

SCIENTIFIC REPORTS



OPEN

Prognostic significance of TCF21 mRNA expression in patients with lung adenocarcinoma

Jian Xiao¹, Aibin Liu², Xiaoxiao Lu¹, Xi Chen³, Wei Li⁴, Shuya He⁵, Bixiu He¹ & Qiong Chen¹

Several prognostic indicators have shown inconsistencies in patients of different genders with lung adenocarcinoma, indicating that these variations may be due to the different genetic background of males and females with lung adenocarcinoma. In this study, we first used the Gene-Cloud of Biotechnology Information (GCBi) bioinformatics platform to identify differentially expressed genes (DEGs) that eliminated gender differences between lung adenocarcinoma and normal lung tissues. Then, we screened out that transcription factor 21 (TCF21) is a hub gene among these DEGs by creating a gene co-expression network on the GCBi platform. Furthermore, we used the comprehensive survival analysis platforms Kaplan-Meier plotter and Prognoscan to assess the prognostic value of TCF21 expression in lung adenocarcinoma patients. Finally, we concluded that decreased mRNA expression of TCF21 is a predictor for poor prognosis in patients with lung adenocarcinoma.

Lung cancer is a malignant tumor characterized by uncontrolled cell growth in tissues of the lung and bronchus. Except for prostate cancer in males and breast cancer in females, lung cancer is the second most commonly diagnosed cancer in men and women¹. It is the most common cause of cancer death in humans and accounts for more than one-quarter of all cancer deaths¹. Moreover, the overall age-adjusted 5-year relative survival for lung cancer is 15.5% in males and 20.3% in females². Therefore, further researches are needed to develop treatments for lung cancer, including the identification of prognostic factors for lung cancer patients.

Non-small cell lung cancer (NSCLC) accounts for 85% of total lung cancer cases³. According to the latest statistics from the SEER Cancer Statistics Review, from 2006 to 2012, the average 5-year relative survival rate of NSCLC in males is 19.2% and in females is 26.8%⁴. The major histological phenotypes of NSCLC are classified as adenocarcinoma, squamous cell carcinoma and large cell carcinoma. However, lung adenocarcinoma (approximately 50% of cases) is the most important pathological type of NSCLC⁵. In recent years, many prognostic factors for the survival of lung adenocarcinoma patients have been reported by researchers^{6–12}. However, some of these indicators have inconsistent prognostic value in patients with lung adenocarcinoma according to different gender^{13–17}. We hypothesized that one of the causes of this inconsistency is due to the different genetic background of lung adenocarcinoma in males and females.

In this study, we used the Gene-Cloud of Biotechnology Information (GCBi) bioinformatics platform to identify differentially expressed genes (DEGs) that eliminated gender differences between lung adenocarcinoma and normal lung tissues. Among these DEGs, we found that transcription factor 21 (TCF21) is a hub gene. Furthermore, using the comprehensive survival analysis platforms Kaplan-Meier plotter and Prognoscan, we showed that the decreased mRNA expression of TCF21 is an unfavorable prognostic factor for lung adenocarcinoma patients.

Results

Major sample characteristics. We used two Gene Expression Omnibus (GEO) datasets (GSE40791¹⁸ and GSE10072¹⁹) in this study. Both of these datasets were included in the GCBi bioinformatics analysis platform. GSE40791 included 94 lung adenocarcinoma frozen tissues (53 derived from males, 41 derived from females) and 100 non-neoplastic lung samples (58 derived from males, 42 derived from females) (Supplementary Table S1).

¹Department of Geriatrics, Respiratory Medicine, Xiangya Hospital of Central South University, Changsha, China.

²Department of Geriatrics, Xiangya Hospital of Central South University, Changsha, China. ³Department of Respiratory Medicine, Xiangya Hospital of Central South University, Changsha, China. ⁴Department of Geriatrics, Clinical Laboratory, Xiangya Hospital of Central South University, Changsha, China. ⁵Department of Biochemistry & Biology, University of South China, Hengyang, China. Correspondence and requests for materials should be addressed to Q.C. (email: qiongch@163.com)

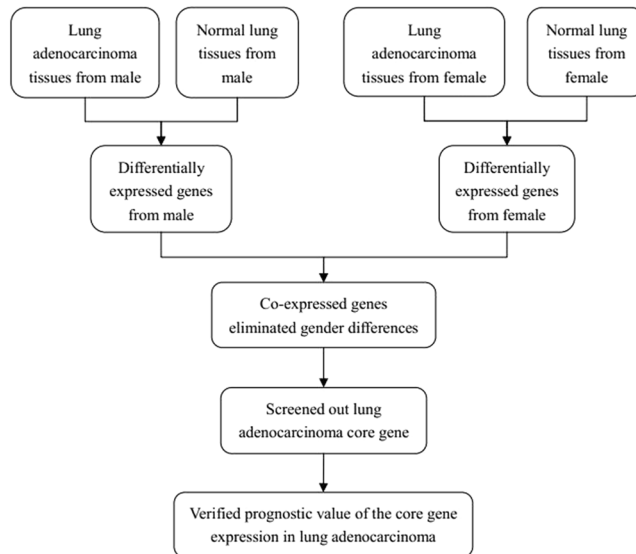


Figure 1. Flow diagram of the study design.

GSE10072 included 58 tumor samples (35 derived from males, 23 derived from females) and 49 non-tumor tissues (34 derived from males, 15 derived from females) (Supplementary Table S2).

Research design. In our study, we wanted to identify the DEGs that eliminated gender differences between lung adenocarcinoma and normal lung tissues. Therefore, we first divided both lung adenocarcinoma and non-neoplastic lung samples into two groups according to gender. Using the GCBI bioinformatics analysis platform, we then obtained two DEG datasets from males and females. Then, by selecting the intersection of these two datasets, we obtained the co-expressed DEGs that eliminated gender differences. Next, we created a gene co-expression network to screen out hub genes in the co-expressed DEGs. Finally, we verified the prognostic value of core gene expression for patients with lung adenocarcinoma using the comprehensive survival analysis platforms Kaplan-Meier plotter and PrognScan. The flow diagram of our study design is shown in Fig. 1.

