Identification and validation of two LUAD-development characteristic gene sets for diagnosing lung adenocarcinoma and predicting prognosis

Cheng Liu
The Fourth Affiliated Hospital of Harbin Medical University
Xiang Li (a19ak0f@163.com)
The Fourth Affiliated Hospital of Harbin Medical University
Hua Shao
The Fourth Affiliated Hospital of Harbin Medical University
Dan Li
The Fourth Affiliated Hospital of Harbin Medical University

Research article

Keywords: Lung adenocarcinoma, Prognostic stratification system, TCGA, Gene set

Posted Date: January 27th, 2020

DOI: https://doi.org/10.21203/rs.2.21884/v1

Abstract

Purpose

Lung adenocarcinoma (LUAD) is one of the main types of lung cancer, the low rates for early diagnosis and a bad prognosis for advanced stage lead to a higher mortality rate. Therefore, it is of great significance to identify the related genes that promote its development.

Patients and methods

512 LUADs from The Cancer Genome Atlas were used to performed differentially expressed gene (DEG) analysis and short-term time-series expression miner to identify the LUAD-development characteristic gene sets. Survival analysis was applied to identified LUAD-unfavorable gene set and LUAD-favorable gene set. Gene set variation analysis (GSVA) was performed to score individual samples against the two gene sets. ROC curve analysis, univariate and multivariate cox regression analysis were used to evaluate the diagnostic and prognostic value of the two GSVA score systems. Two independent data sets from GEO were used for verifying the results. Functional enrichment analysis was used to explore the potential biological functions of LUAD-unfavorable gene set.

Results

With the development of LUAD, 185 DEGs were gradually up-graduated, including 84 genes associated with survival and classed as LUAD-unfavorable gene set; 237 DEGs were gradually down-graduated, including 39 genes associated with survival and classed as LUAD-favorable gene set. ROC curve analysis and univariate/multivariate Cox proportional hazards analyses indicated both of LUAD-unfavorable GSVA score and LUAD-favorable GSVA score were biomarkers for diagnosing LUAD and independent biomarkers for predicting prognosis. The LUAD-unfavorable genes were involved in multiple cancer-related pathways, such as p53 signaling pathway and cell cycle.

Conclusion

We identified and validated two LUAD-development characteristic gene sets that not only have diagnostic value but also prognostic value. It may provide new insight for further research on LUAD.

Introduction

Lung cancer is the highest incidence of cancer in men and women worldwide (11.6% of the total cancer cases), which is also the major dead cause of cancer (18.4% of the total cancer deaths)[1]. Non-small cell lung cancer (NSCLC) accounts for 85% of all lung Cancer cases[2], and lung adenocarcinoma (LUAD) is one of the main subtypes of NSCLC. At present, it is generally believed that smoking is the dominant cause of lung cancer, while LUAD is more likely to occur in non-smoking women, and titles younger age showing trends[3, 4]. Early LUAD is highly curable by surgery, but its early diagnosis rate is low, patients are

often diagnosed with advanced LUAD[5] with a high mortality rate. Although target therapy is effective for selected advanced LUAD, the overall survival of patients is poor due to the emergence of drug resistance. The looking for LUAD diagnosis and prognosis biomarkers remains one of the hot spots in clinical research.

In recent years, high-throughput sequencing techniques and gene databases have been widely performed in studying of the cancer diagnosis and prognostic features[6 – 9]. For example, EGFR, KRAS, BRAF, and ERBB 2 have been shown to be associated with treatment efficacy and prognosis[10 – 12]. However, most studies only focus on a single gene, and do not take the simultaneous changes of multiple genes into account. Moreover, there are few studies on the LUAD-development characteristic gene sets.

Hereon,we identified two LUAD-development characteristic gene sets named as LUAD-unfavorable gene set and LUAD-favorable gene set. Gene set variation analysis (GSVA) is used to score individual samples against the two gene sets. Both of LUAD-unfavorable GSVA score and LUAD-favorable GSVA score were reliable biomarkers for diagnosing LUAD and independent biomarkers for predicting prognosis.

Material And Methods

The Cancer Genome Atlas (TCGA, https://www.cancer.gov/)[13] and Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/)[14] are the international genetic databases that are publicly accessible and freely available to researchers. In TCGA, there are 281 LUADs with stage I, 121 LUADs with stage II, 84 LUADs with stage III, 26 LUADs with stage IV and 57 healthy lung tissue samples. In addition, GSE10072 based on GPL96 platform was download from GEO, included 58 LUAD samples and 49 healthy lung tissue samples. GSE31210 based on GPL570 platform, included 226 LUAD samples and 20 healthy lung tissue samples. The two data set was used to verify the prognostic value. And GSE31210 was also used to validated the prognosis value. The “normalizeBetweenArrays” function in the limma package[15] was used to normalize the gene expression profiles. If a gene responds to a multiple probes, the average value of these probes is considered to be the expression value of the corresponding gene.

Differentially expressed gene (DEG) analysis and short time-series expression miner (STEM)

The RNA sequencing expression profile of LUAD in TCGA displayed as read counts was normalized by voom function[16] in limma package. Then, we identified DEG using limma package for 4 LUAD stages and healthy lung tissue samples respectively. DEG was set as $P < 0.01$ after FDR correction and $|logFC| > 1.5$ as the threshold. In the developing of LUAD, if a DEG was gradually up-regulated ($logFC_{stage\ I\ vs\ control} < logFC_{stage\ II\ vs\ control} < logFC_{stage\ III\ vs\ control} < logFC_{stage\ IV\ vs\ control}$) or gradually down-regulated ($logFC_{stage\ I\ vs\ control} > logFC_{stage\ II\ vs\ control} > logFC_{stage\ III\ vs\ control} > logFC_{stage\ IV\ vs\ control}$),then it was considered to be LUAD-development characteristic gene These genes were organized into different clusters based on expression patterns using STEM[17].

