invasion, migration, tumor progression, and metastasis, and might be a potential therapeutic target [27–31]. MELTF, also known as MTF (Melanotransferrin) or MTF1 (metal regulatory transcription factor 1), as an iron (Fe) binding transferrin homolog, is mainly expressed in melanoma and is low expression in normal tissues. Previous studies indicated that MTF plays a key role in cell invasion and migration [32, 33]. Subsequent studies indicated that it could promote carcinoma cell invasion,

migration, proliferation, and EMT progression and be an attractive target [34–38].

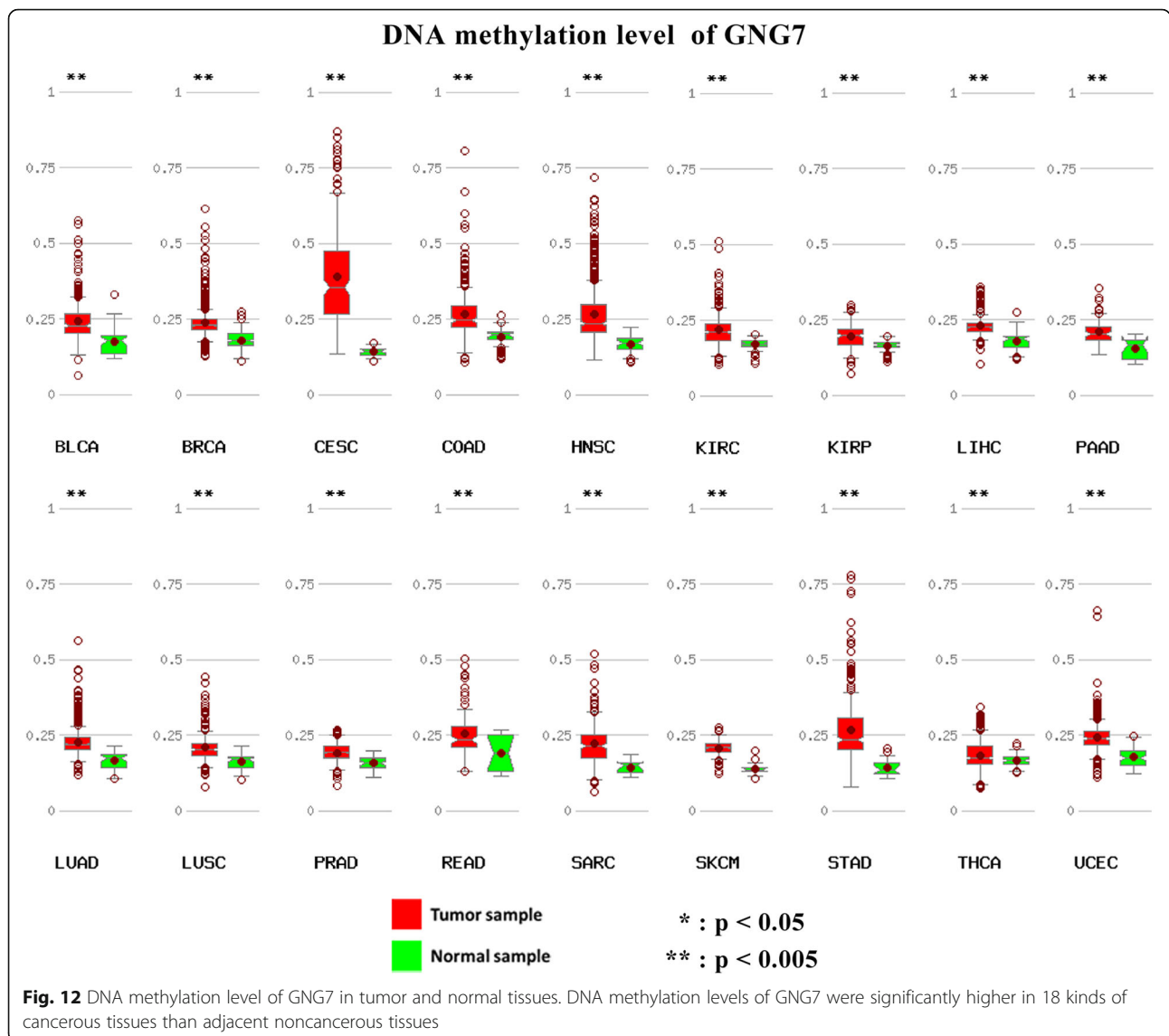
GNG7 (G protein subunit gamma 7), a novel possible tumor suppressor gene, is proved to be down-regulated in various carcinoma, including head and neck squamous cell carcinoma, clear cell carcinoma of kidney, pancreatic cancer, oesophageal cancer, lung adenocarcinoma [39–43]. However, the mechanism of its role in tumorigenesis and progression is still little known. We further validated the expression of the four genes using



