Gene expression-based clinical predictions in lung adenocarcinoma

Yanlu Xiong[1,*], Jie Lei[1,*], Jinbo Zhao[1,*], Yangbo Feng[1], Tianyun Qiao[1], Yongsheng Zhou[1], Tao Jiang[1,&], Yong Han[1,2]

[1]Department of Thoracic Surgery, Tangdu Hospital, Fourth Military Medical University, Xi'an, China
[2]Department of Thoracic Surgery, Air Force Medical Center, PLA, Beijing, China
*Equal contribution

Correspondence to: Tao Jiang, Yong Han; **email:** jiangtaochest@163.com; hanyong_td@163.com, https://orcid.org/0000-0002-5325-1722

ABSTRACT

Mining disease-related genes contributes momentously to handling lung adenocarcinoma (LUAD). But genetic complexity and tumor heterogeneity severely get in the way. Fortunately, new light has been shed by dramatic progress of bioinformatic technology in the past decades. In this research, we investigated relationships between gene expression and clinical features of LUAD via integrative bioinformatic analysis. First, we applied limma and DESeq2 packages to analyze differentially expressed genes (DEGs) of LUAD from GEO database and TCGA project (tumor tissues versus normal tissues), and acquired 180 down-regulated DEGs and 52 up-regulated DEGs. Then, we investigated genetic and biological assignment of theses DEGs by Bioconductor packages and STRING database. We found these DEGs were distributed dispersedly among chromosomes, enriched observably in extracellular matrix-related processes, and weighted hierarchically in interaction network. Finally, we established DEGs-based statistical models for evaluating TNM stage and survival status of LUAD. And these models (logistic regression models for TNM parameter and Cox regression models for survival probability) all possessed fine predictive efficacy (C-indexes: T, 0.740; N, 0.687; M, 0.823; overall survival, 0.678; progression-free survival, 0.611). In summary, we have successfully established gene expression-based models for assessing clinical characteristics of LUAD, which will assist its pathogenesis investigation and clinical intervention.

INTRODUCTION

Around the world, lung cancer possesses the most frequent new cases and deaths among cancers [1]. Lung adenocarcinoma (LUAD) accounts for a large portion of lung cancer, urgently calling for effective treatment [2, 3].

Cancer-related genes are well-deserved targets for understanding and treating cancer, and LUAD is no exception. Indeed, almost every revolutionary breakthrough in the battle with LUAD could not run without findings of cancer-related genes [4–6]. For example, cytotoxic chemical therapies chiefly aim at genes functioning critically in basic cellular activities, and molecular targeted therapies depend principally on tumor driver genes, while immune checkpoint genes are core targets for immune therapies [6, 7]. Furthermore, tumor clinical characteristics are made up of a series of biological processes, of which underpinnings attribute to gene function [8], that is to say, cancer-related genes could also help evaluate tumor clinicopathological parameters. However, huge quantities of genes and intensive heterogeneity of LUAD both seriously curb the way to find these key players [9].

Fortunately, remarkable progress in cancer genomics and bioinformatic technology endows us with possibilities to crack such hard nut [10]. On the one hand, more well-rounded and accurate genomic

profiling becomes easier to obtain [11, 12]. On the other hand, systematic and integrated analysis confers more rationality on understanding carcinogenesis, for single-gene investigation seems quite stretched facing such complex cancer pathogenesis [12, 13]. In this research, we used bioinformatic analysis to establish gene expression-based models for evaluating TNM parameters of LUAD and predicting its survival probability, which could provide insight into malignant etiology and handling methods.

RESULTS

Up-regulated and down-regulated DEGs were derived from LUAD transcriptome profiling

Cancer-related differentially expressed genes (DEGs) are highly correlated with tumor initiation and progression, which is quite conductive for evaluating clinical characteristics of malignancy. We acquired DEGs from three LUAD microarrays (GSE32863, GSE43458, GSE10072) using limma package (fold change >2 or fold change <0.5, Adjusted P-value <0.05) (Figure 1A–1C). Analogously, we used DESeq2 package to get DEGs of LUAD RNA-sequencing data from The Cancer Genome Atlas (TCGA) program (fold change >2 or fold change <0.5, Adjusted P-value <0.05) (Figure 1D). Then, we applied intersection analysis of these results and ultimately obtained 52 upregulated genes and 180 downregulated genes (tumor versus normal) (Figure 1E, 1F).

Genetic annotation, biological assignment and interaction function of DEGs

To further comprehend these DEGs, we focused on their elementary hallmarks. First, genetic mapping atlas were drawn to show chromosome locations of these
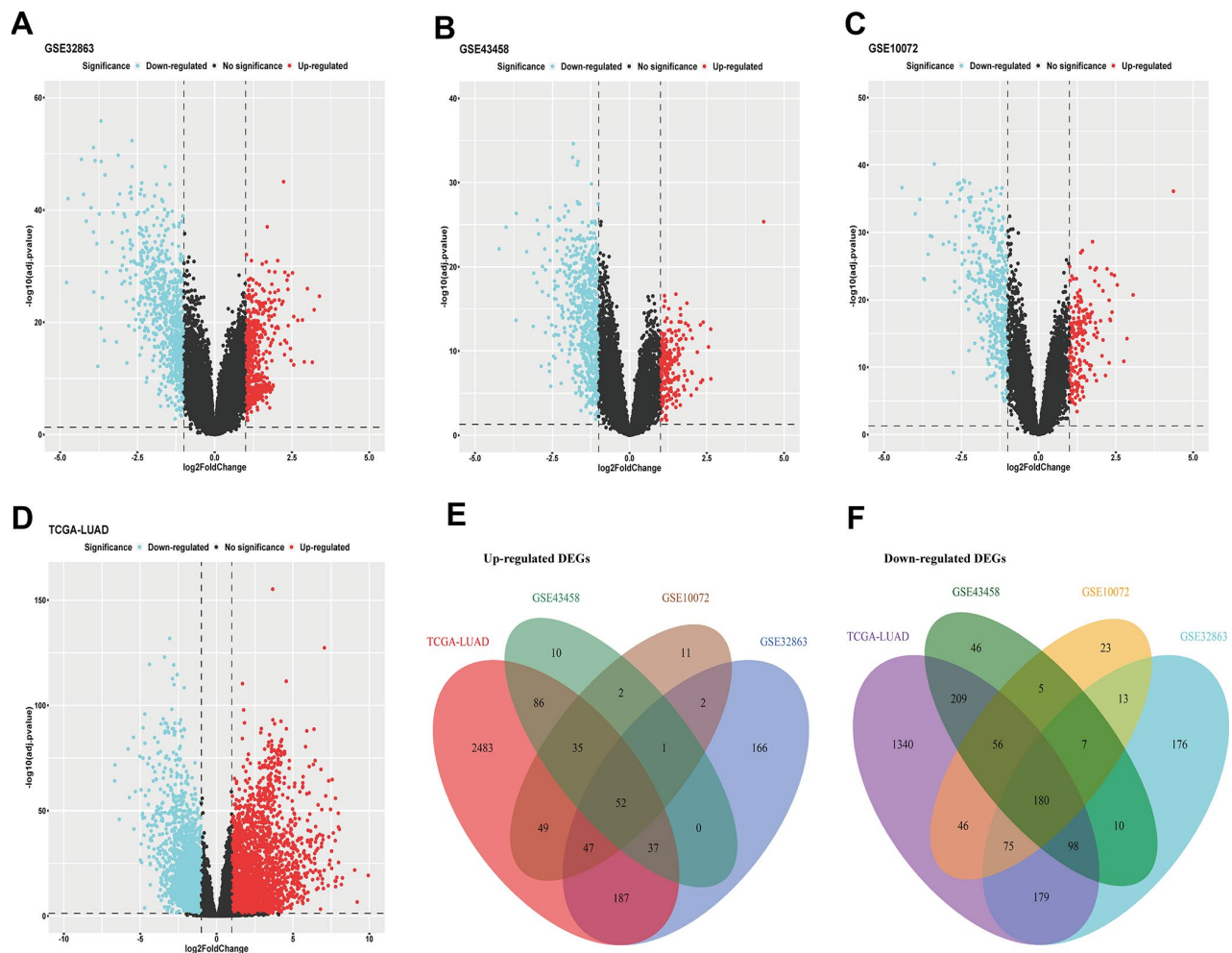