Survival analysis and LUAD-development characteristic gene set

The median expression value for each LUAD-development characteristic gene as the cutoff to dichotomize patients into high-expression and low-expression groups. To determine if the expression level of LUAD-development characteristic gene is associated with overall survival times, we applied a Kaplan–Meier survival analysis with the log-rank method. Survival analysis was performed using survival package (https://CRAN.R-project.org/package=survival) in R. $P < 0.01$ was considered to be significant. A LUAD-development characteristic gene that is gradually up-regulated in LUAD development and associated with poor prognosis is defined as LUAD-unfavorable gene, while a LUAD-development characteristic gene that is gradually down-regulated in lung cancer development and associated with good prognosis is defined as LUAD-favorable gene. Subsequently, two LUAD-development characteristic gene sets were collected,including LUAD-unfavorable gene set and LUAD-favorable gene set.

Calculation of LUAD-development characteristic GSVA score

GSVA is popular approach to score individual samples against molecular signatures (or gene sets). GSVA package[18] in R was used to calculate LUAD-unfavorable GSVA score and LUAD-favorable GSVA score for individual sample.

Receiver operating characteristic (ROC) curve analysis and univariate/multivariate Cox proportional hazards analyses

The pROC package [19] was used to conduct ROC curve analysis of LUAD-unfavorable GSVA score and LUAD-unfavorable GSVA score to evaluate their ability to diagnose LUAD. Univariate/multivariate Cox proportional hazards analyses were used to compare the relative prognostic value of the two GSVA score systems with that of routine clinicopathological features.

Functional enrichment analysis

To explore the biology of these genes, Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses for the two LUAD-development characteristic gene sets were performed respectively using the clusterProfiler package [20] in R. $P < 0.05$ was considered significant.

Validation of the up-regulation of LUAD-unfavorable genes compared to normal lung tissue at protein level

In order to confirm the up-regulation of LUAD-unfavorable genes compared to normal lung tissue at protein level, we randomly selected 10 genes of LUAD-unfavorable gene set and scanned the Human

Protein Atlas (https://v15.proteinatlas.org/) [21] web tool to validate the differential expression of these ten genes at the protein level.

Results

In the present study, LUAD samples were used to carry out differentially expressed gene analysis and survival analysis and identified LUAD-development characteristic gene sets. Functional enrichment analysis was performed for the two gene sets to explore their biological function. LUAD-unfavorable GSVA score and LUAD-favorable GSVA score were calculated for each samples. The two GSVA score systems were significantly associated with LUAD prognosis.Two independent data sets was used to validate our findings. (Fig. 1).

Multiple genes were defined as LUAD-development characteristic genes

Compared to control samples, there were 3082 DEGs in stage I LUADs, 3437 DEGs in stage II LUADs, 3518 DEGs in stage III LUADs and 3510 DEGs stage IV LUADs (Fig. 2A). The results indicated the gene expression patterns were various with the LUAD development. Thus, identification of LUAD-development characteristic genes is crucial. A total of 2658 common DEGs was in stage I-IV LUADs (Fig. 2B). Among of them, 185 DEGs were gradually up-regulated and 237 DEGs were gradually down-regulated with LUAD development, these genes maybe play a crucial role in the LUAD development. STEM identified that two gene profiles (clusters) were significantly presenting an upward trend and two gene profiles (clusters) were significantly presenting a downward trend in the LUAD development (Fig. 2C).

LUAD-development characteristic genes were associated with LUAD prognosis

A total of 84 LUAD-development characteristic genes that are gradually up-regulated in LUAD development and associated with poor prognosis, and these 84 genes formed the LUAD-unfavorable gene set. Thirty-nine LUAD-development characteristic genes that are gradually down-regulated in LUAD development and associated with good prognosis and these 39 genes formed the LUAD-favorable gene set (Table 1). This means that not all LUAD-development characteristic genes are associated with the prognosis of LUAD. In the LUAD-unfavorable gene set, NEK2, CENPK, CDC25C, PLK4, LYPD3, FAM72D, NEIL3, GTSE1, CDK1 and KIF14 were the ten genes with most significant association with poor prognosis (Fig. 3A). While in the LUAD-favorable gene set, OR7E47P, MS4A2, RAB44, BMP5, ARHGEF6, JAML, TRPC2, HPGDS, HPSE2, and KLK11 were the ten genes with most significant association with good prognosis (Fig. 3B). Moreover, ten genes (ASPM, BLM, CDC25C, CDK1, DEPDC1, KIF11, KIF14, LYPD3, NEK2, PLK4) of LUAD-unfavorable gene set were included in The Human Protein Atlas, and high expressed in LUAD compared to normal lung (Fig. 3C).

LUAD-unfavorable GSVA score and LUAD-favorable GSVA score are biomarkers of LUAD and LUAD prognosis

Each sample was calculated LUAD-unfavorable GSVA score and LUAD-favorable GSVA score using GSVA package. Obviously, LUAD-favorable GSVA score was decreasing, while LUAD-unfavorable GSVA score was increasing with LUAD development (Fig. 4A). ROC curve analysis indicated that both LUAD-unfavorable GSVA score and LUAD-favorable GSVA score are biomarker of LUAD with AUC = 0.982 and AUC = 0.994, respectively (Fig. 4B). Furthermore, they were validated in two independent data sets GSE10072 (Fig. 4C) and GSE31210 (Fig. 4D). All LUAD patients in TCGA were separated into low- and high-score groups according to median GSVA score. Both the two GSVA score systems were significant associated with LUAD prognosis (Fig. 4E). Patients with high LUAD-unfavorable GSVA score had more poor prognosis, and patients with high LUAD-favorable GSVA score had better prognosis. Univariate and multivariate Cox analyses showed the two GSVA score systems for prognosis were significant independent factors adjusted for clinicopathological features (Table 2–3). Moreover, the two GSVA score systems were still significantly associated with prognosis in an independent test set (GSE31210) (Fig. 4F).