the lung adenocarcinoma dataset GSE75037 dataset, and the results were consistent with the analysis of the TCGA-LUAD dataset. Kaplan-Meier survival curves indicated that high expression of DKK1, LDHA, and MELTF and low expression of GNG7 were associated with a poor OS for LUAD.

We further explored the predictive value of the four-gene panel for patients with EGFR, KRAS and TP53 mutation. The results showed that the expression of GNG7 is lower in TP53 mutant than wild-type patients, but expression of MELTF was the reverse. It is suggested that TP53 may play an opposite role in the expression regulation of the two genes. In addition, the four-gene panel was still excellent predictors for OS in LUAD patients with TP53 and KRAS mutations. It is suggested that the four-gene panel have useful predictive value and are not affected by mutations in these key genes.

The results of GSEA suggested that the absence of GNG7 expression and the increase of DKK1, LDHA and MELTF expression may be significantly related to the metabolism of genetic material, especially in the regulation of cell cycle pathway. This provides a sound theoretical basis for the future design of targeted therapy drugs for these 4 genes from the perspective of genetic material metabolism. DNA methylation is part of the common mechanisms of regulating genes expression. Our results showed that DNA methylation levels of the GNG7 were significantly higher in multiple tumors than in normal tissues. Furthermore, methylation level of the promoter and CpG Island region was negatively correlated with mRNA expression of GNG7. It indicated DNA methylation of GNG7 may involves in regulation of its expression.