TCF21 is the core gene in the DEGs obtained from GSE40791. We identified 321 potential DEGs from males (Fig. 2a and Supplementary Table S3) and 463 potential DEGs from females (Fig. 2b and Supplementary Table S4). To remove duplicate expressions from the same gene and expressions from unannotated genes, 223 DEGs (Supplementary Table S5) and 330 DEGs (Supplementary Table S6) were selected from the male and female groups. By taking the intersection of these two DEGs datasets, 211 co-expressed DEGs that eliminated gender differences were found (Fig. 2c). Among them, 131 DEGs were incorporated into the gene co-expression network (Fig. 3a). As shown in Table 1 and Fig. 3b, we found that TCF21 had the most gene connections in the gene co-expression network. Therefore, we hypothesized that TCF21 is the core gene in the DEGs between lung adenocarcinoma and normal lung tissues.

TCF21 is also the core gene in the DEGs obtained from GSE10072. To verify the results obtained from GSE40791, we then used the same process in GCBI to study the DEGs between lung adenocarcinoma and normal lung tissues from GSE10072. We identified 76 potential DEGs from males (Fig. 4a and Supplementary Table S7) and 87 potential DEGs from females (Fig. 4b and Supplementary Table S8). To remove duplicate expressions from the same gene and expressions from unannotated genes, 52 DEGs (Supplementary Table S9) and 65 DEGs (Supplementary Table S10) were selected from the male and female groups, respectively. By taking the intersection of these two DEGs datasets, 46 co-expressed DEGs that eliminated gender differences were found (Fig. 4c). Among them, 16 DEGs were incorporated into the gene co-expression network (Fig. 5a). As shown in Table 2 and Fig. 5b, we again found that TCF21 has the most gene connections in the gene co-expression network. Consequently, we confirmed that TCF21 is the core gene in the DEGs between lung adenocarcinoma and normal lung tissues.

Decreased mRNA expression of TCF21 is an unfavorable prognostic factor for lung adenocarcinoma. Compared with non-neoplastic lung tissues, TCF21 was decreased in the lung adenocarcinoma samples (Supplementary Tables S3–S10). Furthermore, TCF21 is the core gene in the DEGs between lung adenocarcinoma and normal lung tissues (Figs 3 and 5). Accordingly, we hypothesized that TCF21 expression may be a prognostic factor for patients with lung adenocarcinoma. Using the comprehensive Kaplan-Meier survival analysis platform, we discovered that decreased mRNA expression of TCF21 is an unfavorable prognostic factor of overall survival for patients with lung adenocarcinoma (hazard ratio, HR = 0.50; 95% confidence interval, 95% CI: 0.39–0.65; $P = 4.7e^{-08}$; $n = 673$) (Fig. 6a). When further considering the gender difference, decreased mRNA expression of TCF21 has similar prognostic value in both males (HR = 0.57; 95% CI: 0.41–0.81; $P = 0.0014$; $n = 328$) and females (HR = 0.47; 95% CI: 0.31–0.72; $P = 0.00045$; $n = 287$) (Fig. 6b and c). Simultaneously, results