LUAD-unfavorable genes involved in multiple cancer-related pathways

In order to explore the biological functions of LUAD-unfavorable genes, LUAD-unfavorable genes were performed functional enrichment analysis. The results show that these genes are mainly related to organelle fission,nuclear division□nuclear chromosome segregation□chromosome segregation and other biological functions (Fig. 5A). LUAD-unfavorable genes significantly involved in multiple cancer-related pathways, sunch as progesterone – mediated oocyte maturation, Oocyte meiosis□Cell cycle and p53 signaling pathway. (Fig. 5B)

Discussion

Around the world, lung cancer is the main cause of cancer-related deaths, and even with surgical treatment, the recurrence rate is very high[22]. Therefore, it is of great significance to explore biomarkers that can accurately diagnose lung cancer and predict prognosis for the treatment and management of lung cancer. Numerous studies have shown that abnormally expressed genes in lung cancer (including LUAD) are closely related to prognosis and can be considered as the potential prognostic biomarkers[23 – 25].

In the present study, we found the number of DEG is various in different stages of LUAD-development. This indicated gene expression patterns were various with the LUAD development. Compared to healthy lung tissue, a gene may be differentially expressed in early LUAD, but not in advanced stage. We identified 422 LUAD-development characteristic genes, including 185 genes gradually up-regulated and 237 genes

gradually down-regulated with LUAD-development. The development of LUAD results from synergistic effects of multiple genes. Notably, not all LUAD-development characteristic genes are associated with the prognosis of LUAD. LUAD-unfavorable gene set contained 84 gradually up-regulated DEGs and LUAD-favorable gene set contained 39 gradually down-regulated DEGs. Unsurprisingly, Previous studies have suggested that some of them are associated with LUAD development. NEK2 is overexpressed in a variety of malignant tumors, and is closely related to tumor drug resistance, rapid recurrence and poor prognosis[26 – 28]. KIF14 has also been found to be associated with poor prognosis in a variety of cancers[29, 30]. While in the LUAD-favorable gene set, genes that were significantly associated with LUAD survival included OR7E47P, MS4A2, RAB44, BMP5, ARHGEF6, KLK11and etc. Among them, KLK11 was found to be a diagnostic and prognostic indicator of NSCLC[31]. This just more confirms the possibility that the LUAD-unfavorable gene set and LUAD-unfavorable gene set can be used as a prognostic model for LUAD.

All samples were calculated LUAD-unfavorable GSVA scores and LUAD-favorable GSVA scores. This is obviously different from the gene signatures in other previous studies[32 – 34]. In the previous studies, a gene often got a coefficient from a Cox regression analysis or other method in the training set. However, due to the limitations of the sample size and the heterogeneity of the tumor, we may never know the true coefficient of a gene. Therefore, GSVA was used to score individual samples against gene sets (LUAD-unfavorable gene set and LUAD-favorable gene set) in our study. ROC curve analysis suggested that both LUAD-unfavorable GSVA score and LUAD-favorable GSVA score exhibited strong diagnostic capacity of LUAD and which was verified by two independent data sets. Univariate and multivariate Cox regression analysis suggested that LUAD-unfavorable GSVA score and LUAD-unfavorable gene set were independent prognostic factors for LUAD's overall survival. This resultS was also verified by a independent data set.

Functional enrichment analysis indicates that LUAD-unfavorable genes are significantly involved in p53 signaling pathway, cell cycle and other pathways. It is suggested that LUAD-unfavorable genes may be involved in the occurrence and development of LUAD through these channels. However, further studies are needed to investigate and validate the functions of these genes.

In the present study, although we provided new insights into the LUAD prognostic stratification system, several limitations were notable. Firstly, the two gene sets may be too large. Their application to the clinic still needs to wait for further decline in sequencing costs. Secondly, the synergy between the genes of these two gene sets to promote LUAD development still requires molecular experimental validation.

Conclusion

In conclusion, we identified and validated two LUAD-development characteristic gene sets that not only have diagnostic value but also prognostic value. It may provide new insight for further research on LUAD.

List Of Abbreviations

lung adenocarcinoma (LUAD)

differentially expressed gene (DEG)

gene set variation analysis (GSVA)

non-small cell lung cancer (NSCLC)

The Cancer Genome Atlas (TCGA)

short time-series expression miner (STEM)

receiver operating characteristic (ROC)

Gene Ontology (GO)

Kyoto Encyclopedia of Genes and Genomes (KEGG)

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The authors declare that they have no competing interests

Competing interests

The authors declare that they have no competing interests

Funding

The study was supported by Subject of Education Department of Heilongjiang Provincial (No. 12531258 and 12511264).

Authors' contributions

Cheng Liu conceived and designed the study, and drafted the manuscript. Xiang Li and Hua Shao collected, analyzed and interpreted the experimental data. Dan Li contributed to the data collection, data interpretation and manuscript reviewing. All authors read and approved the final manuscript.