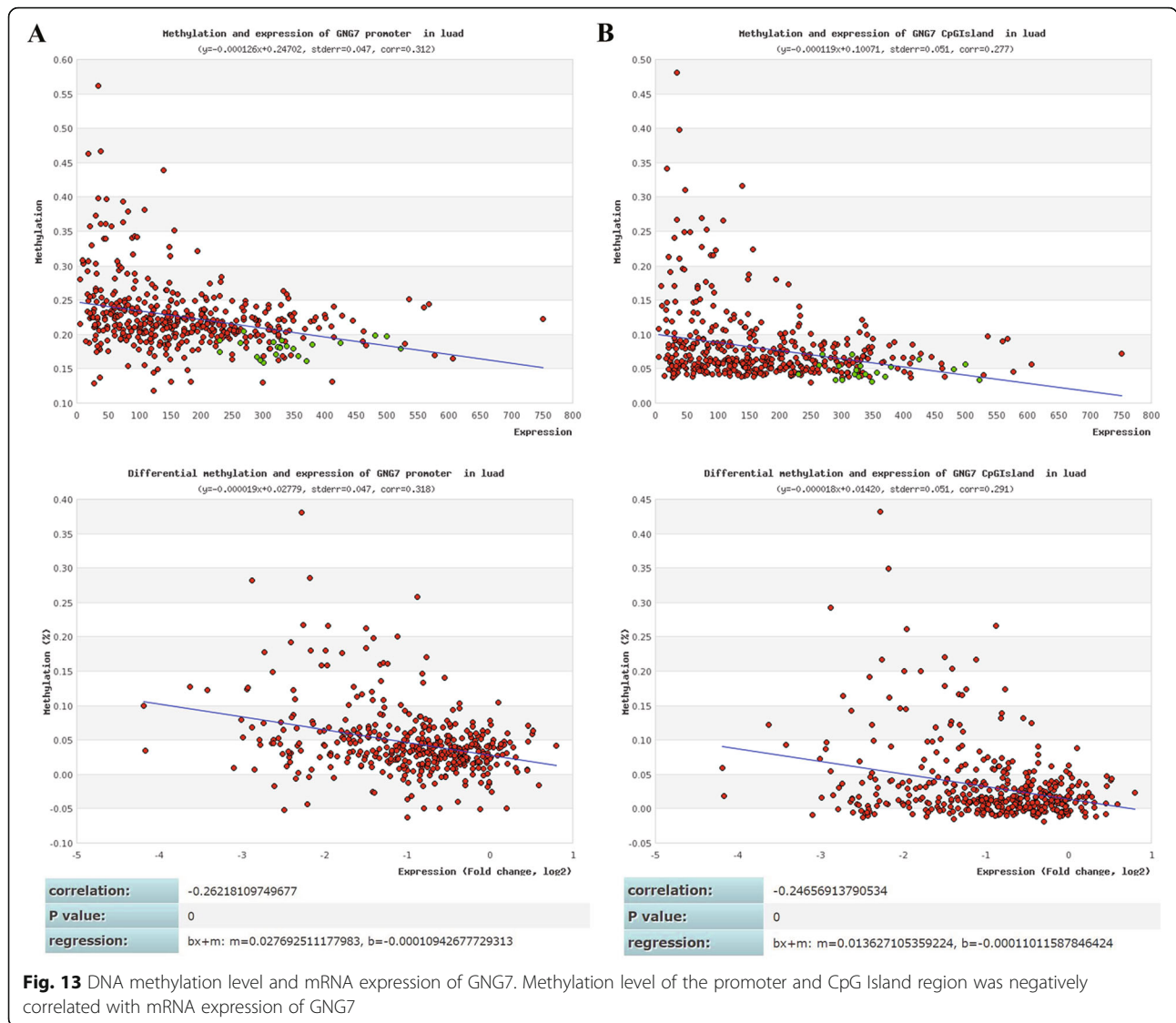


Overall, our study established an accurate and effective four-gene panel prognostic model for OS in LUAD patients. Risk scores based on this four-gene panel can be used to determine the OS of LUAD patients. Nomogram combining our signature with clinical parameters like pharmaceutical, age, TNM stage can be utilized to predict 1-year, 3-year, and 5-year survival in LUAD patients. Therefore, it will be useful for prognosis and follow-up monitoring of LUAD patients and reducing the extra cost for molecular diagnosis such as whole-genome sequencing. Besides, as a possible novel tumor suppressor gene, the elucidating mechanism of GNG7 in tumor genesis and progression will deepen our understanding of carcinomas including lung cancer and have great theoretical and scientific significance. However, it should be noted

that there are still some limitations to our study. Firstly, the data in our study mainly came from TCGA and GEO databases, and it was necessary to further verify the expression and prognostic value of the four genes at mRNA and protein level in a large independent clinical cohort. Secondly, the nomogram requires further external calibration and validation to improve predictive effectivity and accuracy. Thirdly, the potential biological mechanisms of the four genes in LUAD need to be further illuminated using functional studies.

Conclusions

Our study proposed a novel four-gene panel and nomogram to predict the OS for patients with LUAD, which may contribute to predicting prognosis accurately and



making clinical decisions of individual therapy for LUAD patients. The four genes may be significantly related to the metabolism of genetic material, especially in the regulation of cell cycle pathway. This provides a reliable theoretical basis for the future design of targeted therapy drugs for these 4 genes from the perspective of genetic material metabolism.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12885-020-07657-9>.

- Additional file 1.** List of differentially expressed genes.
- Additional file 2: Figure S1.** Heatmap of differentially expressed genes.
- Additional file 3: Figure S2.** Volcano plot of differentially expressed genes.

Additional file 4. List of differentially expressed genes associated with OS for LUAD.

Additional file 5: Figure S3. LASSO profiles of the 523 prognostic genes in LUAD. (A) LASSO coefficient profiles of the 523 prognostic genes in LUAD. (B) Lasso deviance profiles of the 523 prognostic genes in LUAD.