Figure 1. DEGs of LUAD. (**A–C**) DEGs acquired from GEO microarrays (GSE32863, GSE43458, GSE10072); (**D**) DEGs obtained from TCGA LUAD RNA-sequencing; (**E**) intersection of up-regulated DEGs; (**F**) intersection of down-regulated DEGs. DEGs, differentially expressed genes; GEO, Gene Expression Omnibus; TCGA, The Cancer Genome Atlas; LUAD, lung adenocarcinoma.

DEGs, and these DEGs had a scattered distribution in genomes (Figure 2A) (not all genes can be labeled in the figure, genomic information of the whole DEGs could be seen in Supplementary Table 1). Further, gene ontology (GO) analysis and pathway enrichment analysis both displayed significant enrichment of these DEGs in cell matrix-associated gene assemblies, which strongly indicated prominent weight of micro-environment upon carcinogenesis (Figure 2B, 2C) (Supplementary Table 2). Subsequently, we investigated interaction between DEGs in STRING database, and results suggested these DEGs had hierarchical function in malignant progression, demonstrating necessity for further filtrating (Figure 2D).

Estimating TNM parameters of LUAD by multiple gene analysis

TNM staging, an assessment for tumor growth, lymph node invasion and distant metastasis, constitutes momentous facets of LUAD clinical features. We tried to unearth relationships between DEGs and TNM parameters. We first transferred TNM records to two-category dimension (T: T3-4 for higher risk indicator and T1-2 for lower indicator; N: lymph node invasion happening or not; M: having distant metastasis or not), and normalized corresponding RNA-sequencing data by Z score. Then we selected preliminarily 11 T-related DEGs, 58 N-related DEGs and 21 M-related DEGs by univariate logistic regression analysis respectively ($P<0.05$) (Figure 3A–3C). However, potent interaction among these genes indicated confounding factors existed (Supplementary Figure 1–3). So we put these primary screening genes in multivariate logistic regression models, but all showed poor significance, calling for further modification (Supplementary Figure 4–6). After step regression by Akaike information criterion (AIC), we finally built optimized models for TNM parameters. As to T parameter, we chose four genes, one gene (GPC3) functioned as a protective factor, the other three genes (CAV1, LDLR, LIMCH1) functioned as hazard factors (C-index, 0.740; R^2, 0.173) (Figure 3D). The area under the curve (AUC) of T-related model was 0.740, indicating fine predictive effect
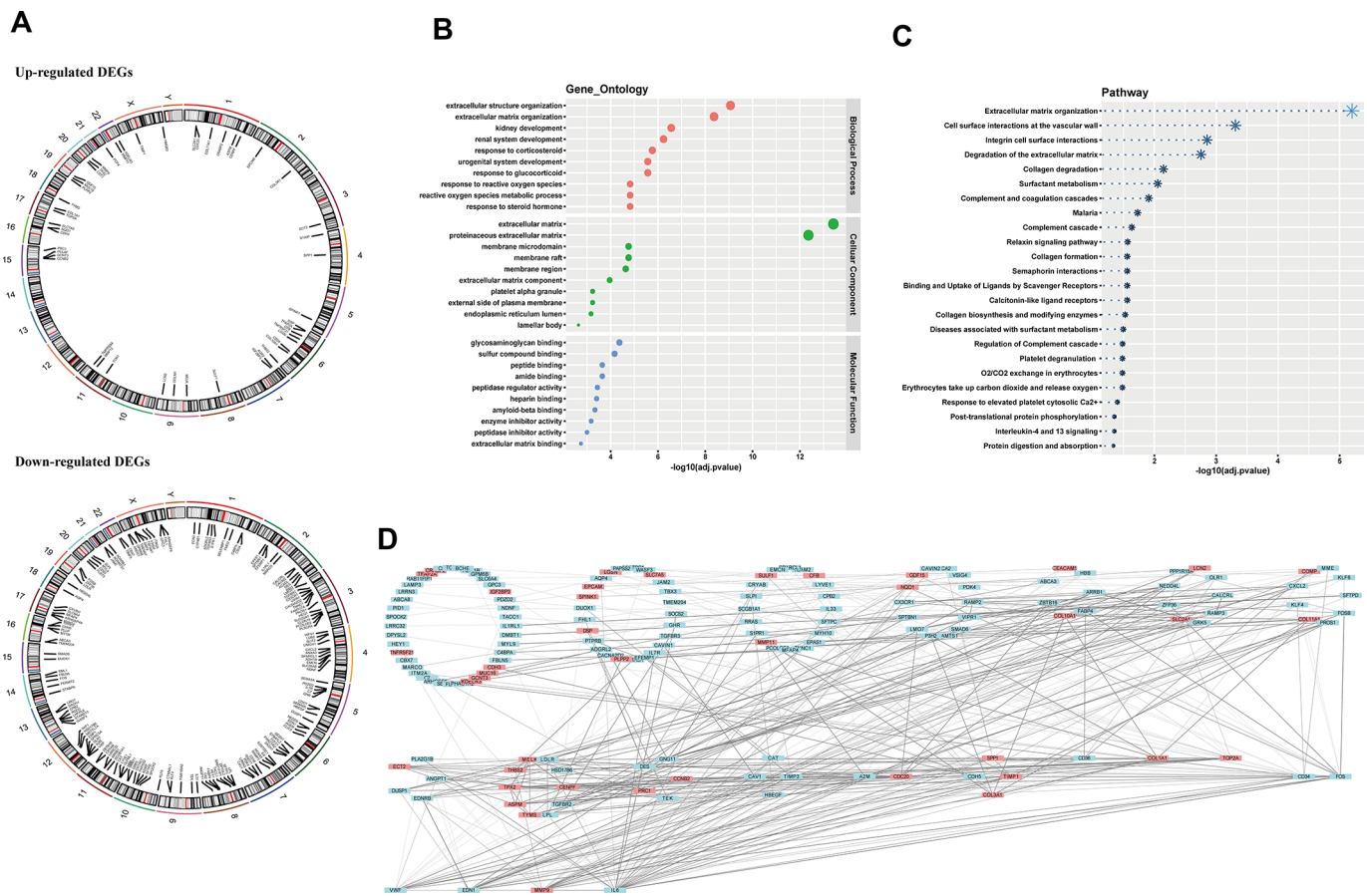

