Mathematical Biosciences and Engineering

MBE, 18(5): 6608–6619.
DOI: 10.3934/mbe.2021327
Received: 01 April 2021
Accepted: 07 June 2021
Published: 03 August 2021

http://www.aimspress.com/journal/MBE

Research article

Analysis of the role of METTL5 as a hub gene in lung adenocarcinoma based on a weighted gene co-expression network

Xinwang Yan [1], Xiaowen Zhao [2], Qing Yan [2], Ye Wang [2,*] and Chunling Zhang[2,*]

[1] Medical College of Qingdao University, Jining NO.1 People's Hospital, Qingdao, Shandong Province 266042, China
[2] Clinical Laboratory, Qingdao Central Hospital, The Second Affiliated Hospital of Medical College of Qingdao University, Qingdao, Shandong Province 266042, China

Correspondence: E-mail: qdzcl2011@163.com; yewang@qdu.edu.cn; Tel: (86)-0532-84855397; (86)-0532-84961583.

Abstract: Lung adenocarcinoma (LUAD) is a frequently diagnosed malignant tumor that is highly invasive and lethal. The prognosis of patients with LUAD still needs to be improved, as conventional treatment is remarkably well tolerated. In this study, the expression profile of LUAD in the TCGA database was used for differential expression analysis, and differential expression genes were determined to construct a weighted gene co-expression network analysis (WGCNA) for dividing and finding the gene modules with the highest correlation with tumor stage. Here, METTL5, DDX23, GPSM2, CEP95, WDCP, and METL17 were identified as hub genes. According to the relation degree, METTL5 was determined as the candidate gene in this study. Difference analysis and receiver operating characteristic (ROC) curve were applied to identify the predictive performance of METTL5 in LUAD, and Kaplan-Meier (KM) analysis showed that the prognosis of LUAD patients with high METTL5 expression was poor. Further GSEA analysis showed that high-expressed METTL5 was related to epithelial-mesenchymal transition and other pathways. Therefore, METTL5 may be involved in the occurrence and malignant progression of LUAD. The current findings provide an effective molecular target for early diagnosis of LUAD, helping monitor the malignant progression of LUAD and improve the prognosis of LUAD patients.

Keywords: weighted gene co-expression network analysis; lung adenocarcinoma; METTL5; DDX23; cancer genome map

1. Introduction

Cancer remains the second most common cause of death in the world, and the incidence of cancer will continue to increase rapidly in the future. Cancer is expected to be a major drag on global life expectancy in the 21st century, and lung cancer is one of the leading causes of cancer morbidity and mortality worldwide, as deaths from lung cancer account for nearly one-fifth of all cancer deaths each year [1,2]. Lung cancer is mainly divided into adenocarcinoma, squamous cell carcinoma, small cell carcinoma and large cell carcinoma [3], and lung adenocarcinoma (LUAD) is the most common histological subtype, accounting for about 40% of lung cancer cases [4]. LUAD is highly aggressive and fatal. The overall survival period of LUAD patients is shorter than 5 years, and patients are prone to develop high tolerance when receiving conventional radiotherapy and chemotherapy [5]. Discovering specific tumor driver genes and pathways in LUAD and promoting personalized targeted therapy related to tumor progression may facilitate the clinical diagnosis and treatment of LUAD and accurate prediction of clinical treatment outcome [4,5].

Weighted gene co-expression network analysis (WGCNA) is a scientific technique for constructing co-expression networks based on mRNA expression profiles [6]. WGCNA connects corresponding genes that are significantly co-expressed in tissue samples using nodes and soft threshold assignment. Gene pairs sharing corresponding weights and their adjacency functions ensure that their connections conform to scale-free network analysis. WGCNA helps identify pivotal genes in the cancer process, and is widely applied to screen tumor markers for diagnosis and treatment of cancers [6,7].