Abbreviations

OS: Overall survival; LUAD: Lung adenocarcinoma; TCGA: The Cancer Genome Atlas; LASSO: Least Absolute Shrinkage and Selection Operator; ROC: Receiver operating characteristic; GEO: Gene Expression Omnibus; GSEA: Gene set enrichment analysis; TCGA-LIHC: The Cancer Genome Atlas-Liver Hepatocellular Carcinoma Dataset; TCGA-LUAD: The Cancer Genome Atlas-Lung Adenocarcinoma Dataset; DEGs: Differential expression genes; C-index: The concordance index; MethHC: A database of DNA Methylation and gene expression in Human Cancer; FDR: False discovery rate; DDK1: Dickkopf WNT signaling pathway inhibitor 1; GNG7: G protein subunit gamma 7; LDHA: Lactate dehydrogenase A; MELTF: Melanotransferrin; MTF1: Metal regulatory transcription factor 1; AUC: Area under the ROC curve

Acknowledgements

Not applicable.

Authors' contributions

XZ and CL designed this study; CL performed most data collection and analysis; QL, DZ, JL helped to perform analysis and collected the data; CL and XZ wrote and revised the manuscript. All authors read and approved the final manuscript.

Funding

This project was supported by grants from National Natural Science Foundation of China (81660018). The funder is the corresponding author of the study and is responsible for the overall design and planning of this study.

Availability of data and materials

The datasets supporting the conclusions of this article are available in the TCGA-GDC (<https://portal.gdc.cancer.gov/>) repository.

Ethics approval and consent to participate

Any repository data used in this study are open access and do not require any permissions. Ethics approval and consent to participate are not applicable for them.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 3 March 2020 Accepted: 17 November 2020