Figure 2. Chromosome location, biological annotation and interaction function. (**A**) genetic mapping of up-regulated and down-regulated DEGs; (**B**) gene ontology annotation of DEGs; (**C**) pathway enrichment of DEGs; (**D**) hierarchy of gene interaction. DEGs, differentially expressed genes.
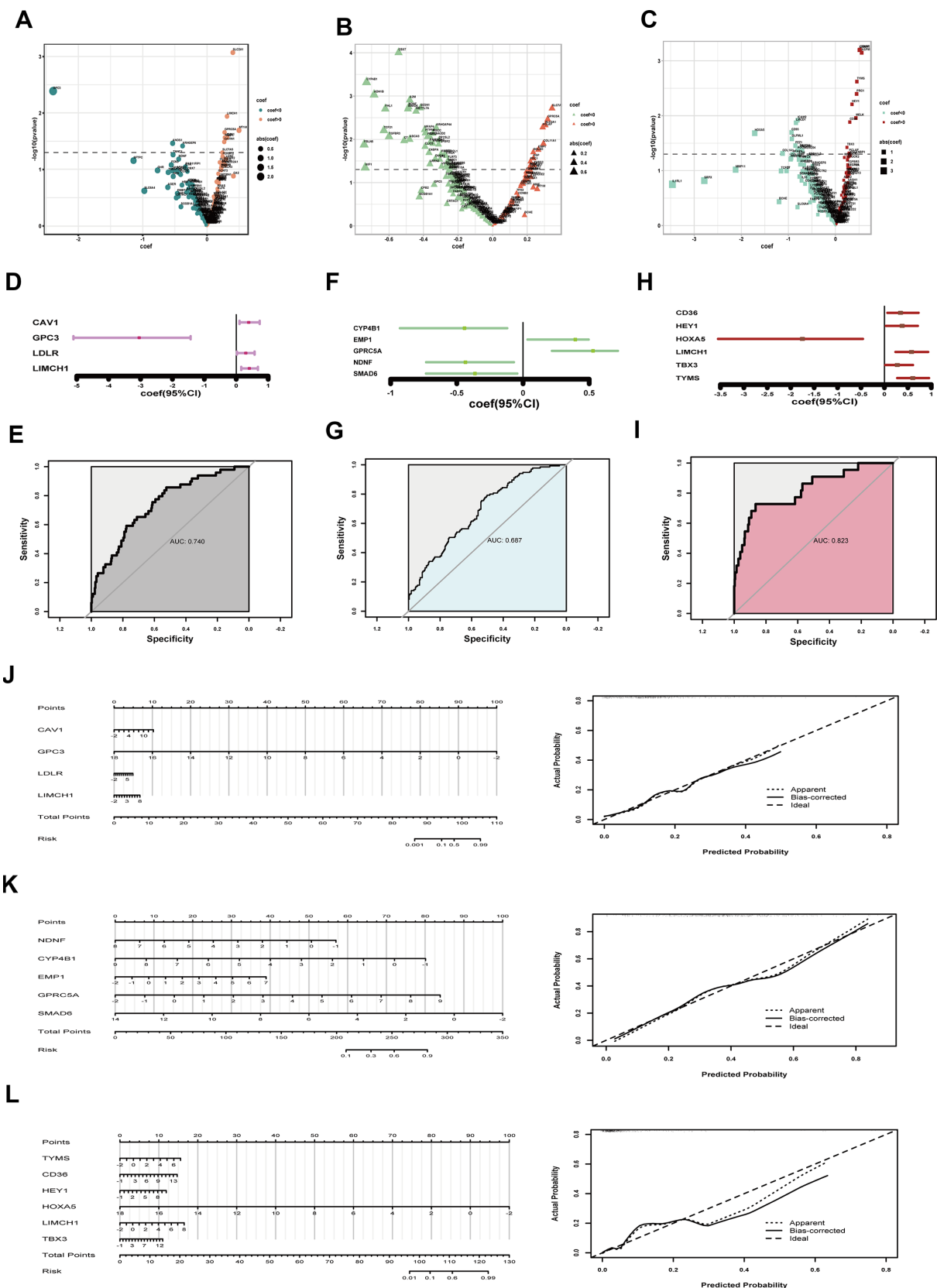
Figure 3. Assessing tumor size, lymph node invasion and distant metastasis. (**A–C**) coef and p value in univariate logistic regression analysis for tumor growth (**A**) lymph node invasion (**B**) and distant metastasis (**C**) respectively; (**D**) coef and 95% CI derived from the optimized model for tumor size; (**E**) ROC curve to show predictive potential of T-related model; (**F**) coef and 95% CI of the improved model for

(Figure 3E). In the improved model for lymph node invasion, we got five genes where three genes (CYP4B1, NDNF, SMAD6) showed protective efficacy, and two genes (EMP1, GPRC5A) presented risk function (C-index, 0.687; R^2, 0.170) (Figure 3F). Similarly, the optimized N-related model possessed fine predictive potency (AUC, 0.687) (Figure 3G). For model predicting distant metastasis, or named hematogenous metastasis, we ultimately obtained six genes, one (HOXA5) showed protective ability, five (CD36, HEY1, LIMCH1, TBX3, TYMS) exhibited hazardous effect (Figure 3H) (C-index, 0.823; R^2, 0.308). The optimized model of M parameter also had high consistency with reality (AUC, 0.823) (Figure 3I). Additionally, we plotted nomograph to predict risk probability of tumor growth, and calibration curve proved reasonable efficacy (Figure 3J). Analogously, lymph node invasion risk probability was presented, and the model was

in good agreement with actual situation (Figure 3K). Furthermore, risk probability for distant metastasis also demonstrated fine predictive efficacy (Figure 3L).