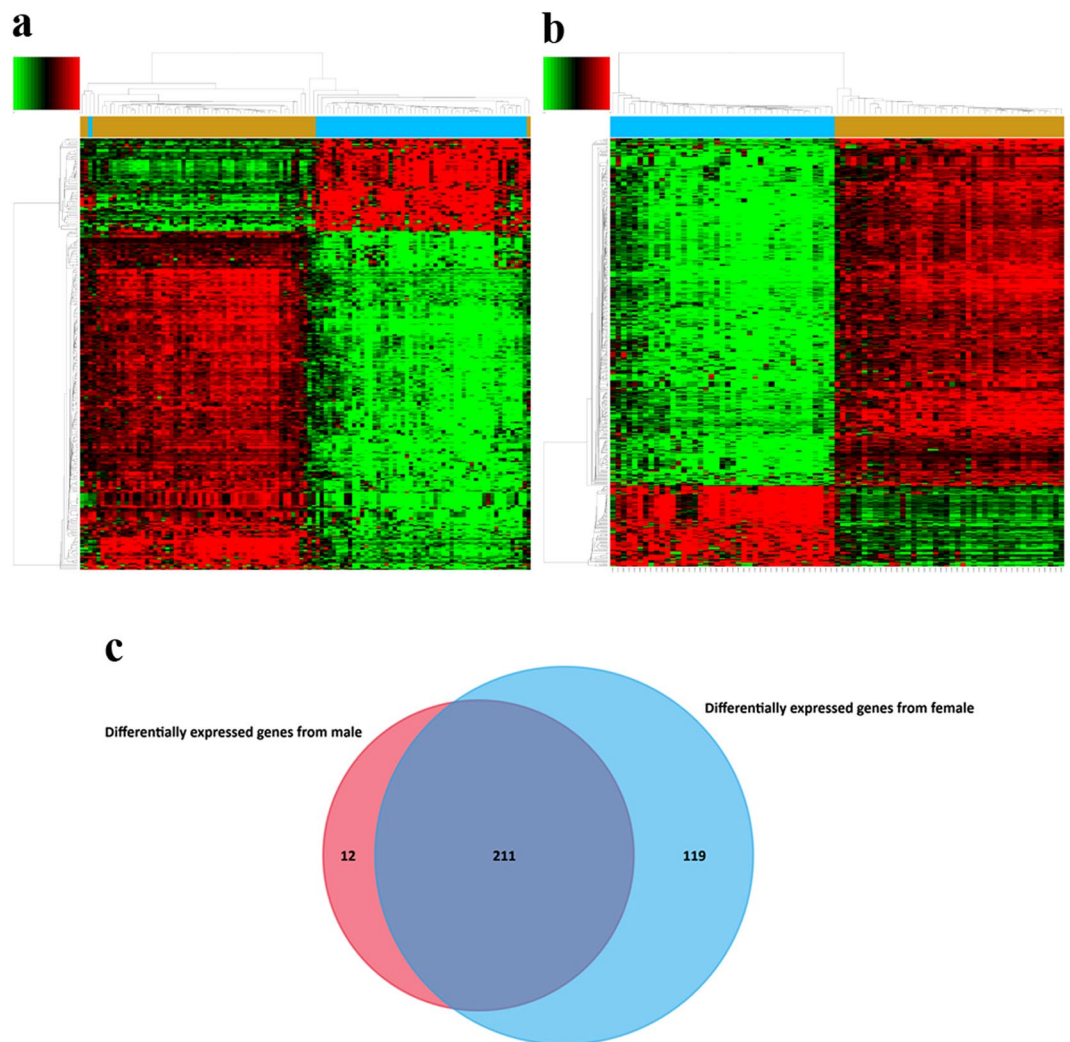


Figure 2. DEGs between lung adenocarcinoma and normal lung tissues for GSE40791. (a) Heat map for potential DEGs (n = 321) from males (lung adenocarcinoma n = 53, in blue; normal lung tissues n = 58, in yellow). (b) Heat map for potential DEGs (n = 463) from females (lung adenocarcinoma n = 41, in blue; normal lung tissues n = 42, in yellow). (c) Duplicate expressions with the same gene symbol and those without a specific gene symbol were removed; the Venn diagram shows DEGs (n = 223) from males, DEGs (n = 330) from females and the co-expressed genes (n = 211) that eliminated gender differences.

generated from PrognScan based on the GEO dataset of GSE31210 indicated that decreased mRNA expression of TCF21 is not only an unfavorable prognostic factor for lung adenocarcinoma patients' relapse-free survival (HR = 0.59; 95% CI: 0.46–0.76; P = 0.000032; n = 204) (Fig. 7a) but also for overall survival (HR = 0.71; 95% CI: 0.55–0.91; P = 0.006584; n = 204) (Fig. 7b). However, also using the comprehensive survival analysis platforms Kaplan-Meier plotter and PrognScan, we found that TCF21 expression has no prognostic value in patients with lung squamous cell carcinoma (data not shown).

Discussion

To assess the prognostic value of gene expression for lung cancer patient survival, the expression of the genes was normally detected by immunohistochemistry in protein level^{20–22}. However, the expression level of mRNA detected by polymerase chain reaction also performed reliable prognostic value for lung cancer patients^{23,24}. In this study, we extracted gene expression array data from GEO datasets using the GCBi bioinformatics analysis platform and further assessed these data using the comprehensive survival analysis platform Kaplan-Meier plotter and PrognScan. We finally found that decreased TCF21 mRNA expression is an unfavorable prognostic factor for patients with lung adenocarcinoma. Our study indicated that comprehensive utilization of bioinformatics analyses is a good strategy to study prognosis of lung adenocarcinoma and may be useful to assess prognosis for other cancer types.

Previous studies did not consider the effect of gender differences in DEGs between lung and normal tissues at the design stage. Because of the different genetic background between males and females, we considered that lung adenocarcinomas with different gender origins may also have significant genetic difference. In our study,

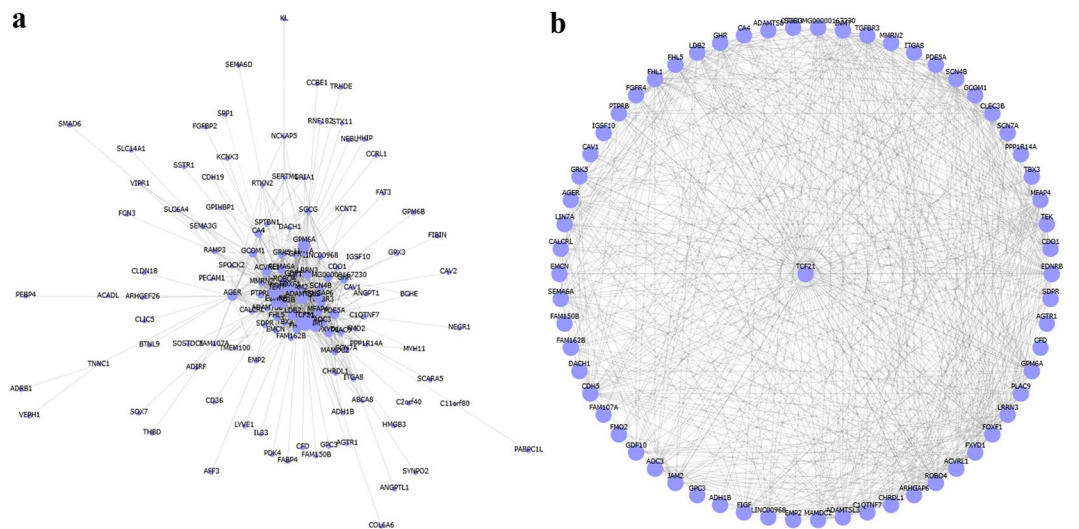


Figure 3. Gene co-expression network of the DEGs that eliminated gender differences for GSE40791. **(a)** The gene co-expression network. **(b)** Sub-graph shows that TCF21 has the most gene connections in the gene co-expression network.

Gene Symbol	Degree	Description
TCF21	62	transcription factor 21
GPM6A	49	glycoprotein M6A
TEK	46	TEK tyrosine kinase, endothelial
ADAMTSL3	40	ADAMTS-like 3
FOXF1	39	forkhead box F1
CDH5	38	cadherin 5, type 2 (vascular endothelium)
LDB2	38	LIM domain binding 2
EDNRB	37	endothelin receptor type B
ARHGAP6	36	Rho GTPase activating protein 6
ROBO4	35	roundabout, axon guidance receptor, homolog 4 (Drosophila)

Table 1. Top 10 hub genes in the gene co-expression network of the DEGs that eliminated gender differences for GSE40791.

to eliminate the impact of gender differences, we divided both lung adenocarcinoma and non-neoplastic lung samples into two groups according to gender. Using comprehensive bioinformatics analyses, we reported that the decreased mRNA expression of TCF21 is an unfavorable prognostic factor for patients with lung adenocarcinoma. To our knowledge, this is the first study to investigate the prognostic value of TCF21 in patients with this specific lung cancer type.

TCF21 (also known as Capsulin, Pod-1 or Epicardin) is a member of the basic-helix-loop-helix transcription factor family^{25–27}. During embryogenesis, TCF21 is crucial for the development of a number of cell types in the spleen²⁸, heart²⁹, kidney³⁰ and lung³⁰. However, TCF21 is also important in cancer development. *In vitro*, it can affect tumor cell cycle balance³¹ and suppress cancer cell proliferation, migration³² and invasion^{32,33}. Additionally, mouse model experiments showed that TCF21 can significantly reduce tumor growth *in vivo*^{34,35}. Consequently, TCF21 is considered to be a potential tumor suppressor^{34,36}.