Acknowledgements

Not applicable

References

1. Bray F, Ferlay J, Soerjomataram I, et al. 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. 2018/09/13. DOI: 10.3322/caac.21492.
2. Govindan R, Page N, Morgensztern D, et al. Changing epidemiology of small-cell lung cancer in the United States over the last 30 years: analysis of the surveillance, epidemiologic, and end results database. *J Clin Oncol* 2006; 24: 4539-4544. 2006/09/30. DOI: 10.1200/JCO.2005.04.4859.
3. Hecht SS. Tobacco smoke carcinogens and lung cancer. *J Natl Cancer Inst* 1999; 91: 1194-1210. 1999/07/21. DOI: 10.1093/jnci/91.14.1194.
4. Donner I, Katainen R, Sipila LJ, et al. Germline mutations in young non-smoking women with lung adenocarcinoma. *Lung Cancer* 2018; 122: 76-82. 2018/07/24. DOI: 10.1016/j.lungcan.2018.05.027.
5. Ding L, Getz G, Wheeler DA, et al. Somatic mutations affect key pathways in lung adenocarcinoma. *Nature* 2008; 455: 1069-1075. 2008/10/25. DOI: 10.1038/nature07423.
6. Zhao K, Li Z and Tian H. Twenty-gene-based prognostic model predicts lung adenocarcinoma survival. *Onco Targets Ther* 2018; 11: 3415-3424. 2018/06/22. DOI: 10.2147/OTT.S158638.
7. He SY, Xi WJ, Wang X, et al. Identification of a Combined RNA Prognostic Signature in Adenocarcinoma of the Lung. *Med Sci Monit* 2019; 25: 3941-3956. 2019/05/28. DOI: 10.12659/MSM.913727.
8. Dama E, Melocchi V, Dezi F, et al. An Aggressive Subtype of Stage I Lung Adenocarcinoma with Molecular and Prognostic Characteristics Typical of Advanced Lung Cancers. *Clin Cancer Res* 2017; 23: 62-72. 2016/07/01. DOI: 10.1158/1078-0432.CCR-15-3005.
9. Feng A, Tu Z and Yin B. The effect of HMGB1 on the clinicopathological and prognostic features of non-small cell lung cancer. *Oncotarget* 2016; 7: 20507-20519. 2016/02/04. DOI: 10.18632/oncotarget.7050.
10. Guan JL, Zhong WZ, An SJ, et al. KRAS mutation in patients with lung cancer: a predictor for poor prognosis but not for EGFR-TKIs or chemotherapy. *Ann Surg Oncol* 2013; 20: 1381-1388. 2012/12/05. DOI: 10.1245/s10434-012-2754-z.
11. Naoki K, Chen TH, Richards WG, et al. Missense mutations of the BRAF gene in human lung adenocarcinoma. *Cancer Res* 2002; 62: 7001-7003. 2002/12/04.
12. Mendelsohn J and Baselga J. Status of epidermal growth factor receptor antagonists in the biology and treatment of cancer. *J Clin Oncol* 2003; 21: 2787-2799. 2003/07/16. DOI: 10.1200/JCO.2003.01.504.

13. Tomczak K, Czerwinska P and Wiznerowicz M. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemp Oncol (Pozn)* 2015; 19: A68-77. 2015/02/19. DOI: 10.5114/wo.2014.47136.
14. Barrett T, Wilhite SE, Ledoux P, et al. NCBI GEO: archive for functional genomics data sets–update. *Nucleic Acids Res* 2013; 41: D991-995. 2012/11/30. DOI: 10.1093/nar/gks1193.
15. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015; 43: e47. 2015/01/22. DOI: 10.1093/nar/gkv007.
16. Law CW, Chen Y, Shi W, et al. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol* 2014; 15: R29. 2014/02/04. DOI: 10.1186/gb-2014-15-2-r29.
17. Ernst J and Bar-Joseph Z. STEM: a tool for the analysis of short time series gene expression data. *BMC Bioinformatics* 2006; 7: 191. 2006/04/07. DOI: 10.1186/1471-2105-7-191.
18. Hanzelmann S, Castelo R and Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* 2013; 14: 7. 2013/01/18. DOI: 10.1186/1471-2105-14-7.
19. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 2011; 12: 77. 2011/03/19. DOI: 10.1186/1471-2105-12-77.
20. Yu G, Wang LG, Han Y, et al. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 2012; 16: 284-287. 2012/03/30. DOI: 10.1089/omi.2011.0118.
21. Colwill K, Renewable Protein Binder Working G and Graslund S. A roadmap to generate renewable protein binders to the human proteome. *Nat Methods* 2011; 8: 551-558. 2011/05/17. DOI: 10.1038/nmeth.1607.
22. Scott WJ, Howington J, Feigenberg S, et al. Treatment of non-small cell lung cancer stage I and stage II: ACCP evidence-based clinical practice guidelines (2nd edition). *Chest* 2007; 132: 234S-242S. 2007/10/06. DOI: 10.1378/chest.07-1378.
23. Cui Y, Liu J, Yin HB, et al. Fibulin-1 functions as a prognostic factor in lung adenocarcinoma. *Jpn J Clin Oncol* 2015; 45: 854-859. 2015/07/18. DOI: 10.1093/jjco/hyv094.
24. Giatromanolaki A, Kalamida D, Sivridis E, et al. Increased expression of transcription factor EB (TFEB) is associated with autophagy, migratory phenotype and poor prognosis in non-small cell lung cancer. *Lung Cancer* 2015; 90: 98-105. 2015/08/13. DOI: 10.1016/j.lungcan.2015.07.008.
25. Xu P, Liu L, Wang J, et al. Genetic variation in BCL2 3'-UTR was associated with lung cancer risk and prognosis in male Chinese population. *PLoS One* 2013; 8: e72197. 2013/08/27. DOI: 10.1371/journal.pone.0072197.
26. Li G, Zhong Y, Shen Q, et al. NEK2 serves as a prognostic biomarker for hepatocellular carcinoma. *Int J Oncol* 2017; 50: 405-413. 2017/01/20. DOI: 10.3892/ijo.2017.3837.
27. Fang Y and Zhang X. Targeting NEK2 as a promising therapeutic approach for cancer treatment. *Cell Cycle* 2016; 15: 895-907. 2016/03/29. DOI: 10.1080/15384101.2016.1152430.

28. Zhou W, Yang Y, Xia J, et al. NEK2 induces drug resistance mainly through activation of efflux drug pumps and is associated with poor prognosis in myeloma and other cancers. *Cancer Cell* 2013; 23: 48-62. 2013/01/19. DOI: 10.1016/j.ccr.2012.12.001.
29. Zhang Y, Yuan Y, Liang P, et al. Overexpression of a novel candidate oncogene KIF14 correlates with tumor progression and poor prognosis in prostate cancer. *Oncotarget* 2017; 8: 45459-45469. 2017/05/20. DOI: 10.18632/oncotarget.17564.
30. O'Hare M, Shadmand M, Sulaiman RS, et al. Kif14 overexpression accelerates murine retinoblastoma development. *Int J Cancer* 2016; 139: 1752-1758. 2016/06/09. DOI: 10.1002/ijc.30221.
31. Xu CH, Zhang Y and Yu LK. The diagnostic and prognostic value of serum human kallikrein-related peptidases 11 in non-small cell lung cancer. *Tumour Biol* 2014; 35: 5199-5203. 2014/02/11. DOI: 10.1007/s13277-014-1674-x.
32. Li X, Shi Y, Yin Z, et al. An eight-miRNA signature as a potential biomarker for predicting survival in lung adenocarcinoma. *J Transl Med* 2014; 12: 159. 2014/06/05. DOI: 10.1186/1479-5876-12-159.
33. Liu C, Li Y, Wei M, et al. Identification of a novel glycolysis-related gene signature that can predict the survival of patients with lung adenocarcinoma. *Cell Cycle* 2019; 18: 568-579. 2019/02/08. DOI: 10.1080/15384101.2019.1578146.
34. Shi X, Tan H, Le X, et al. An expression signature model to predict lung adenocarcinoma-specific survival. *Cancer Manag Res* 2018; 10: 3717-3732. 2018/10/06. DOI: 10.2147/CMAR.S159563.