Establishing gene-based model for predicting overall survival probability of LUAD

Overall survival (OS) time is a momentous indicator for cancer prognosis. We tried to establish a gene expression model to predict OS probability in LUAD. We tentatively acquired 15 up-regulated genes and 34 down-regulated genes by log-rank analysis ($P<0.05$) (Supplementary Figure 7). Noticeably, univariate survival analysis of single gene is susceptible to be confounded, the correlation analysis showed signs (Figure 4A). So we applied multivariate analysis of all these genes by Cox proportional hazard regression,
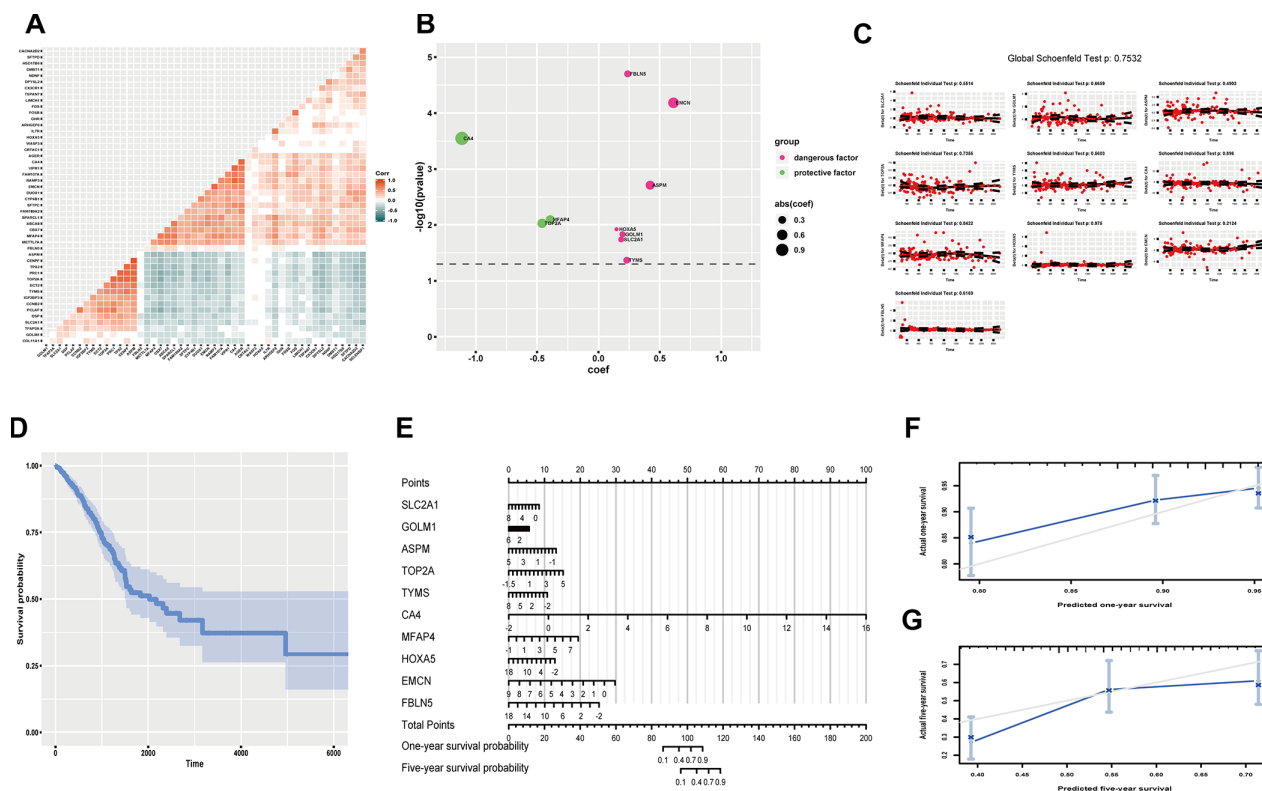

Figure 4. Predicting OS probability of LUAD. (**A**) Heatmap to exhibit correlation of genes; (**B**) coef and p value of genes in the optimized Cox proportional hazards regression model; (**C**) Residual plots to test proportional hazard assumption; (**D**) K-M curve exhibiting cumulative OS probability based on mean covariables; (**E**) Nomograph for predicting one-year and five-year OS probability; (**F**) Examining predicting efficacy of one year OS probability; (**G**) Examining predicting efficacy of five year OS probability. Corr, correlation coefficient; OS, overall survival; DEGs, differentially expressed genes; coef, coefficient; K-M, Kaplan-Meier.

however, little significance reappeared (Supplementary Figure 8). We optimized the model by stepwise regression. Finally, ten genes were contained, three had protective efficacy (CA4, MFAP4, TOP2A), seven possessed hazardous potential (FBLN5, EMCN, ASPM, HOXA5, GOLM1, SLC2A1, TYMS) (C-index, 0.678; R^2, 0.149) (Figure 4B). Proportional hazards examination meets the premise condition (Figure 4C). Meanwhile, Kaplan-Meier (K-M) curve was utilized to exhibit cumulative survival probability based on means of concomitant variables (Figure 4D). Later, we used the model to predict one-year and five-year OS probability of LUAD (Figure 4E). And calibration curve proved fine predictive potency (Figure 4F, 4G).

Reflecting progression-free survival probability of LUAD via multivariate analysis

As recurrence is an important indicator of worsening, progression-free survival (PFS) time functions as a crucial reference for clinical intervention. We screened out several related DEGs to reflect PFS probability of

LUAD. At first, we found 9 up-regulated genes and 20 down-regulated genes by log-rank test ($P<0.05$) (Supplementary Figure 9). And confounding effect still existed (Figure 5A). We then included all these genes in a multivariate Cox proportional hazard regression model, but none genes showed significance (Supplementary Figure 10). So, optimization was done by stepwise regression. At last, we chose three genes (IGF2BP3, SLC2A1, GOLM1), which were all regarded as hazardous factors (C-index, 0.611; R^2, 0.063) (Figure 5B). And the model abided by proportional hazards hypothesis well (Figure 5C). Further, we showed cumulative survival probability based on average scores of these genes (Figure 5D). Later, we presented nomograph to show one-year and five-year PFS probability (Figure 5E). And predictive efficacy of this model was good (Figure 5F, 5G).