Unfortunately, TCF21 expression is commonly deregulated by promoter hypermethylation in different types of cancer^{32,33,37–40}. This phenomenon is detrimental for cancer patients. Study reported that TCF21 promoter methylation is correlated with decreased survival in patients with metastatic skin melanoma³³. Low expression of TCF21 is an independent prognostic factor for poor survival in patients with clear cell renal cell carcinoma⁴¹ or gastric cancer⁴².

For lung cancer, previous studies demonstrated that aberrant TCF21 promoter methylation existed in 9 of 10 lung cancer cell lines³⁹ and the majority of lung cancer tissues (70–81%)^{34,38,39}. Richards KL and colleagues³⁸ reported that 84% of non-small cell lung cancer samples showed decreased TCF21 protein expression. TCF21 expression was not correlated with gender and was lower in adenocarcinoma than in squamous cell carcinoma³⁸, indicating that TCF21 is a crucial tumor suppressor for lung adenocarcinoma in both males and females. We corroborated these findings in the present study. Recently, Wu H and colleagues³⁵ found that low TCF21 expression

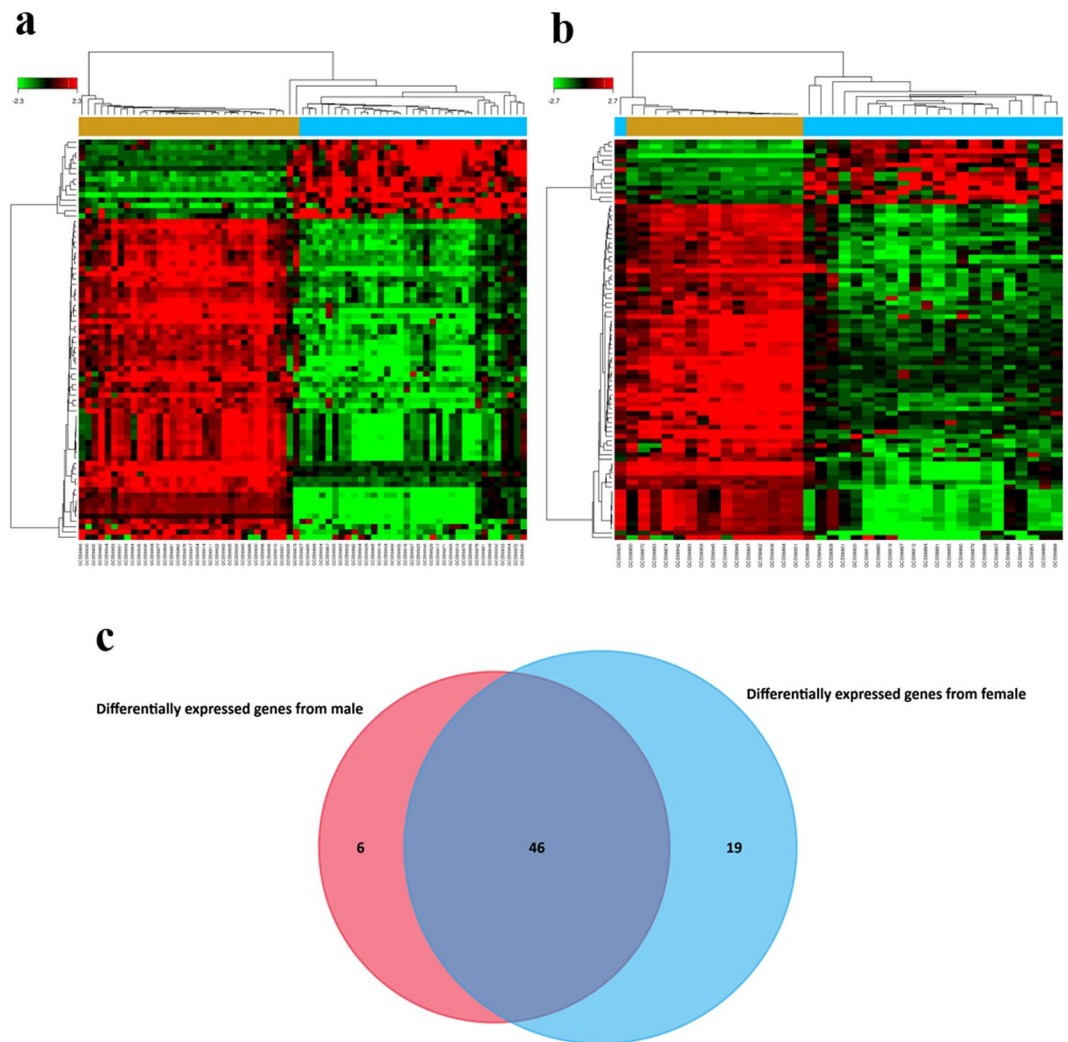


Figure 4. DEGs between lung adenocarcinoma and normal lung tissues for GSE10072. (a) Heat map for potential DEGs ($n = 76$) from males (lung adenocarcinoma $n = 35$, in blue; normal lung tissues $n = 34$, in yellow). (b) Heat map for potential DEGs ($n = 87$) from females (lung adenocarcinoma $n = 23$, in blue; normal lung tissues $n = 15$, in yellow). (c) Duplicate expressions with the same gene symbol and those without a specific gene symbol were removed; the Venn diagram shows DEGs ($n = 52$) from males, DEGs ($n = 65$) from females and the co-expressed genes ($n = 46$) that eliminated gender differences.

is associated with poor survival for lung cancer using the survival analysis platform Kaplan-Meier plotter. In our study, we comprehensively used both Kaplan-Meier plotter and PrognScan to first conclude that decreased mRNA expression of TCF21 is an unfavorable prognostic factor in lung adenocarcinoma patients without gender difference and that TCF21 expression showed no prognostic value in patients with lung squamous cell carcinoma. We consider that maybe TCF21 is a specific prognostic factor in lung adenocarcinoma rather than in lung squamous cell carcinoma, for which this point has not taken into account in previous studies.