Published online: 07 December 2020

References

- Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2018;68(6):394–424.
- Travis WD. Pathology of lung cancer. *Clin Chest Med*. 2011;32(4):669–92.
- Travis WD, Brambilla E, Noguchi M, Nicholson AG, Geisinger KR, Yatabe Y, Beer DG, Powell CA, Riely GJ, Van Schil PE, et al. International association for the study of lung cancer/american thoracic society/european respiratory society international multidisciplinary classification of lung adenocarcinoma. *J Thorac Oncol*. 2011;6(2):244–85.
- Lortet-Tieulent J, Soerjomataram I, Ferlay J, Rutherford M, Weiderpass E, Bray F. International trends in lung cancer incidence by histological subtype: adenocarcinoma stabilizing in men but still increasing in women. *Lung Cancer*. 2014;84(1):13–22.
- Toyoda Y, Nakayama T, Ioka A, Tsukuma H. Trends in lung cancer incidence by histological type in Osaka, Japan. *JPN J CLIN ONCOL*. 2008;38(8):534–9.
- Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. *CA Cancer J Clin*. 2016;66(1):7–30.
- Liu GM, Zeng HD, Zhang CY, Xu JW. Identification of a six-gene signature predicting overall survival for hepatocellular carcinoma. *Cancer Cell Int*. 2019;19:138.
- Diboun I, Wernisch L, Orengo CA, Koltzenburg M. Microarray analysis after RNA amplification can detect pronounced differences in gene expression using limma. *BMC Genomics*. 2006;7:252.
- Tibshirani R. The lasso method for variable selection in the cox model. *Stat Med*. 1997;16(4):385–95.
- Heagerty PJ, Lumley T, Pepe MS. Time-dependent ROC curves for censored survival data and a diagnostic marker. *BIOMETRICS*. 2000;56(2):337–44.
- Tang H, Xiao G, Behrens C, Schiller J, Allen J, Chow CW, Suraokar M, Corvalan A, White M, Wistuba I, et al. A 12-gene set predicts survival benefits from adjuvant chemotherapy in non-small cell lung cancer patients. *Clin Cancer Res*. 2013;19(6):1577–86.
- Iasonos A, Schrag D, Raj GV, Panageas KS. How to build and interpret a nomogram for cancer prognosis. *J Clin Oncol*. 2008;26(8):1364–70.
- Huang WY, Hsu SD, Huang HY, Sun YM, Chou CH, Weng SL, Huang H-D. MethHC: a database of DNA methylation and gene expression in human cancer. *Nucleic Acids Res*. 2015;43(Database issue):D856–61.
- Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102(43):15545–50.
- Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal*. 2013;6(269):p11.
- Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *CANCER DISCOV*. 2012;2(5):401–4.
- Girard L, Rodriguez-Canales J, Behrens C, Thompson DM, Botros IW, Tang H, Xie Y, Rekhman N, Travis WD, Wistuba I, et al. An expression signature as an aid to the histologic classification of non-small cell lung cancer. *Clin Cancer Res*. 2016;22(19):4880–9.
- Zuo S, Zhang X, Wang L. A RNA sequencing-based six-gene signature for survival prediction in patients with glioblastoma. *Sci Rep*. 2019;9(1):2615.
- Wang Z, Wang Z, Niu X, Liu J, Wang Z, Chen L, Qin B. Identification of seven-gene signature for prediction of lung squamous cell carcinoma. *Oncotargets Ther*. 2019;12:5979–88.
- Lyros O, Lamprecht AK, Nie L, Thieme R, Götzel K, Gasparri M, Haasler G, Rafiee P, Shaker R, Gockel I. Dickkopf-1 (DKK1) promotes tumor growth via Akt-phosphorylation and independently of Wnt-axis in Barrett's associated esophageal adenocarcinoma. *Am J Cancer Res*. 2019;9(2):330–46.
- Zhuang X, Zhang H, Li X, Li X, Cong M, Peng F, Yu J, Zhang X, Yang Q, Hu G. Differential effects on lung and bone metastasis of breast cancer by Wnt signalling inhibitor DKK1. *Nat Cell Biol*. 2017;19(10):1274–85.
- Park H, Jung HY, Choi HJ, Kim DY, Yoo JY, Yun CO, Min JK, Kim YM, Kwon YG. Distinct roles of DKK1 and DKK2 in tumor angiogenesis. *ANGIOGENESIS*. 2014;17(1):221–34.
- Chen L, Li M, Li Q, Wang CJ, Xie SQ. DKK1 promotes hepatocellular carcinoma cell migration and invasion through beta-catenin/MMP7 signaling pathway. *Mol Cancer*. 2013;12:157.
- Zhang P, Li S, Lv C, Si J, Xiong Y, Ding L, Ma Y, Yang Y. BPI-9016M, a c-met inhibitor, suppresses tumor cell growth, migration and invasion of lung adenocarcinoma via miR203-DKK1. *THERANOSTICS*. 2018;8(21):5890–902.
- Yamabuki T, Takano A, Hayama S, Ishikawa N, Kato T, Miyamoto M, Ito T, Ito H, Miyagi Y, Nakayama H, et al. Dickkopf-1 as a novel serologic and prognostic biomarker for lung and esophageal carcinomas. *Cancer Res*. 2007;67(6):2517–25.
- Zhang J, Zhang X, Zhao X, Jiang M, Gu M, Wang Z, Yue W. DKK1 promotes migration and invasion of non-small cell lung cancer via beta-catenin signaling pathway. *Tumour Biol*. 2017;39(7):1393385844.
- Sheng SL, Liu JJ, Dai YH, Sun XG, Xiong XP, Huang G. Knockdown of lactate dehydrogenase suppresses tumor growth and metastasis of human hepatocellular carcinoma. *FEBS J*. 2012;279(20):3898–910.
- Hou XM, Yuan SQ, Zhao D, Liu XJ, Wu XA. LDHA promotes malignant behavior via activation of epithelial-to-mesenchymal transition in lung adenocarcinoma. *Biosci Rep*. 2019;39(1):BSR20181476.
- Li L, Kang L, Zhao W, Feng Y, Liu W, Wang T, Mai H, Huang J, Chen S, Liang Y, et al. miR-30a-5p suppresses breast tumor growth and metastasis through inhibition of LDHA-mediated Warburg effect. *CANCER LETT*. 2017;400:89–98.
- Jin L, Chun J, Pan C, Alesi GN, Li D, Magliocca KR, Kang Y, Chen ZG, Shin DM, Khuri FR, et al. Phosphorylation-mediated activation of LDHA promotes cancer cell invasion and tumour metastasis. *ONCOGENE*. 2017;36(27):3797–806.
- Xie H, Hanai J, Ren JG, Kats L, Burgess K, Bhargava P, Signoretti S, Billiard J, Duffy KJ, Grant A, et al. Targeting lactate dehydrogenase—a inhibits tumorigenesis and tumor progression in mouse models of lung cancer and impacts tumor-initiating cells. *Cell Metab*. 2014;19(5):795–809.
- Demeule M, Bertrand Y, Michaud-Levesque J, Jodoin J, Rolland Y, Gabathuler R, Béliveau R. Regulation of plasminogen activation: a role for melanotransferrin (p97) in cell migration. *BLOOD*. 2003;102(5):1723–31.
- Michaud-Levesque J, Demeule M, Béliveau R. Stimulation of cell surface plasminogen activation by membrane-bound melanotransferrin: a key phenomenon for cell invasion. *Exp Cell Res*. 2005;308(2):479–90.
- Suryo RY, Dunn LL, Richardson DR. Identification of distinct changes in gene expression after modulation of melanoma tumor antigen p97 (melanotransferrin) in multiple models in vitro and in vivo. *CARCINOGENESIS*. 2007;28(10):2172–83.