The Cancer Genome Atlas (TCGA) is a large-scale public project established by the National Institutes of Health. Through genome sequencing and comprehensive multi-dimensional analysis of more than 11,000 human tumor samples from 33 different cancers, TCGA explores and classifies the major oncogenes in the human tumor population, thus improving the existing clinical diagnosis and treatment methods, and ultimately preventing or treating cancers [8,9]. In this study, a weighted gene co-expression network was constructed based on the expression profile of LUAD from TCGA and clinical data, aiming at determining pivotal genes in the occurrence and development of LUAD. This study provides potential biological markers for the diagnosis and treatment of LUAD.

2. Materials and methods

2.1. Data filtering and preprocessing

LUAD-related mRNA expression profiles and clinical data were downloaded from the TCGA database (https://www.cureline.com/the-cancer-genome-atlas.html), including 513 tumor samples and 59 normal samples. From GEO database (https://www.ncbi.nlm.nih.gov/), we searched and screened mRNA expression data (GSE138682) in LUAD and adult normal lung tissues and sorted them.

2.2. Difference analysis

The edgeR package was used to perform differential analysis on the expression profiles of mRNAs collected from TCGA. The screening conditions were $|\log2FC| > 1$, $P < 0.05$, and the R package ggplot2 was used to draw the volcano map.

2.3. Weighted gene co-expression network analysis (WGCNA)

The R package WGCNA was used to construct the weighted co-expression network, and β was to 5 to ensurea scale-free network, which was then transformed into a topological overlap matrix (TOM) by after establishing adjacency matrix. The hierarchical clustering tree was visualized according to the TOM. The minimum number of genes in the module was 30, the cutting height was 0.25, and the threshold ≥ 0.2 was the screening condition for output.

2.4. Function and pathway enrichment analysis

To determine the biological functions of trait-related modules, the R package ClusterProfiler was applied to perform functional annotation and pathway enrichment analysis on the obtained module genes from the Gene Ontology (GO) database and Kyoto Encyclopedia of Genes and Genomes (KEGG). The data were visualized by the R package ggplot2.

2.5. Screening of hub genes

According to the intra-module connectivity, the target modules were filtered by WGCNA, visualized by Cytoscape software (version 3.7.1), and classified based on degree.

2.6. Identification of hub genes

The expressions of hub genes in LUAD tumor samples and normal samples from TCGA were compared by the edgeR analysis and further verified by GSE138682 data set. The LUAD specimens in TCGA were divided into high- and low-expression groups according to the median gene expression level of the hub genes. Kaplan-Meier (KM) analysis was performed to verify 484 patients with GDCRNAtools package, and log-rank test was used. Receiver operating characteristic (ROC) curve was plotted to evaluate the accuracy of hub genes as biomarkers, and Area Under Curve (AUC) reflected the predictive performance of the genes. The chi-square test was used to analyze the correlation of clinical traits among different subgroups. $P<0.05$ was considered as statistically significant.

2.7. Single gene set enrichment analysis (GSEA)

To further clarify the biological role of the hub genes in LUAD, the GSEA function in the R package enrichplot was applied to analyze the target genes identified from TCGA. The log2FC in the mRNA expression profile was pre-sorted and the gene list was generated according to ENTREZID. The MSigDB database was employed to match the H gene set in the gene list to calculate the enrichment score (ES), according to the filter condition of FDR<25%, |NES|>1.5, $P<0.05$.

3. Results

3.1. Identification of DEGs

The R package edge was used to screen differentially expressed mRNAs (DEmRNAs) from the TCGA expression profile. A total of 9550 DEmRNAs (including 3072 up-regulated genes and 6,478 down-regulated genes) were identified, as shown in Figure 1.