DISCUSSION

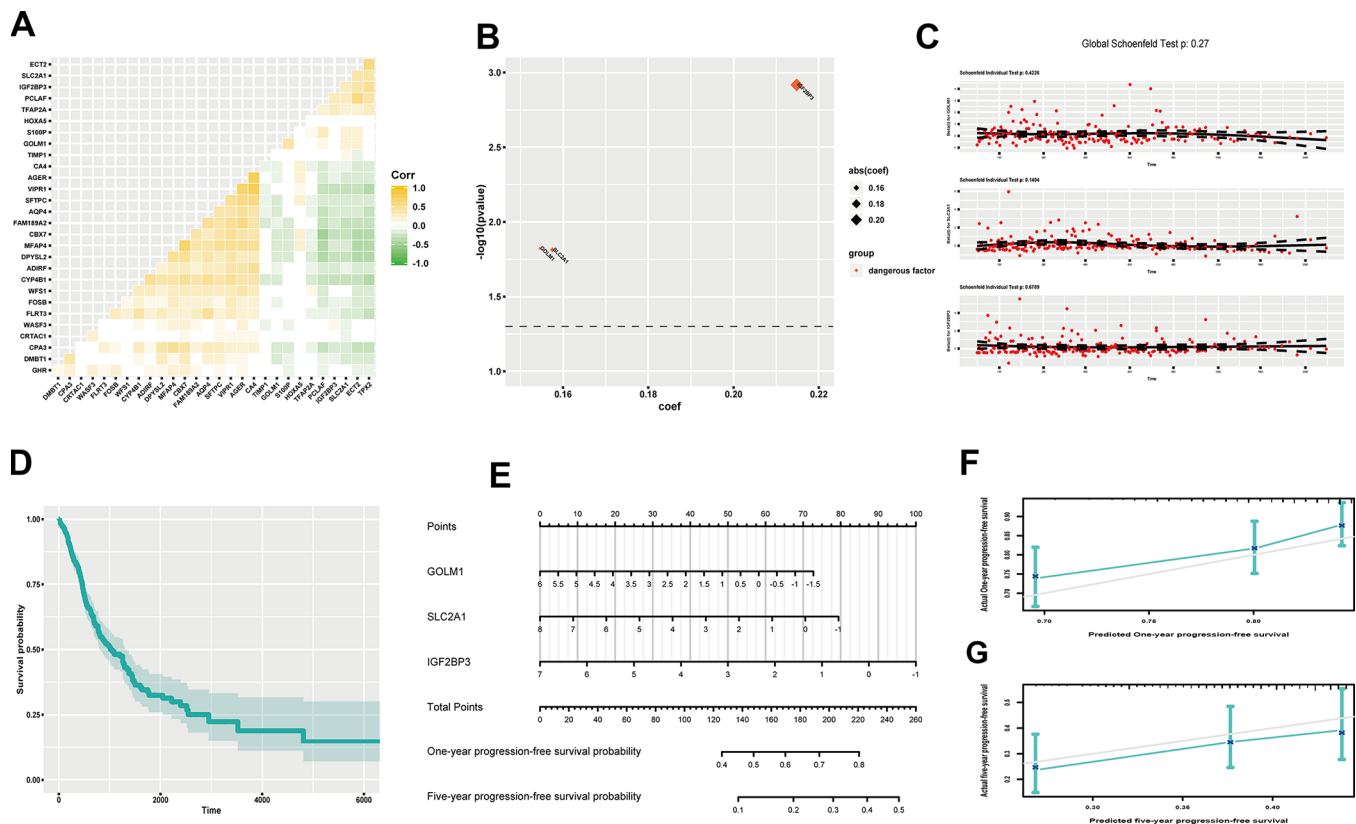Disorder of signaling network caused by aberrant genes triggers malignant transformation and progression,

Figure 5. Estimating PFS probability of LUAD. (**A**) Heatmap of gene correlation; (**B**) coef and p value of genes in the improved Cox proportional hazards regression model; (**C**) Residual plots to examine proportional hazard hypothesis; (**D**) K-M curve exhibiting cumulative PFS probability based on average covariables; (**E**) Nomograph for estimating one-year and five-year PFS probability; (**F**) Examining predictive ability of one year PFS probability; (**G**) Examining predictive ability of five year PFS probability. Corr, correlation coefficient; PFS, progression-free survival; DEGs, differentially expressed genes; coef, coefficient; K-M, Kaplan-Meier.

which provides theoretical basis for acquiring association between genes and clinical characteristics of tumor [13, 14]. We used gene expression data to establish models reflecting tumor growth risk, metastasis hazard, and survival probability of LUAD, which will avail clinical intervention and provide insight for understanding pathogenesis.

As we all know, findings of tumor driver genes facilitate revolutionary advances in therapeutic handling of cancer. However, core genes predominating carcinogenesis are difficult to dig out. Cancer-related genes probably present different expression in cancerous tissue compared to normal tissue. The good news is that filtrating DEGs from tens thousands of genes becomes convenient, with transcriptomic and bioinformatic technology developing rapidly [12]. We first derived DEGs of LUAD from four datasets (Figure 1). Next, we investigated genetic assignment, biological enrichment and interaction function of these DEGs (Figure 2). We found these genes have not concentrated on several specific chromosomes and presented dispersive distribution. Pathway enrichments highlighted importance of microenvironments, which were tightly related to almost all malignant hallmarks of lung cancer [15]. Furthermore, result of gene interaction showed hierarchy, suggesting further digging for dominant genes was imperative.

It goes without saying that tumor growth and metastasis constitute central aspects of cancer biological characters. Tumor TNM staging remains canonical criterion for clinical evaluation and intervention [8, 16]. At present, TNM staging is mainly based on surgical exploration and imaging examination, which both could have invasive effect. However, gene expression data, acquired handily by current genomic technology, could great help to evaluate TNM parameters by appropriate statistical models. Therefore, we tried to establish correlations between TNM scores and genes by logistic regression analysis in LUAD. And optimized models showed fine potential to assess TNM parameters (Figure 3). Noticeably, survival status is regarded as an ultimate indicator for prognosis. OS and PFS both occupy crucial parts in tumor investigation and handling strategies. We utilized multivariate Cox proportional hazards regression models to predict OS and PFS based on gene expression data. Both models had reasonable predictive efficacy (Figures 4, 5).