We know that TCF21 is a crucial tumor suppressor, and TCF21 expression is commonly deregulated by aberrant promoter methylation in different types of cancer. Additionally, decreased expression of TCF21 is a poor prognostic factor for cancer patients. However, we know little about the tumor suppressor mechanism of TCF21. Therefore, in the future, more studies should be conducted to elucidate the mechanisms of TCF21 tumor suppressor function. In addition, the bioinformatics analysis platforms of GCBI, Kaplan-Meier plotter and PrognScan used in this study were based on the method of unsupervised analysis of gene expression profiles (named the “one-step-clustering” approach by Li J *et al.*⁴³). Because of the extremely high variability in gene expression profiles between individual tumors and because “passenger signals” may mask the “real” cancer gene signals, our results derived from this “one-step-clustering” approach may be less robust and accurate^{43,44}.

In summary, we first used the GCBI bioinformatics analysis platform to identify DEGs that eliminated gender differences between lung adenocarcinoma and normal lung tissues, which showed that TCF21 is the hub gene. Then, using the comprehensive survival analysis platforms of Kaplan-Meier plotter and PrognScan, we concluded that decreased mRNA expression of TCF21 is a poor prognostic factor in patients with lung adenocarcinoma.

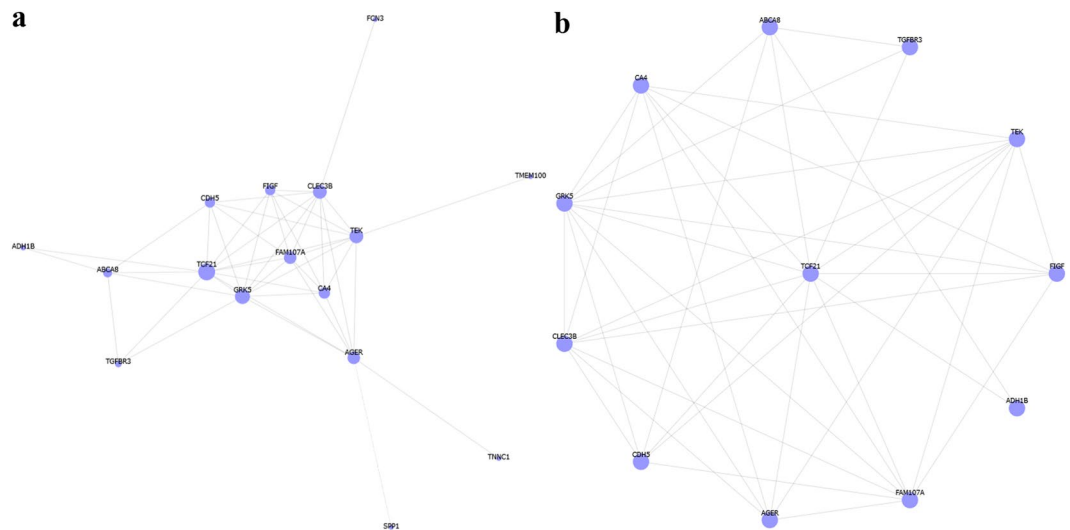


Figure 5. Gene co-expression network of the DEGs that eliminated gender differences for GSE10072. **(a)** The gene co-expression network. **(b)** The sub-graph shows that TCF21 has the most gene connections in the gene co-expression network.

Gene Symbol	Degree	Description
TCF21	11	transcription factor 21
GRK5	10	G protein-coupled receptor kinase 5
CLEC3B	9	C-type lectin domain family 3, member B
TEK	9	TEK tyrosine kinase, endothelial
FAM107A	8	family with sequence similarity 107, member A
AGER	8	advanced glycosylation end product-specific receptor
CA4	7	carbonic anhydrase IV
FIGF	6	c-fos induced growth factor (vascular endothelial growth factor D)
CDH5	6	cadherin 5, type 2 (vascular endothelium)
ABCA8	5	ATP-binding cassette, sub-family A (ABC1), member 8

Table 2. Top 10 hub genes in the gene co-expression network of the DEGs that eliminated gender differences for GSE10072.

Methods

Gene-Cloud of Biotechnology Information (GCBI). GCBI (Shanghai, China, <https://www.gcbi.com.cn>) is a platform that combines a variety of research findings, genetic information, sample information, data algorithms and bioinformatics to create a “gene knowledge base,” which encompasses biology, medicine, informatics, computer science, mathematics, graphics and other disciplines. GCBI includes more than 120 million copies of genomic samples, approximately 90,000 copies of tumor samples and more than 17 million copies of genetic information. Therefore, GCBI is a good bioinformatics analysis platform and has provided data analysis support for many studies on cancer research^{45–49}. In this study, we used GCBI to identify DEGs between lung adenocarcinoma and normal lung tissues and finally screened out that TCF21 is the hub gene in the DEG co-expression network. In the Differential Gene Expression Analysis module on the GCBI platform (Supplementary File), we chose a fold expression change >5 at cutoff values $Q < 0.05$ and $P < 0.05$ to screen out DEGs. Then, we selected the Gene Co-expression Network module on the GCBI platform to create a gene co-expression network for the DEGs that eliminated gender differences.

Gene Expression Omnibus (GEO) DataSets. GEO DataSets (<https://www.ncbi.nlm.nih.gov/gds>) is the public repository for storing high throughput gene expression datasets at National Center of Biotechnology Information. In this study, we selected datasets according to the following inclusion criteria: (1) Human lung cancer specimens containing the pathological type of adenocarcinoma; (2) Normal lung tissues used as the controls; (3) Specimens had gender information; (4) Number of samples no less than 100; (5) Expression profiling by