Tables

Table 1 LUAD-unfavorable gene set and LUAD-favorable gene set

Gene set	Gene symbol
LUAD-unfavorable gene set	ARHGAP11A␣ASPM␣BLM␣C5orf34␣CA9␣CCNA2␣CDC25C␣CDC6␣CDCA2␣CDK1␣CENPF␣CENPK␣CHAF1B␣CLSPN␣DDX11-AS1␣DEPDC1␣DNMT3B␣DTL␣E2F7␣ECT2␣EGLN3␣ESCO2␣EXO1␣FAM111B␣FAM57B␣FAM72D␣FAM83D␣FANCI␣FBXO43␣GAL␣GTSE1␣HASPIN␣HELLS␣HMMR␣KIF11␣KIF14␣KIFC1␣KNL1␣KNTC1␣KREMEN2␣KRT6A␣KRT81␣LINC01269␣LOC101929128␣LYPD3␣MAD2L1␣MELK␣MIR924HG␣MKI67␣MYO19␣NCAPG␣NDC80␣NEIL3␣NEK2␣NUF2␣NUSAP1␣OIP5␣ORC1␣ORC6␣PAICS␣PARPBP␣PCLAF␣PIMREG␣PLK1␣PLK4␣POLQ␣PRC1␣PTPRN␣RAD51␣RRM2␣SGO1␣SLC2A1-AS1␣SPAG5␣SPOCK1␣TEDC2␣TESMIN␣TGFBR3L TICRR␣TROAP␣TTK␣TYMS␣UBE2T␣UCA1␣ZWINT
LUAD-favorable gene set	ACKR1␣ADAMTS8␣ADGRF5␣ARHGEF6␣ATP13A4␣BMP5␣CASS4␣CCDC69␣CLEC3B␣COL6A6␣CTSG␣FAM189A2␣FBP1␣FCER1A␣FLI1␣GCSAML␣GIMAP4␣GIMAP7␣HPGDS␣HPSE2␣INMT␣JAML␣KLK11␣LSAMP␣LY86␣MAL␣MS4A2␣OR7E47P␣P2RY12␣RAB44␣RTN1␣SCN2B␣SIGLEC17P␣SLCO4C1␣SPN␣TM6SF1␣TRPC2␣UNC45B␣ZEB2

Table 2 Univariate and multivariate analyses of LUAD-unfavorable GSVA score

Factor	Univariate Cox analysis			Multivariate Cox analysis		
	β	P	HR(95% CI)	β	P	HR(95% CI)
Gender (female / male)	0.025	0.867	0.763-1.378			
Age (>65years/ >65 years)	0.178	0.243	0.886-1.610			
T stage (T3-4 / T1-2)	0.821	0.000	1.543-3.346	0.589	0.016	1.114-2.914
Lymph node stage (N2-3 / N0-1)	0.818	0.000	1.582-3.243	0.121	0.757	0.523-2.437
Metastasis (M1 / M0)	0.749	0.006	1.234-3.626	0.109	0.799	0.482-2.580
Pathological stage (III-IV / I-II)	0.967	0.000	1.924-3.592	0.509	0.211	0.749-3.695
LUAD-unfavorable GSVA score (High / Low)	0.614	0.000	1.365-2.500	0.575	0.002	1.230-2.565

Table 3 Univariate and multivariate analyses of LUAD-favorable GSVA score

Factor	Univariate Cox analysis			Multivariate Cox analysis		
	β	P.Value	HR(95% CI)	β	P.Value	HR(95% CI)
Gender (female / male)	0.025	0.867	0.763-1.378			
Age(>65 years/<=65 years)	0.178	0.243	0.886-1.610			
T stage (T3-4/ T1-2)	0.821	0.000	1.543-3.346	0.557	0.028	1.062-2.869
Lymph node stage (N2-3/N0-1)	0.818	0.000	1.582-3.243	0.420	0.254	0.739-3.134
Metastasis (M1/ M0)	0.749	0.006	1.234-3.626	0.263	0.518	0.586-2.886
Pathological stage (III-IV/I-II)	0.967	0.000	1.924-3.592	0.365	0.361	0.659-3.152
LUAD-favorable GSVA score (High / Low)	0.614	0.000	1.365-2.500	-0.434	0.017	0.453-0.926