Of course, limitations of our research still exist. First, sample size remains not adequate enough as to great heterogeneity of LUAD, which could inevitably cause confounding effect, that is why our R^2 scores seemed not very perfect. Second, post-transcriptional regulation, post-translational modification and non-coding RNAs all contribute weightily to carcinogenesis, whereupon

transcriptomic profile of only protein coding content seems not comprehensive enough to read cancer [17–19]. Third, the structure of gene-encoded products significantly affects the function of genes, which in turn affects the regulation of genes on important biological activities, such as DNA replication [20, 21], cell migration [22, 23], and etc., while nanomedicine based on molecular structure also plays an increasingly important role in cancer prevention and treatment [24–26]. Therefore, the clinical prediction based on gene expression alone, without gene structure, may not be extremely thorough and thoughtful. At last, gene expression profiling was derived from tissues, which could cause more damage compared to emerging liquid biopsies, predictive models based of which will bring much blessedness in the battle against cancer [27, 28].

In summary, we established gene expression-based models for evaluating clinical features of LUAD via integrative analysis, which will assist diagnosis and treatment of LUAD as well as enlightening investigation of cancer pathogenesis.

MATERIALS AND METHODS

Transcriptomic and clinical data

LUAD microarrays reflecting transcriptome profiling in tumor tissue and normal tissue were obtained from Gene Expression Omnibus (GEO) database (GSE32863, 58 tumor tissues versus 58 normal tissues; GSE43458, 80 tumor tissues versus 30 normal tissues; GSE10072, 58 tumor tissues versus 49 normal tissues) [29–31]; LUAD RNA-sequencing data and corresponding clinicopathologic annotation were derived from TCGA program, where 58 tumors and 58 normal tissues were applied for screening out DEGs, 402 samples containing availably prognostic records for survival analysis and 371 samples possessing complete TNM scorings for risk analysis respectively [32, 33].

Genetic mapping, enrichment analysis and interaction investigation

BiomaRt, org.Hs.eg.db and RCircos packages were applied to annotate and map genetic information of genes [34–36]. And GO analysis was used to describe molecular function, cellular component, and biological process via clusterProfiler package [37]. Subsequently, clusterProfiler and ReactomePA packages were employed for pathway enrichment analysis based on Kyoto Encyclopedia of Genes and Genomes (KEGG) and Reactome database [37, 38]. Then STRING database was utilized for investigating gene interaction [39]. Adjusted P-value <0.05 was considered statistically significant.

Statistical methods

DEGs derived from different transcriptome data were acquired by suitable methods respectively (Adjusted *P*-value <0.05, fold change >2 or <0.5). That is, limma package was used for microarrays, and DESeq2 package was applied for RNA-sequencing [40, 41]. Univariate and multivariate logistic regression model were used to handle two-category data. K-M curve was utilized to establish cumulative survival probability. The survival impact of single gene was estimated by log-rank test, while the Cox proportional hazards regression model was applied for multivariable analysis. Proportional hazards assumption in Cox regression was assessed by Schoenfeld residual tests. Pearson correlation analysis was used to investigate correlation. AIC was used to select and optimize models. Likelihood ratio test, Wald test, scoring test were applied for statistical hypothesis testing. *P*<0.05 was considered significant. All related arithmetic functions were practiced in R language [42].

Abbreviations

LUAD: lung adenocarcinoma; DEGs: differentially expressed genes; GEO: Gene Expression Omnibus; TCGA: The Cancer Genome Atlas; GO: gene ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; K-M: Kaplan-Meier; AIC: Akaike information criterion; AUC: area under the curve; OS: overall survival; PFS: progression-free survival; coef: coefficient; CI: confidence interval; ROC: receiver operating characteristic; Corr: correlation coefficient

AUTHOR CONTRIBUTIONS

Yanlu Xiong: conceptualization, data curation, software and writing - original draft; Jie Lei and Jinbo Zhao: data curation, methodology, writing - review and editing; Yangbo Feng, Tianyun Qiao and Yongsheng Zhou: validation and formal analysis; Tao Jiang and Yong Han: project administration, supervision, funding acquisition and writing - review and editing. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

We sincerely thanked Prof. Lei Shang, an expert in statistics and Prof. Lintao Jia, an expert in tumor biology, at the Fourth Military Medical University, for their professional guidance.

CONFLICTS OF INTEREST

The authors have declared no conflicts of interest.

Editorial note

FUNDING

REFERENCES

1. 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; 68:394–424.
https://doi.org/10.3322/caac.21492
PMID:30207593

2. Thomas A, Liu SV, Subramaniam DS, Giaccone G. Refining the treatment of NSCLC according to histological and molecular subtypes. Nat Rev Clin Oncol. 2015; 12:511–26.
https://doi.org/10.1038/nrclinonc.2015.90
PMID:25963091

3. Chen Z, Fillmore CM, Hammerman PS, Kim CF, Wong KK. Non-small-cell lung cancers: a heterogeneous set of diseases. Nat Rev Cancer. 2014; 14:535–46.
https://doi.org/10.1038/nrc3775
PMID:25056707

4. Saito M, Suzuki H, Kono K, Takenoshita S, Kohno T. Treatment of lung adenocarcinoma by molecular-targeted therapy and immunotherapy. Surg Today. 2018; 48:1–8.
https://doi.org/10.1007/s00595-017-1497-7
PMID:28280984

5. Herbst RS, Morgensztern D, Boshoff C. The biology and management of non-small cell lung cancer. Nature. 2018; 553:446–54.
https://doi.org/10.1038/nature25183
PMID:29364287

6. Chalela R, Curull V, Enríquez C, Pijuan L, Bellosillo B, Gea J. Lung adenocarcinoma: from molecular basis to genome-guided therapy and immunotherapy. J Thorac Dis. 2017; 9:2142–58.
https://doi.org/10.21037/jtd.2017.06.20
PMID:28840016

7. Goldstraw P, Ball D, Jett JR, Le Chevalier T, Lim E, Nicholson AG, Shepherd FA. Non-small-cell lung cancer. Lancet. 2011; 378:1727–40.
https://doi.org/10.1016/S0140-6736(10)62101-0

PMID:21565398

8. Amin MB, Greene FL, Edge SB, Compton CC, Gershenwald JE, Brookland RK, Meyer L, Gress DM, Byrd DR, Winchester DP. The eighth edition AJCC cancer staging manual: continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging. CA Cancer J Clin. 2017; 67:93–99.
https://doi.org/10.3322/caac.21388
PMID:28094848

9. Marusyk A, Polyak K. Tumor heterogeneity: causes and consequences. Biochim Biophys Acta. 2010; 1805:105–17.
https://doi.org/10.1016/j.bbcan.2009.11.002
PMID:19931353