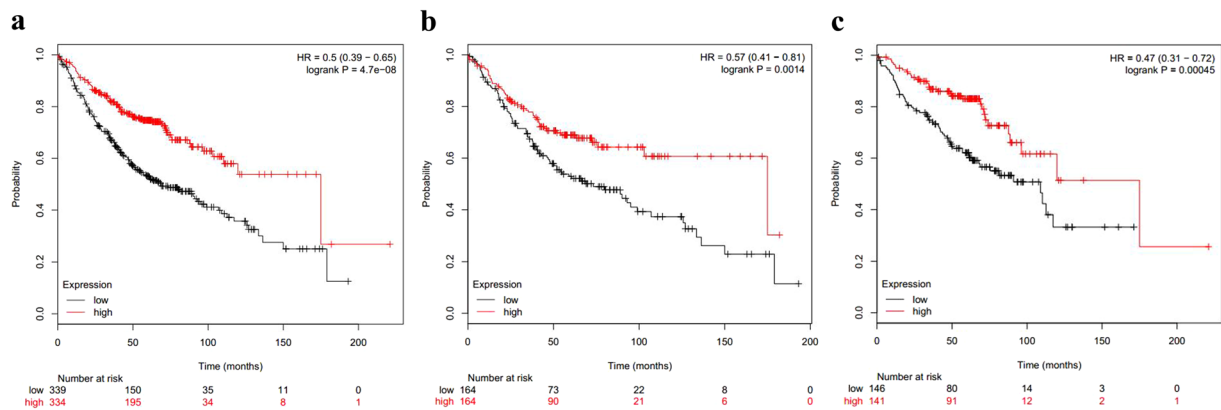


Figure 6. Kaplan–Meier survival curves generated from Kaplan-Meier plotter for TCF21 mRNA expression in patients with lung adenocarcinoma. Overall survival curves for lung adenocarcinoma patients that ignored gender differences (a), in male patients (b) and in female patients (c). HR = hazard ratio.

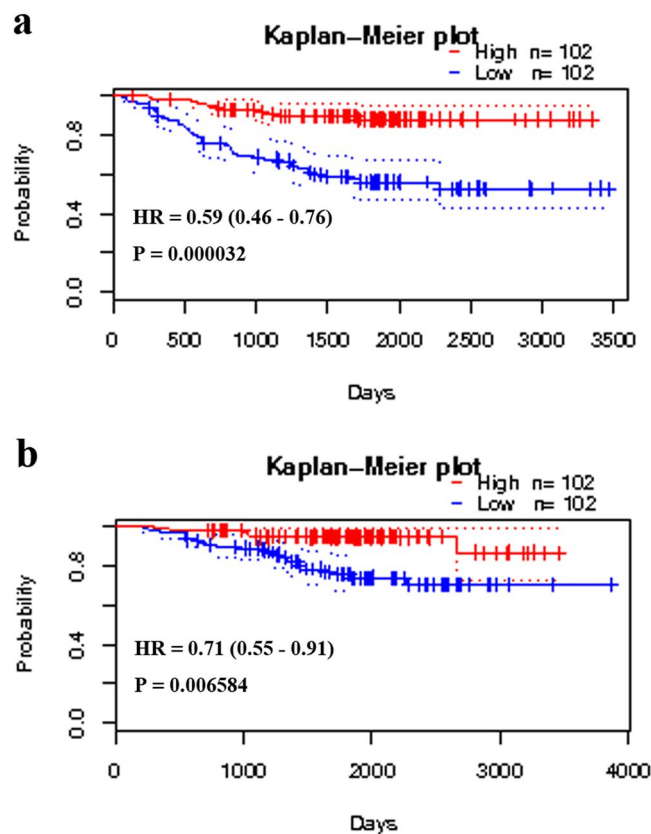


Figure 7. Kaplan–Meier survival curves generated from Prognoscan for TCF21 mRNA expression in patients with lung adenocarcinoma. Relapse-free survival curves (a) or overall survival curves (b) for lung adenocarcinoma patients in the GEO dataset GSE31210. HR = hazard ratio.

array and raw data had the CEL format. First, we used the search terms “Lung cancer [MeSH Terms] AND Homo sapiens [Organism] AND Adenocarcinoma [Description] CEL AND [Supplementary Files] AND Expression profiling by array [DataSet Type]” in the GEO DataSets to identify potential datasets. Then, we further screened these datasets according to the above inclusion criteria. Finally, two GEO datasets, GSE40791 and GSE10072, were included into our study.

Kaplan-Meier plotter. The Kaplan-Meier plotter (<http://kmplot.com/analysis/>) is a comprehensive online platform that can assess the effect of 54,675 genes on survival based on 10,293 cancer samples. These include 2,437 lung cancer patients with a mean follow-up of 49 months. With a meta-analysis of published microarray

datasets, this online tool is suitable for in silico validation of new biomarkers related to survival for patients with non-small cell lung cancer⁵⁰. We used the Kaplan-Meier plotter to assess the prognostic value of TCF21 expression in patients with lung adenocarcinoma or squamous cell carcinoma.

Prognoscan. Prognoscan (<http://www.prognoscan.org/>) is a comprehensive platform for evaluating potential tumor biomarkers and therapeutic targets. Based on a large collection of cancer microarray datasets with clinical annotation, Prognoscan is a useful online tool for assessing the association between specific gene expression and prognosis in patients with cancer⁵¹. We used the Prognoscan platform to validate the prognostic value of TCF21 expression in patients with lung adenocarcinoma or squamous cell carcinoma.