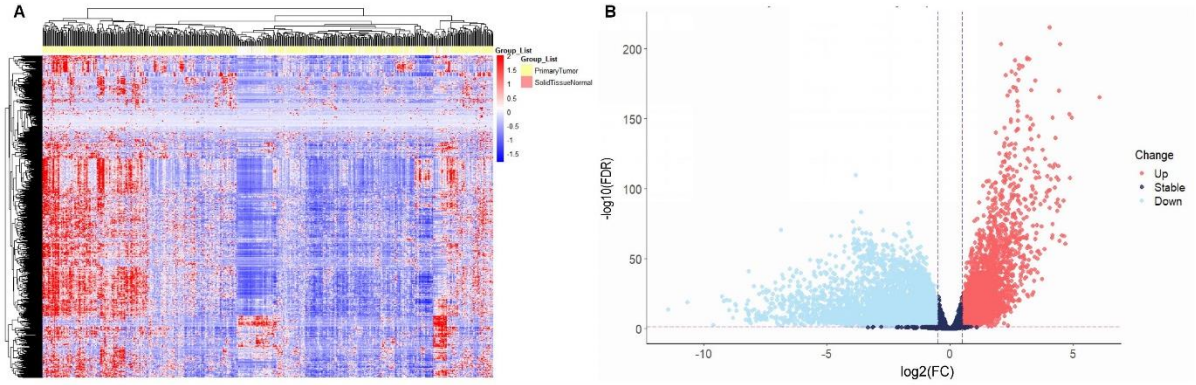

Figure 1. Identification of differentially expressed mRNAs (DEmRNAs) in the LUAD expression profile in TCGA. A) Heat map drawing based on mRNA expression profile in LUAD; B) Volcano map of DEmRNAs.