35. Rolland Y, Demeule M, Fenart L, Beliveau R. Inhibition of melanoma brain metastasis by targeting melanotransferrin at the cell surface. *Pigment Cell Melanoma Res.* 2009;22(1):86–98.
36. Ji L, Zhao G, Zhang P, Huo W, Dong P, Watari H, Jia L, Pfeffer LM, Yue J, Zheng J. Knockout of MTF1 inhibits the epithelial to Mesenchymal transition in ovarian Cancer cells. *J Cancer.* 2018;9(24):4578–85.
37. Sawaki K, Kanda M, Umeda S, Miwa T, Tanaka C, Kobayashi D, Hayashi M, Yamada S, Nakayama G, Omae K, et al. Level of Melanotransferrin in tissue and sera serves as a prognostic marker of gastric Cancer. *Anticancer Res.* 2019;39(11):6125–33.
38. Dunn LL, Sekyere EO, Suryo RY, Richardson DR. The function of melanotransferrin: a role in melanoma cell proliferation and tumorigenesis. *CARCINOGENESIS.* 2006;27(11):2157–69.
39. Ohta M, Mimori K, Fukuyoshi Y, Kita Y, Motoyama K, Yamashita K, Ishii H, Inoue H, Mori M. Clinical significance of the reduced expression of G protein gamma 7 (GNG7) in oesophageal cancer. *Br J Cancer.* 2008;98(2):410–7.
40. Gao LW, Wang GL. Comprehensive bioinformatics analysis identifies several potential diagnostic markers and potential roles of cyclin family members in lung adenocarcinoma. *Onco Targets Ther.* 2018;11:7407–15.
41. Xu S, Zhang H, Liu T, Chen Y, He D, Li L. G protein gamma subunit 7 loss contributes to progression of clear cell renal cell carcinoma. *J Cell Physiol.* 2019;234(11):20002–12.
42. Shibata K, Mori M, Tanaka S, Kitano S, Akiyoshi T. Identification and cloning of human G-protein gamma 7, down-regulated in pancreatic cancer. *Biochem Biophys Res Commun.* 1998;246(1):205–9.
43. Hartmann S, Szaumkessel M, Salaverria I, Simon R, Sauter G, Kiwerska K, Gawecki W, Bodnar M, Marszalek A, Richter J, et al. Loss of protein expression and recurrent DNA hypermethylation of the GNG7 gene in squamous cell carcinoma of the head and neck. *J Appl Genet.* 2012;53(2):167–74.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://www.biomedcentral.com/submissions)