Figures

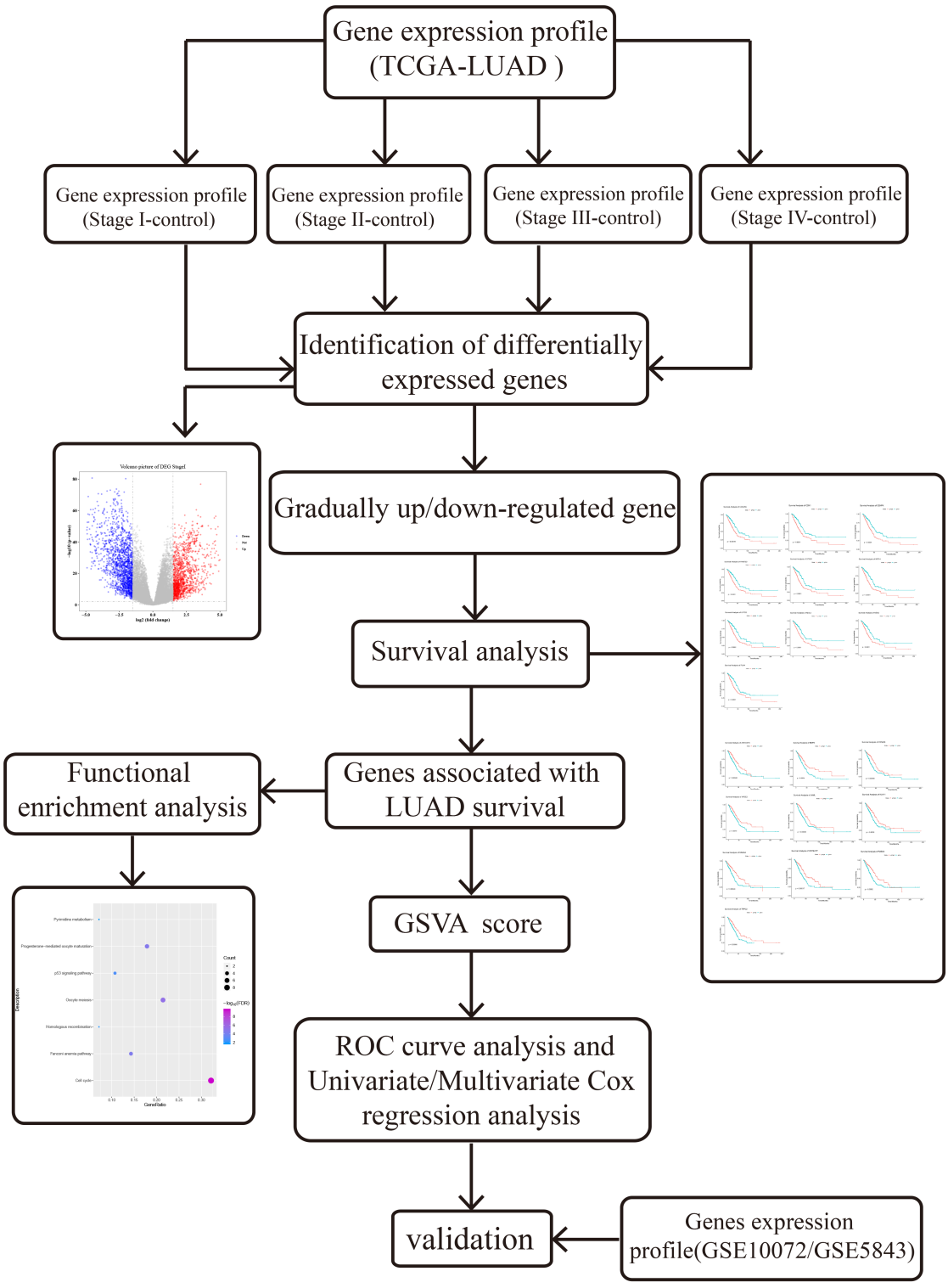

Figure 1

Flow Chart of this study

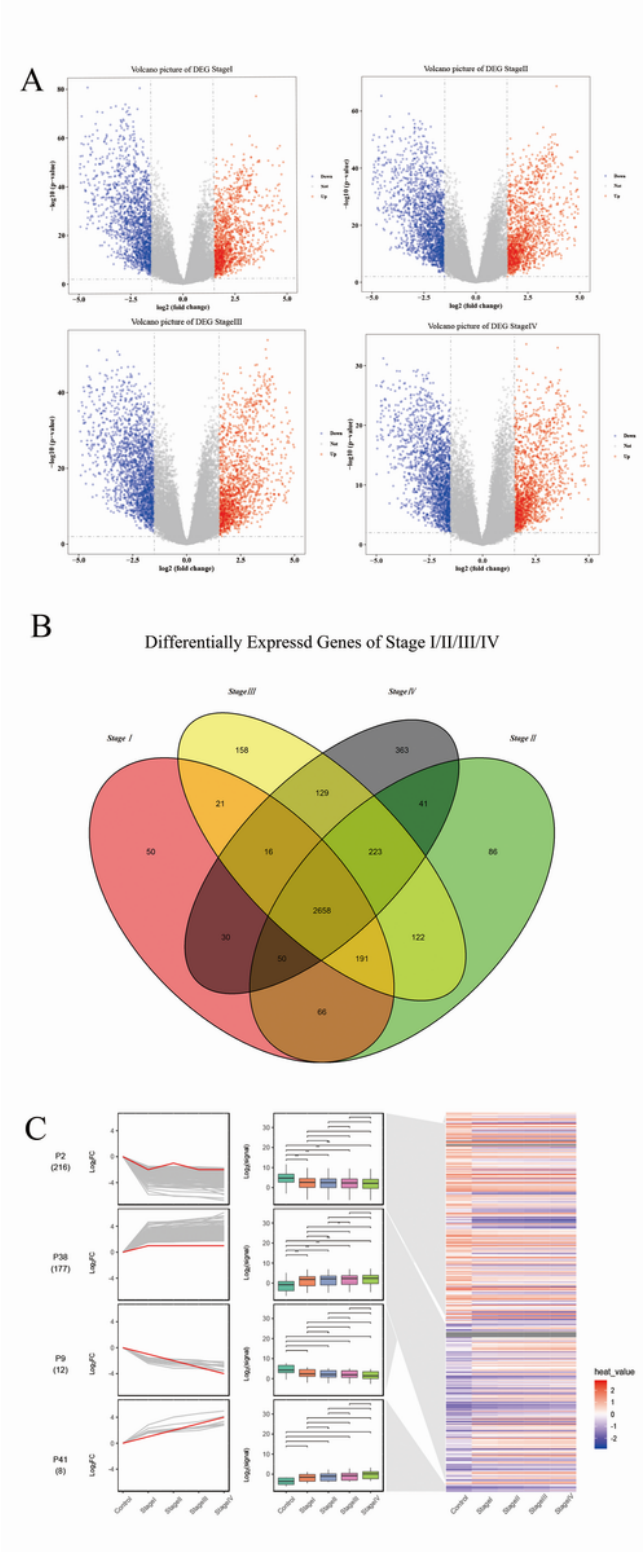

Figure 2

Differential expression gene analysis and STEM. A, LUAD State I、State II、State III、State IV volcano figure、red represents up-regulated genes, blue represents down-regulated genes, and gray represents no significantly differentially expressed genes; B. Common differentially expressed genes in LUAD stage I-IV; C, STEM results. Line plots (left panels) and box plots (right panels) are used to show fold changes (log2

scale) and absolute expression levels (log2 scale), respectively. In each line plot, one representative gene is highlighted in red.