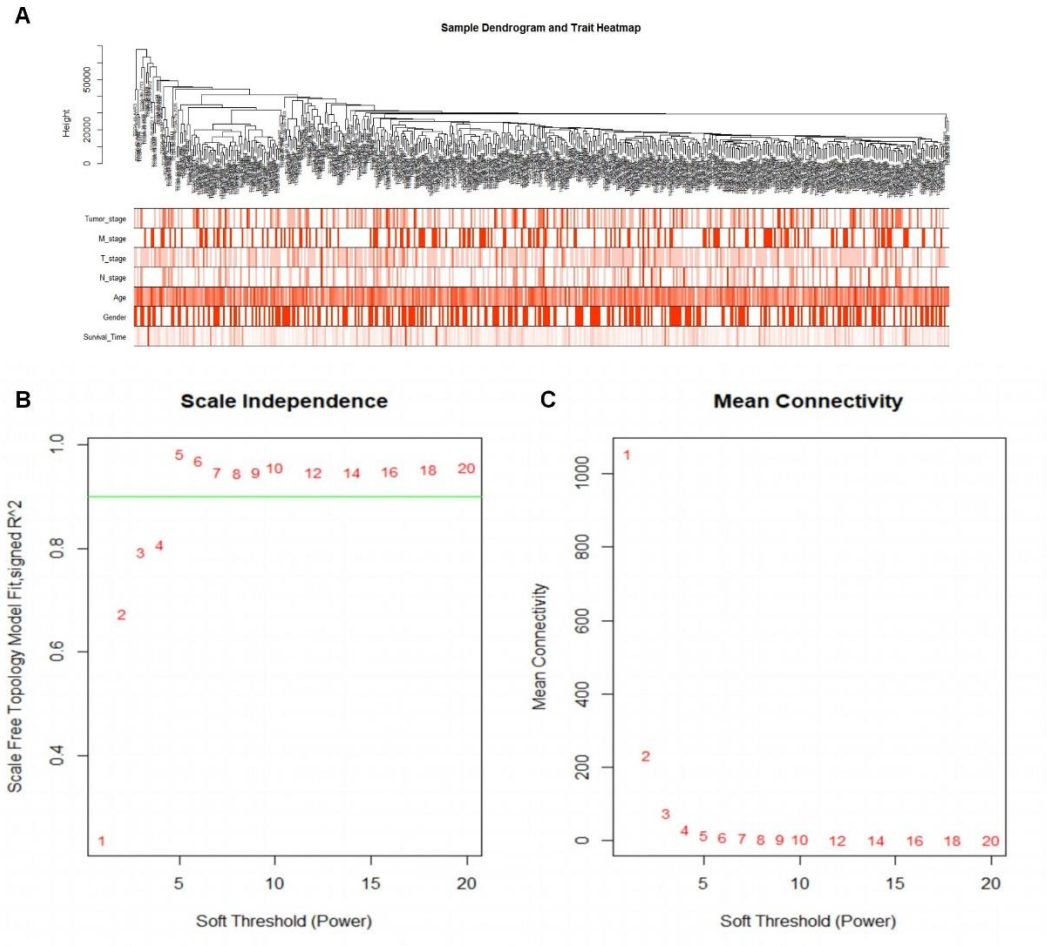

Figure 2. Construction of a co-expression network. A) Dendrograms and trait heat maps of 484 samples, clustered based on expression data in tumor samples and normal samples in LUAD, color intensity and clinical traits (late tumor staging, M staging, T staging, N staging, age, male, longer survival time) are directly proportional; B) scale-free fitting index for each β; C) average connectivity.

3.2. Construction of weighted co-expression network and identification of key modules

After preprocessing and quality evaluation of the TCGA data, an expression matrix of 484 samples was obtained (Figure 2A), and a co-expression network is constructed, with $\beta = 5$, R [2] = 0.98 to ensure the scale-free nature of the network (Figure 2B-C). The dynamic shear tree method was employed to divide and retain 32 modules. We found that the correlation between the blue module and the tumor stage was significantly closer than that of other modules. Thus, the blue module was confirmed as the key module (Figure 3).