References

- Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2016. *CA Cancer J Clin* **66**, 7–30, doi:10.3322/caac.21332 (2016).
- Eberle, A. *et al.* Lung cancer survival in Germany: A population-based analysis of 132,612 lung cancer patients. *Lung Cancer* **90**, 528–33, doi:10.1016/j.lungcan.2015.10.007 (2015).
- Molina, J. R., Yang, P., Cassivi, S. D., Schild, S. E. & Adjei, A. A. Non-small cell lung cancer: epidemiology, risk factors, treatment, and survivorship. *Mayo Clin Proc* **83**, 584–94, doi:10.4065/83.5.584 (2008).
- Howlander, N. *et al.* SEER Cancer Statistics Review, 1975–2013, National Cancer Institute. Bethesda, MD, http://seer.cancer.gov/csr/1975_2013/, based on November 2015 SEER data submission, posted to the SEER web site, April 2016.
- Chen, Z., Fillmore, C. M., Hammerman, P. S., Kim, C. F. & Wong, K. K. Non-small-cell lung cancers: a heterogeneous set of diseases. *Nat Rev Cancer* **14**, 535–46, doi:10.1038/nrc3775 (2014).
- Hoda, M. A. *et al.* High circulating activin A level is associated with tumor progression and predicts poor prognosis in lung adenocarcinoma. *Oncotarget* **7**, 13388–99, doi:10.18632/oncotarget.7796 (2016).
- Li, C. *et al.* SIRT1 expression is associated with poor prognosis of lung adenocarcinoma. *Onco Targets Ther* **8**, 977–84, doi:10.2147/OTT.S82378 (2015).
- Li, P. *et al.* Long Noncoding RNA RGMB-AS1 Indicates a Poor Prognosis and Modulates Cell Proliferation, Migration and Invasion in Lung Adenocarcinoma. *PLoS One* **11**, e0150790, doi:10.1371/journal.pone.0150790 (2016).
- Thomas, A. *et al.* High mesothelin expression in advanced lung adenocarcinoma is associated with KRAS mutations and a poor prognosis. *Oncotarget* **6**, 11694–703, doi:10.18632/oncotarget.3429 (2015).
- Xu, P. *et al.* Elevated nuclear CCND1 expression confers an unfavorable prognosis for early stage lung adenocarcinoma patients. *Int J Clin Exp Pathol* **8**, 15887–94 (2015).
- Zheng, Y. Z. *et al.* ROR1 is a novel prognostic biomarker in patients with lung adenocarcinoma. *Sci Rep* **6**, 36447, doi:10.1038/srep36447 (2016).
- Shi, Y. X. *et al.* Prognostic and predictive values of CDK1 and MAD2L1 in lung adenocarcinoma. *Oncotarget* **7**, 85235–85243, doi:10.18632/oncotarget.13252 (2016).
- Goto, A. *et al.* Loss of TSLC1 expression in lung adenocarcinoma: relationships with histological subtypes, sex and prognostic significance. *Cancer Sci* **96**, 480–6, doi:10.1111/j.1349-7006.2005.00075.x (2005).
- Okano, Y. *et al.* SNP (−617C > A) in ARE-like loci of the NRF2 gene: a new biomarker for prognosis of lung adenocarcinoma in Japanese non-smoking women. *PLoS One* **8**, e73794, doi:10.1371/journal.pone.0073794 (2013).
- Sholl, L. M., Barletta, J. A., Yeap, B. Y., Chirieac, L. R. & Hornick, J. L. Sox2 protein expression is an independent poor prognostic indicator in stage I lung adenocarcinoma. *Am J Surg Pathol* **34**, 1193–8, doi:10.1097/PAS.0b013e3181e5e024 (2010).
- Tanaka, K. *et al.* Prognostic significance of aromatase and estrogen receptor beta expression in EGFR wild-type lung adenocarcinoma. *Am J Transl Res* **8**, 81–97 (2016).
- Yang, C. J. *et al.* Poorer prognosis in Taiwanese female ever smokers with stage IV lung adenocarcinoma who were readministered a tyrosine kinase inhibitor. *Onco Targets Ther* **9**, 1511–8, doi:10.2147/OTT.S100169 (2016).
- Zhang, Y. *et al.* USP44 regulates centrosome positioning to prevent aneuploidy and suppress tumorigenesis. *J Clin Invest* **122**, 4362–74, doi:10.1172/JCI63084 (2012).
- Landi, M. T. *et al.* Gene expression signature of cigarette smoking and its role in lung adenocarcinoma development and survival. *PLoS One* **3**, e1651, doi:10.1371/journal.pone.0001651 (2008).
- Husni, R. E. *et al.* DNMT3a expression pattern and its prognostic value in lung adenocarcinoma. *Lung Cancer* **97**, 59–65, doi:10.1016/j.lungcan.2016.04.018 (2016).
- Tang, Z. *et al.* Elevated expression of FABP3 and FABP4 cooperatively correlates with poor prognosis in non-small cell lung cancer (NSCLC). *Oncotarget* **7**, 46253–46262, doi:10.18632/oncotarget.10086 (2016).
- Yang, G. *et al.* High ABCG4 Expression Is Associated with Poor Prognosis in Non-Small-Cell Lung Cancer Patients Treated with Cisplatin-Based Chemotherapy. *PLoS One* **10**, e0135576, doi:10.1371/journal.pone.0135576 (2015).
- Ruiz-Martinez, M. *et al.* YKT6 expression, exosome release, and survival in non-small cell lung cancer. *Oncotarget* **7**, 51515–51524, doi:10.18632/oncotarget.9862 (2016).
- Mecklenburg, I., Sienel, W., Schmid, S., Passlick, B. & Kufer, P. A Threshold of Systemic MAGE-A Gene Expression Predicting Survival in Resected Non-Small Cell Lung Cancer. *Clin Cancer Res* **23**, 1213–1219, doi:10.1158/1078-0432.CCR-16-0557 (2017).
- Hidai, H., Bardales, R., Goodwin, R., Quertermous, T. & Quertermous, E. E. Cloning of capsulin, a basic helix-loop-helix factor expressed in progenitor cells of the pericardium and the coronary arteries. *Mech Dev* **73**, 33–43, doi:10.1016/S0925-4773(98)00031-8 (1998).
- Miller, C. L. *et al.* Disease-related growth factor and embryonic signaling pathways modulate an enhancer of TCF21 expression at the 6q23.2 coronary heart disease locus. *PLoS Genet* **9**, e1003652, doi:10.1371/journal.pgen.1003652 (2013).
- Quaggin, S. E., Vanden Heuvel, G. B. & Igarashi, P. Pod-1, a mesoderm-specific basic-helix-loop-helix protein expressed in mesenchymal and glomerular epithelial cells in the developing kidney. *Mech Dev* **71**, 37–48, doi:10.1016/S0925-4773(97)00201-3 (1998).
- Lu, J. *et al.* The basic helix-loop-helix transcription factor capsulin controls spleen organogenesis. *Proc Natl Acad Sci USA* **97**, 9525–30, doi:10.1073/pnas.97.17.9525 (2000).
- Acharya, A. *et al.* The bHLH transcription factor Tcf21 is required for lineage-specific EMT of cardiac fibroblast progenitors. *Development* **139**, 2139–49, doi:10.1242/dev.079970 (2012).
- Quaggin, S. E. *et al.* The basic-helix-loop-helix protein pod1 is critically important for kidney and lung organogenesis. *Development* **126**, 5771–83 (1999).
- Franca, M. M., Ferraz-de-Souza, B., Lerario, A. M., Frago, M. C. & Lotfi, C. F. POD-1/TCF21 Reduces SHP Expression, Affecting LRF-1 Regulation and Cell Cycle Balance in Adrenocortical and Hepatocarcinoma Tumor Cells. *Biomed Res Int* **2015**, 841784–9, doi:10.1155/2015/841784 (2015).
- Dai, Y. *et al.* Down-regulation of TCF21 by hypermethylation induces cell proliferation, migration and invasion in colorectal cancer. *Biochem Biophys Res Commun* **469**, 430–6, doi:10.1016/j.bbrc.2015.09.109 (2016).
- Arab, K. *et al.* Epigenetic deregulation of TCF21 inhibits metastasis suppressor KISS1 in metastatic melanoma. *Carcinogenesis* **32**, 1467–73, doi:10.1093/carcin/bgr138 (2011).