10. Karczewski KJ, Snyder MP. Integrative omics for health and disease. Nat Rev Genet. 2018; 19:299–310.
https://doi.org/10.1038/nrg.2018.4
PMID:29479082

11. Roychowdhury S, Chinnaiyan AM. Translating cancer genomes and transcriptomes for precision oncology. CA Cancer J Clin. 2016; 66:75–88.
https://doi.org/10.3322/caac.21329 PMID:26528881

12. Cieślik M, Chinnaiyan AM. Cancer transcriptome profiling at the juncture of clinical translation. Nat Rev Genet. 2018; 19:93–109.
https://doi.org/10.1038/nrg.2017.96
PMID:29279605

13. Chin L, Andersen JN, Futreal PA. Cancer genomics: from discovery science to personalized medicine. Nat Med. 2011; 17:297–303.
https://doi.org/10.1038/nm.2323
PMID:21383744

14. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011; 144:646–74.
https://doi.org/10.1016/j.cell.2011.02.013
PMID:21376230

15. Altorki NK, Markowitz GJ, Gao D, Port JL, Saxena A, Stiles B, McGraw T, Mittal V. The lung microenvironment: an important regulator of tumour growth and metastasis. Nat Rev Cancer. 2019; 19:9–31.
https://doi.org/10.1038/s41568-018-0081-9
PMID:30532012

16. Edge SB, Compton CC. The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM. Ann Surg Oncol. 2010; 17:1471–74.
https://doi.org/10.1245/s10434-010-0985-4
PMID:20180029

17. Krueger KE, Srivastava S. Posttranslational protein modifications: current implications for cancer detection, prevention, and therapeutics. Mol Cell Proteomics. 2006; 5:1799–810.
https://doi.org/10.1074/mcp.R600009-MCP200
PMID:16844681

18. Schmitt AM, Chang HY. Long noncoding RNAs in cancer pathways. Cancer Cell. 2016; 29:452–63.
https://doi.org/10.1016/j.ccell.2016.03.010
PMID:27070700

19. Audic Y, Hartley RS. Post-transcriptional regulation in cancer. Biol Cell. 2004; 96:479–98.
https://doi.org/10.1016/j.biolcel.2004.05.002
PMID:15380615

20. Kou Y, Koag MC, Lee S. Structural and kinetic studies of the effect of guanine N7 alkylation and metal cofactors on DNA replication. Biochemistry. 2018; 57:5105–16.
https://doi.org/10.1021/acs.biochem.8b00331
PMID:29957995

21. Kou Y, Koag MC, Lee S. N7 methylation alters hydrogen-bonding patterns of guanine in duplex DNA. J Am Chem Soc. 2015; 137:14067–70.
https://doi.org/10.1021/jacs.5b10172
PMID:26517568

22. Dedden D, Schumacher S, Kelley CF, Zacharias M, Biertümpfel C, Fässler R, Mizuno N. The architecture of Talin1 reveals an autoinhibition mechanism. Cell. 2019; 179:120–31.e13.
https://doi.org/10.1016/j.cell.2019.08.034
PMID:31539492

23. Yamada KM, Sixt M. Mechanisms of 3D cell migration. Nat Rev Mol Cell Biol. 2019; 20:738–52.
https://doi.org/10.1038/s41580-019-0172-9
PMID:31582855

24. Wen R, Umeano AC, Kou Y, Xu J, Farooqi AA. Nanoparticle systems for cancer vaccine. Nanomedicine (Lond). 2019; 14:627–48.
https://doi.org/10.2217/nnm-2018-0147
PMID:30806568

25. Martin JD, Cabral H, Stylianopoulos T, Jain RK. Improving cancer immunotherapy using nanomedicines: progress, opportunities and challenges. Nat Rev Clin Oncol. 2020; 17:251–266.
https://doi.org/10.1038/s41571-019-0308-z
PMID:32034288

26. Russell LM, Liu CH, Grodzinski P. Nanomaterials innovation as an enabler for effective cancer interventions. Biomaterials. 2020; 242:119926.
https://doi.org/10.1016/j.biomaterials.2020.119926
PMID:32169771

27. Bardelli A, Pantel K. Liquid biopsies, what we do not know (yet). Cancer Cell. 2017; 31:172–79.
https://doi.org/10.1016/j.ccell.2017.01.002

PMID:28196593

28. Haber DA, Velculescu VE. Blood-based analyses of cancer: circulating tumor cells and circulating tumor DNA. Cancer Discov. 2014; 4:650–61. https://doi.org/10.1158/2159-8290.CD-13-1014 PMID:24801577

29. Selamat SA, Chung BS, Girard L, Zhang W, Zhang Y, Campan M, Siegmund KD, Koss MN, Hagen JA, Lam WL, Lam S, Gazdar AF, Laird-Offringa IA. Genome-scale analysis of DNA methylation in lung adenocarcinoma and integration with mRNA expression. Genome Res. 2012; 22:1197–211. https://doi.org/10.1101/gr.132662.111 PMID:22613842

30. Kabbout M, Garcia MM, Fujimoto J, Liu DD, Woods D, Chow CW, Mendoza G, Momin AA, James BP, Solis L, Behrens C, Lee JJ, Wistuba II, Kadara H. ETS2 mediated tumor suppressive function and MET oncogene inhibition in human non-small cell lung cancer. Clin Cancer Res. 2013; 19:3383–95. https://doi.org/10.1158/1078-0432.CCR-13-0341 PMID:23659968

31. Landi MT, Dracheva T, Rotunno M, Figueroa JD, Liu H, Dasgupta A, Mann FE, Fukuoka J, Hames M, Bergen AW, Murphy SE, Yang P, Pesatori AC, et al. Gene expression signature of cigarette smoking and its role in lung adenocarcinoma development and survival. PLoS One. 2008; 3:e1651. https://doi.org/10.1371/journal.pone.0001651 PMID:18297132

32. Weinstein JN, Collisson EA, Mills GB, Shaw KR, Ozenberger BA, Ellrott K, Shmulevich I, Sander C, Stuart JM, and Cancer Genome Atlas Research Network. The cancer genome atlas pan-cancer analysis project. Nat Genet. 2013; 45:1113–20. https://doi.org/10.1038/ng.2764 PMID:24071849

33. Tomczak K, Czerwińska P, Wiznerowicz M. The cancer genome atlas (TCGA): an immeasurable source of knowledge. Contemp Oncol (Pozn). 2015; 19:A68–77. https://doi.org/10.5114/wo.2014.47136 PMID:25691825