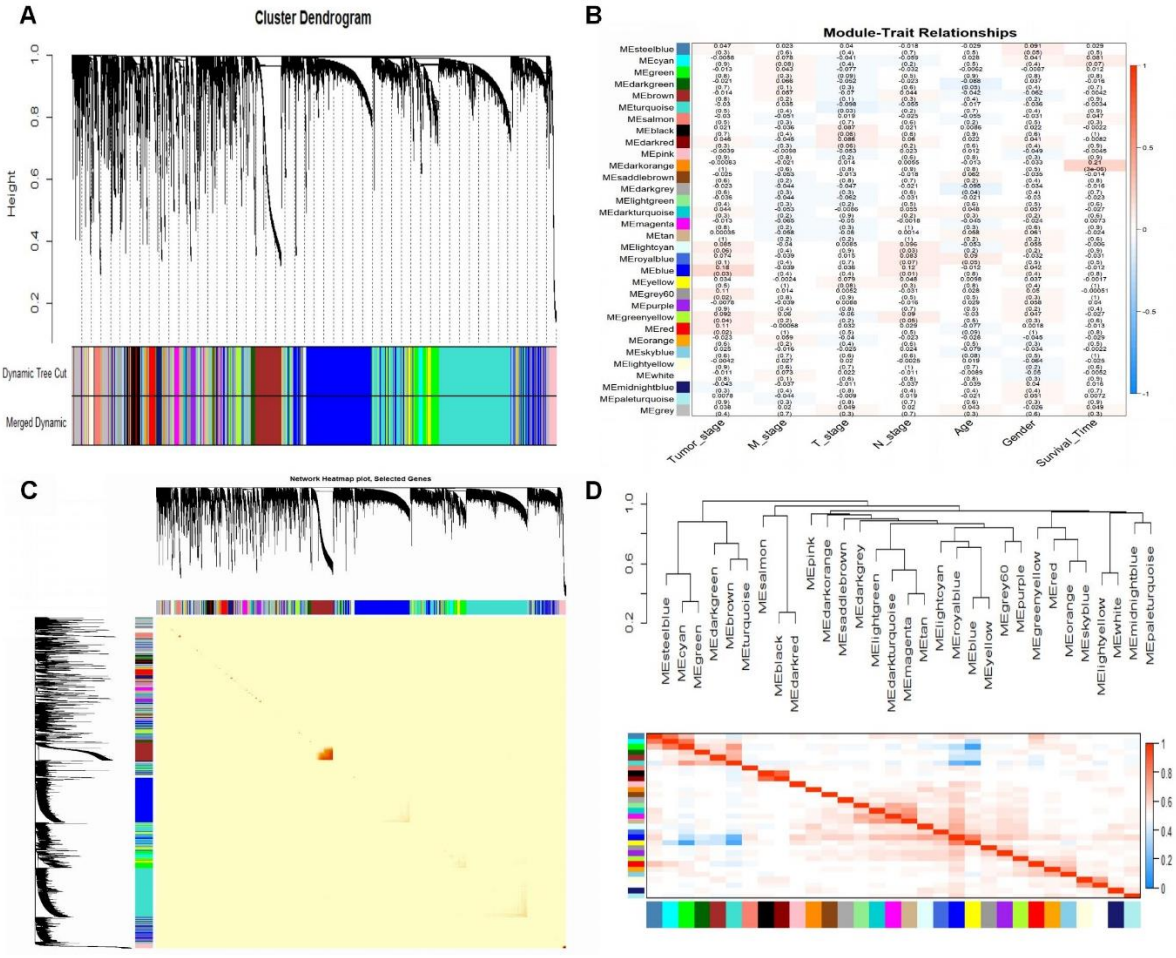

Figure 3. Screening of key modules. A) Dendrogram of the expression levels of all genes based on the difference metrics (1-TOM) clustering; B) Heat map of the correlation between the module characteristic genes and samples, each cell contains the correlation coefficient and P value; C) Total Interaction of expressed genes; D) Correlation diagram between modules.

3.3. Enrichment analysis

The correlation between gene members and gene significance in the blue module was verified (Figure 4A). GO enrichment analysis showed that the genes in the blue module were related to DNA replication, cell cycle checkpoints, and cell cycle G2/M phase transitions, moreover, biological

functions (BP) related to RNA catalytic activity, single-stranded DNA binding, helicase activity and other molecular functions (MF) and chromosomal regions and certain related cell components (CC) were found to be significantly enriched (Figure 4B). KEGG analysis results also demonstrated that the genes in the blue module were enriched in the cell cycle, DNA replication, p53 signaling pathway and some other related pathways (Figure 4C), indicating that the genes in the blue module may be involved in the development of LUAD.