34. Smith, L. T. *et al.* Epigenetic regulation of the tumor suppressor gene TCF21 on 6q23-q24 in lung and head and neck cancer. *Proc Natl Acad Sci USA* **103**, 982–7, doi:10.1073/pnas.0510171102 (2006).
35. Wu, H. *et al.* Curcumin increases exosomal TCF21 thus suppressing exosome-induced lung cancer. *Oncotarget* **7**, 87081–87090, doi:10.18632/oncotarget.13499 (2016).
36. Arab, K. *et al.* Long noncoding RNA TARID directs demethylation and activation of the tumor suppressor TCF21 via GADD45A. *Mol Cell* **55**, 604–14, doi:10.1016/j.molcel.2014.06.031 (2014).
37. Gooskens, S. L. *et al.* TCF21 hypermethylation in genetically quiescent clear cell sarcoma of the kidney. *Oncotarget* **6**, 15828–41, doi:10.18632/oncotarget.4682 (2015).
38. Richards, K. L. *et al.* Methylation of the candidate biomarker TCF21 is very frequent across a spectrum of early-stage nonsmall cell lung cancers. *Cancer* **117**, 606–17, doi:10.1002/cncr.v117.3 (2011).
39. Shivapurkar, N. *et al.* Differential methylation of a short CpG-rich sequence within exon 1 of TCF21 gene: a promising cancer biomarker assay. *Cancer Epidemiol Biomarkers Prev* **17**, 995–1000, doi:10.1158/1055-9965.EPI-07-2808 (2008).
40. Weiss, D., Stockmann, C., Schrodter, K. & Rudack, C. Protein expression and promoter methylation of the candidate biomarker TCF21 in head and neck squamous cell carcinoma. *Cell Oncol (Dordr)* **36**, 213–24, doi:10.1007/s13402-013-0129-5 (2013).
41. Ye, Y. W. *et al.* Down-regulation of TCF21 is associated with poor survival in clear cell renal cell carcinoma. *Neoplasma* **59**, 599–605, doi:10.4149/neo_2012_076 (2012).
42. Yang, Z., Li, D. M., Xie, Q. & Dai, D. Q. Protein expression and promoter methylation of the candidate biomarker TCF21 in gastric cancer. *J Cancer Res Clin Oncol* **141**, 211–20, doi:10.1007/s00432-014-1809-x (2015).
43. Li, J. *et al.* Identification of high-quality cancer prognostic markers and metastasis network modules. *Nat Commun* **1**, 34, doi:10.1038/ncomms1033 (2010).
44. Wang, E. Understanding genomic alterations in cancer genomes using an integrative network approach. *Cancer Lett* **340**, 261–9, doi:10.1016/j.canlet.2012.11.050 (2013).
45. Feng, A., Tu, Z. & Yin, B. The effect of HMGB1 on the clinicopathological and prognostic features of non-small cell lung cancer. *Oncotarget* **7**, 20507–19, doi:10.18632/oncotarget.7050 (2016).
46. Jiang, Y. Z. *et al.* Transcriptome Analysis of Triple-Negative Breast Cancer Reveals an Integrated mRNA-lncRNA Signature with Predictive and Prognostic Value. *Cancer Res* **76**, 2105–14, doi:10.1158/0008-5472.CAN-15-3284 (2016).
47. Kong, F. *et al.* Hepatitis B virus X protein promotes interleukin-7 receptor expression via NF-kappaB and Notch1 pathway to facilitate proliferation and migration of hepatitis B virus-related hepatoma cells. *J Exp Clin Cancer Res* **35**, 172, doi:10.1186/s13046-016-0448-2 (2016).
48. Wang, G. *et al.* Simvastatin induces cell cycle arrest and inhibits proliferation of bladder cancer cells via PPARgamma signalling pathway. *Sci Rep* **6**, 35783, doi:10.1038/srep35783 (2016).
49. Yang, Z. *et al.* MARCKS contributes to stromal cancer-associated fibroblast activation and facilitates ovarian cancer metastasis. *Oncotarget* **7**, 37649–37663, doi:10.18632/oncotarget.8726 (2016).
50. Gyorffy, B., Surowiak, P., Budczies, J. & Lanczky, A. Online survival analysis software to assess the prognostic value of biomarkers using transcriptomic data in non-small-cell lung cancer. *PLoS One* **8**, e82241, doi:10.1371/journal.pone.0082241 (2013).
51. Mizuno, H., Kitada, K., Nakai, K. & Sarai, A. Prognoscan: a new database for meta-analysis of the prognostic value of genes. *BMC Med Genomics* **2**, 18, doi:10.1186/1755-8794-2-18 (2009).

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant No. 81572284) and the Important Research and Development Plan of Hunan Provincial Science and Technology Department (Grant No. 2015SK20662). We thank Donglin Cheng, Lang Ma and Qingqing LYU from Gene-Cloud of Biotechnology Information Center for assisting with the data analysis.

Author Contributions

Conceived and designed the study: Q.C. and J.X.; Searched databases: A.B.L. and X.X.L.; Data analyses: W.L., X.C. and J.X.; Preparation of tables and figures: S.Y.H., B.X.H. and J.X.; Wrote and revised the manuscript: J.X. and Q.C.; All authors reviewed the manuscript.

Additional Information

Supplementary information accompanies this paper at doi:10.1038/s41598-017-02290-2

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017