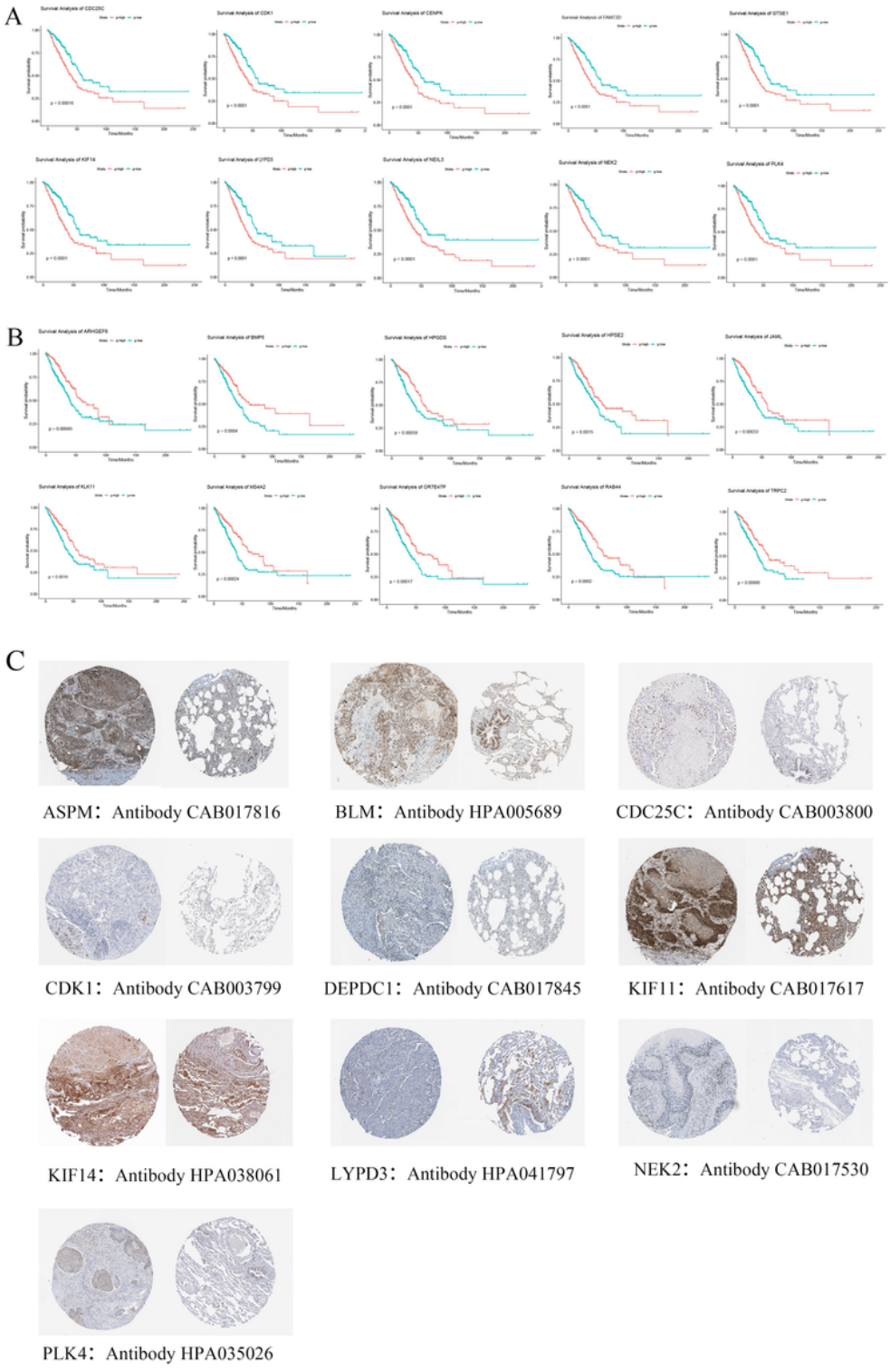

Figure 3

Survival analysis A, Survival curves of 10 genes most significantly correlated with LUAD in LUAD-unfavorable gene set. B, Survival curves of 10 genes most significantly correlated with LUAD in LUAD-favorable gene set. C, The up-regulation of LUAD-unfavorable gene was validated using

immunohistochemistry, lung cancer samples are on the left and normal lung tissue samples are on the right.