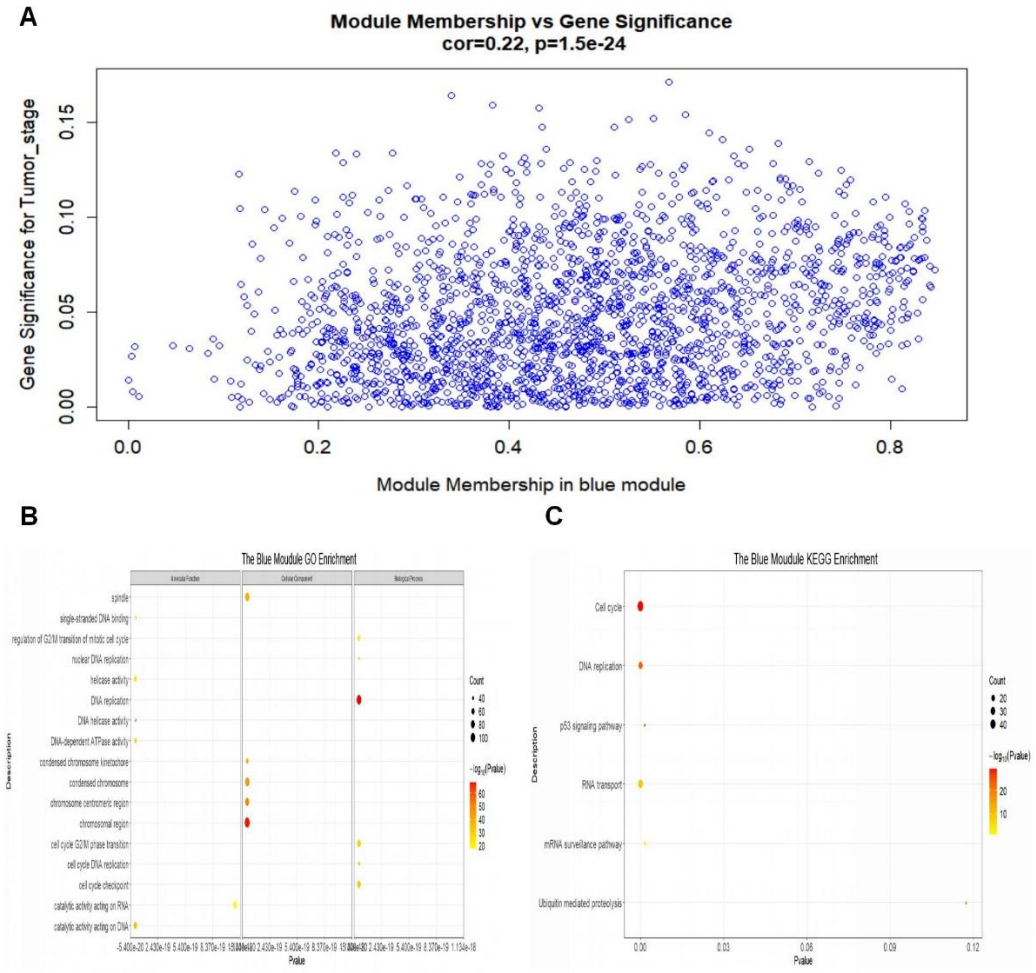

Figure 4. Enrichment analysis of the blue module. A) Scatter plot between members of the blue module and gene significance; B) GO enrichment analysis of genes in the blue module; C) KEGG enrichment analysis of genes in the blue module.

3.4. Identification of hub genes

The blue module contained a total of 2110 genes. Threshold > 0.2 was the filter condition to indicate the correlation edge in the blue module. After importing the data into Cytoscape, we obtained 218 edges and 173 nodes. According to the degree ranking, the degree of the six genes, namely, METTL5, DDX23, GPSM2, CEP95, WDCP, and METL17, was the largest (Figure 5), and therefore they were considered as the hub genes. Among them, as the degree of METTL5 was the highest, it served as the key gene of this study for subsequent analysis.