34. Durinck S, Spellman PT, Birney E, Huber W. Mapping identifiers for the integration of genomic datasets with the r/bioconductor package biomaRt. Nat Protoc. 2009; 4:1184–91.

https://doi.org/10.1038/nprot.2009.97 PMID:19617889

35. Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, Morgan MT, Carey VJ. Software for computing and annotating genomic ranges. PLoS Comput Biol. 2013; 9:e1003118. https://doi.org/10.1371/journal.pcbi.1003118 PMID:23950696

36. Zhang H, Meltzer P, Davis S. RCircos: an R package for circos 2D track plots. BMC Bioinformatics. 2013; 14:244. https://doi.org/10.1186/1471-2105-14-244 PMID:23937229

37. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012; 16:284–87. https://doi.org/10.1089/omi.2011.0118 PMID:22455463

38. Yu G, He QY. ReactomePA: an r/bioconductor package for reactome pathway analysis and visualization. Mol Biosyst. 2016; 12:477–79. https://doi.org/10.1039/c5mb00663e PMID:26661513

39. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, Santos A, Doncheva NT, Roth A, Bork P, Jensen LJ, von Mering C. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. Nucleic Acids Res. 2017; 45:D362–68. https://doi.org/10.1093/nar/gkw937 PMID:27924014

40. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015; 43:e47. https://doi.org/10.1093/nar/gkv007 PMID:25605792

41. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014; 15:550. https://doi.org/10.1186/s13059-014-0550-8 PMID:25516281

42. Chan BK. Data analysis using R programming. Adv Exp Med Biol. 2018; 1082:47–122. https://doi.org/10.1007/978-3-319-93791-5_2 PMID:30357717

SUPPLEMENTARY MATERIALS

Supplementary Figure
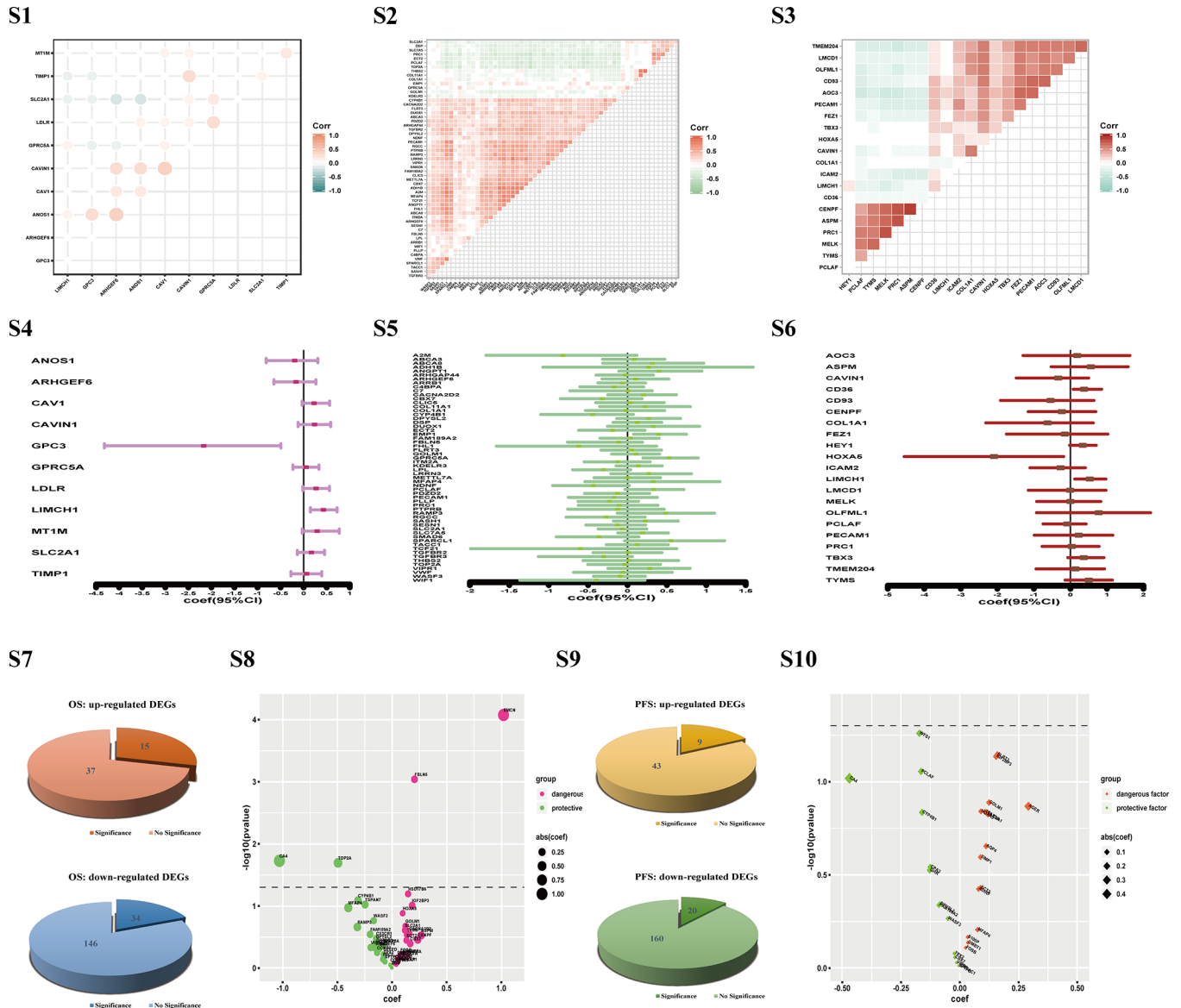
Supplementary Figure. (S1) (S2) (S3) correlation analysis for preliminarily filtrated DEGs for T (S1), N (S2), and M (S3) parameters respectively; (S4) (S5) (S6) coef and 95% CI of the preliminary multivariate logistic regression models of T (S4), N (S5), and M (S6) parameters respectively; (S7) screened OS-related DEGs by log-rank test (*P*<0.05 was considered as significance); (S8) coef and p value of OS-related genes in the preliminary Cox proportional hazards regression model; (S9) sifted PFS-related DEGs by log-rank test (*P*<0.05 was regarded as significance); (S10) coef and p value of PFS-related genes in the initial Cox proportional hazards regression model DEGs, differentially expressed genes; coef, coefficient; Corr, correlation coefficient; CI, confidence interval; OS, overall survival; PFS, progression-free survival.

Supplementary Tables

Please browse Full Text version to see the data of Supplementary Tables 1, 2.

Supplementary Table 1. Genomic information of DEGs selected from comparison of tumor vs. normal tissues.

Supplementary Table 2. Results of GO analysis and pathway enrichment analysis of DEGs.