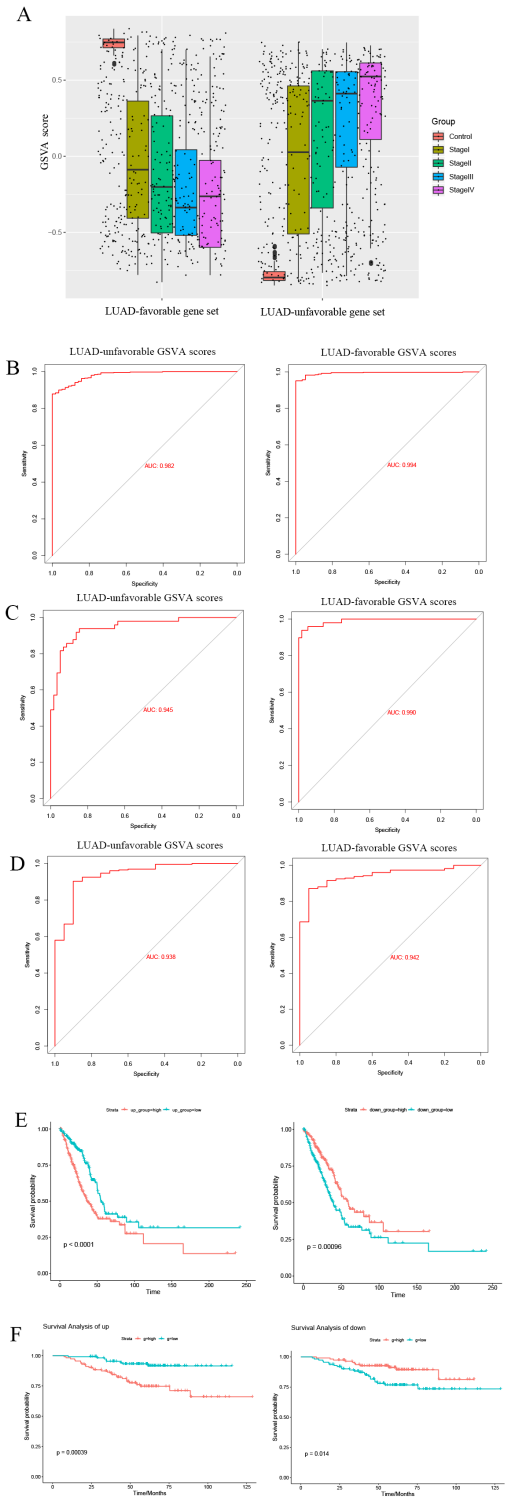

Figure 4

Evaluating the diagnostic and prognostic abilities of LUAD-unfavorable gene set and LUAD-favorable gene set. A, The box plots of LUAD-unfavorable gene set GSVA score and LUAD-favorable gene set GSVA score. B, ROC curves of LUAD-unfavorable gene set GSVA score and LUAD-favorable gene set GSVA

score. C, ROC curves of LUAD-unfavorable GSVA score and LUAD-favorable GSVA score in GSE10072. D, ROC curves of LUAD-unfavorable GSVA score and LUAD-favorable GSVA score in GSE31210. E, Survival analysis of LUAD-unfavorable gene set GSVA score and LUAD-favorable gene set GSVA score in LUAD from TCGA. F, Survival analysis of LUAD-unfavorable GSVA score and LUAD-favorable GSVA score in GSE31210.

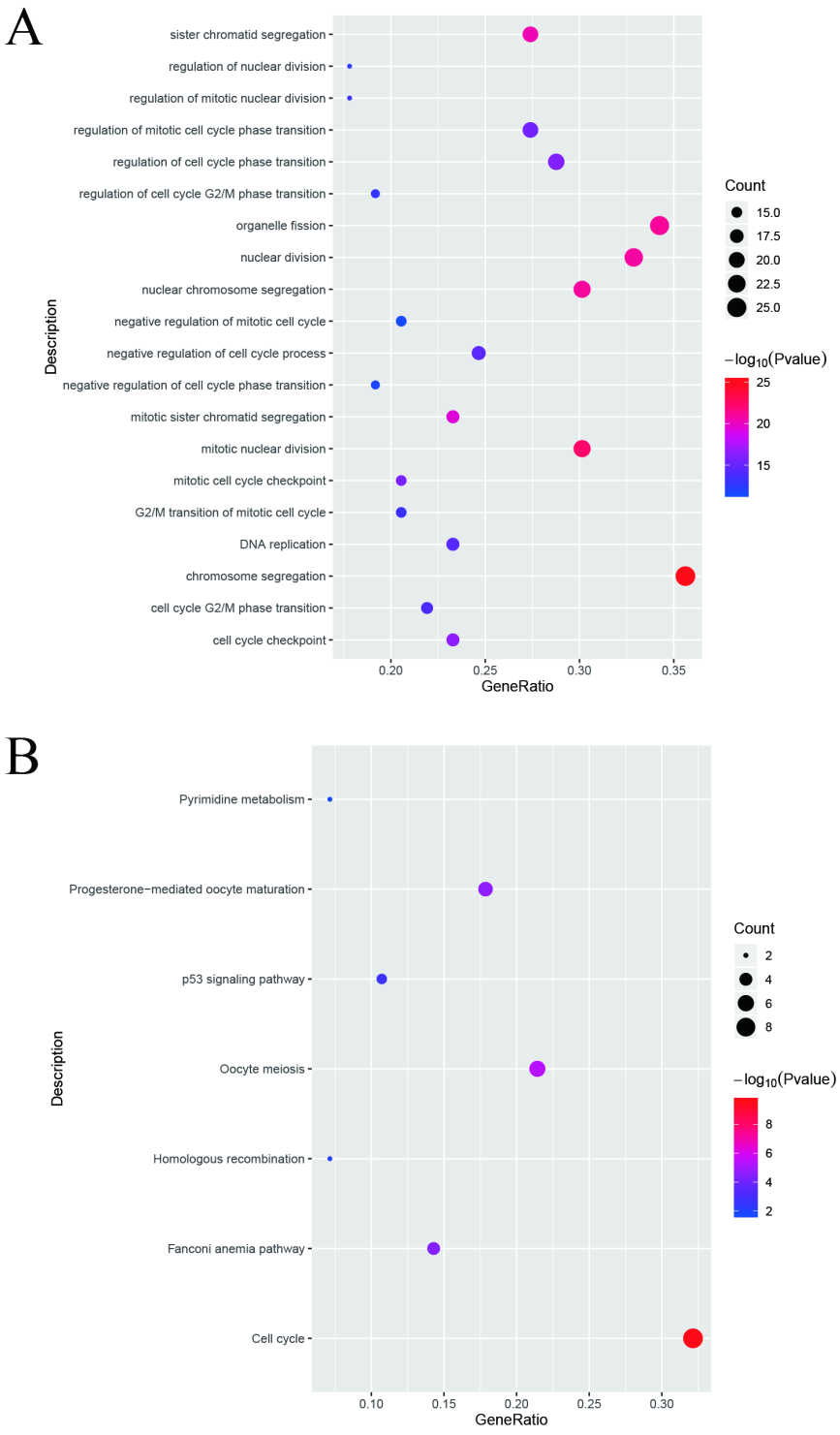

Figure 5

Annotation of the statistical significant enriched Gene Ontology and KEGG pathways of LUAD-unfavorable genes. A. Biological process. B, KEGG pathway analysis.