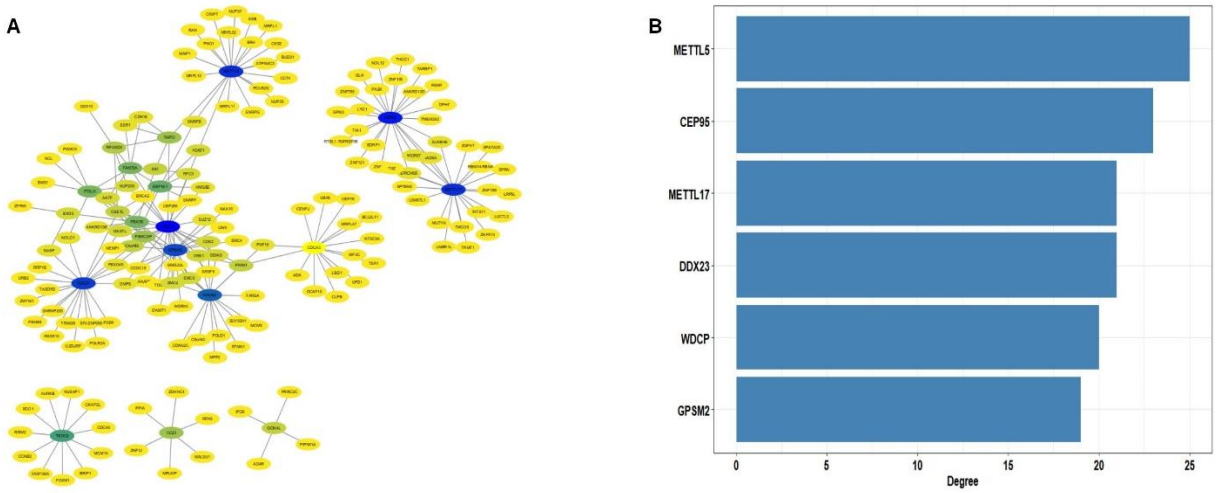

Figure 5. Mining of hub genes. A) Using Cytoscape to graphically depict the blue module, showing the gene pair with the highest intra-mold topological overlap, and each link corresponds to the TOM between the connected nodes; B) The degree value of each hub gene.

3.5. The expression and role of METTL5 in LUAD

A comparative analysis of the expression levels of METTL5 in LUAD tumor samples and normal samples from TCGA and GSE138682 showed that the relative expression of METTL5 in tumor samples was higher than that in normal samples (Figure 6A). KM analysis was performed to plot overall survival of LUAD patients with high and low METTL5 expression. We found that the overall survival of LUAD patients with high METTL5 expression was generally shorter (Figure 6B), suggesting that high expression of METTL5 may resulted in a poor prognosis of LUAD patients. The ROC curve (Figure 6C) demonstrated that the expression level of METTL5 had a high predictive performance in identifying tumor tissues and normal tissues (AUC = 0.854 (0.854-0.900), P<0.001; AUC = 0.880 (0.640-1.000), P= 0.047). However, there was no significant difference in tumor stage, M stage, T stage, N stage, age, or gender of LUAD patients between the two groups, which required further study (Figure 6D). GSEA was used to analyze the potential function of METTL5 in LUAD. It was found that the genes in the LUAD genome with high METTL5 expression were significantly enriched in the genes of "HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION", "HALLMARK_INFLAMMATORY_RESPONSE", and "HALLMARK_KRAS_SIGNALING_UPLU", indicating that METTL5 may be related to tumor metastasis and inflammation, reaction and carcinogenic signals and some other related biological processes.

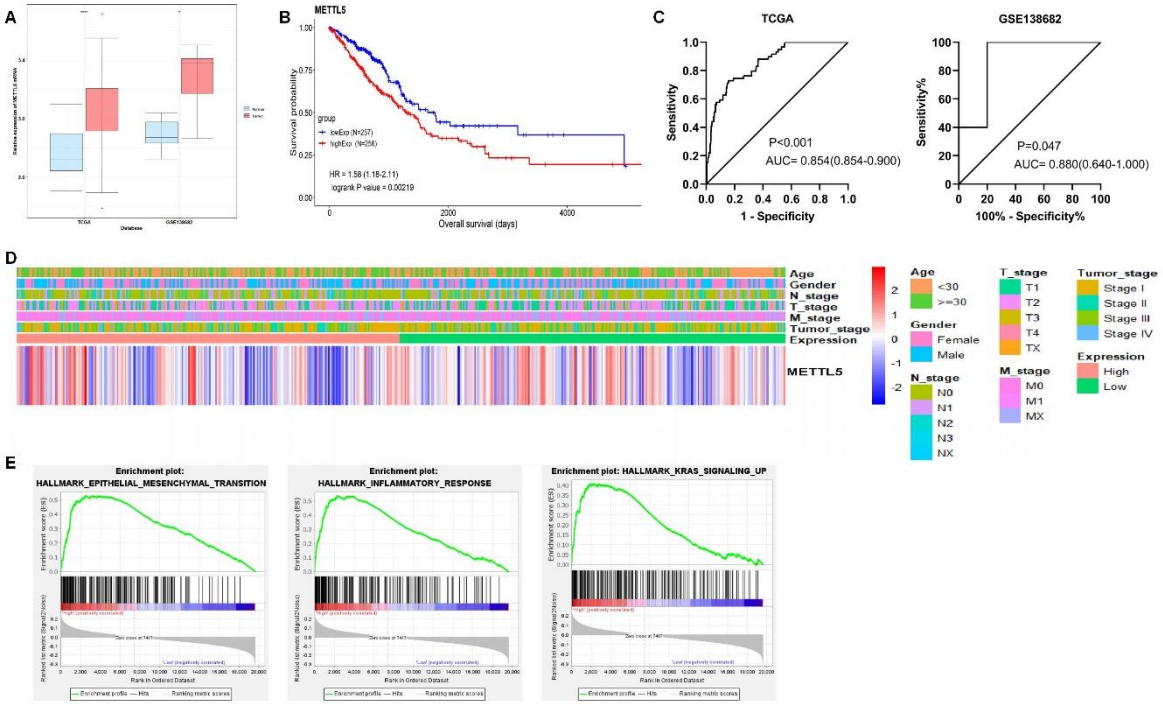

Figure 6. The expression and role of METTL5 in LUAD. A) Comparison of the expression levels of METTL5 in tumor samples and normal samples on TCGA and GSE138682 chip; B) Survival curves of LUAD patients with high and low METTL5 expression in TCGA; C) According to the expression levels of tumor samples and normal samples from TCGA and GSE138682 chips and the receiver operating characteristic (ROC) curve; D) Heat map of the correlation between high and low expression of METTL5 and the traits of LUAD patients; E) GSEA analysis result of METTL5.

4. Discussion

In this study, through differential analysis of the mRNA expression profile of LUAD in TCGA, a total of 9550 DEmRNAs were identified, of which 3072 DEmRNAs were up-regulated and 6,478 DEmRNAs were down-regulated. To further explore the role of these DEmRNAs in LUAD, we constructed a weighted gene co-expression network using WGCNA to identify the gene modules the most relevant to clinical traits, and divided 9550 obtained DEmRNAs into 32 modules. After the correlation heat map between the characteristic genes of the module and the sample was established, we observed that the blue module had the highest correlation with tumor staging. The tumor staging is a key indicator for clinically determining different treatment plans for LUAD patients [10,11]. The gene modules of LUAD provide help for finding the corresponding targets in LUAD treatment.

The results of GO and KEGG enrichment analysis demonstrated that the genes in the blue module were significantly enriched in BP related to DNA replication, cell cycle checkpoints, cell cycle G2/M phase transitions, cell cycle, DNA replication, and p53 signaling pathway, and other pathways. When human cell genes are stimulated by internal and external factors such as radiation and reactive oxygen species, DNA damage and replication defects will occur. Various checkpoints of normal cells will mediate DNA damage responses, activate DNA repair systems or induce cell

apoptosis, and prevent cells from proliferation. Subsequently, loss of cell cycle checkpoints or defects in the recognizing DNA damage in cancer cells will result in unlimited cell proliferation [12,13], but targeting genes that affect DNA replication pressure could ensure the overall integrity of the genome and cell cycle [12–14]. This also suggests that the genes in the blue module may be involved in the occurrence and development of LUAD.

The genes in the blue module were imported into Cytoscape, and the hub genes were screened according to the degree. Here, METTL5 with the highest degree was determined as the candidate hub gene in this study. By comparing the expression levels of METTL5 in LUAD tumor samples and normal samples collected from TCGA and GSE138682, we found that METTL5 was significantly high expressed in LUAD tumor tissues and showed a great predictive performance. In addition, LUAD patients with high expression of METTL5 tended to have a lower overall survival rate, suggesting that METTL5 may be a biomarker for diagnosis and prognosis of patients with LUAD. This is consistent with the findings of Sun [15] et al. METTL5 expression in LUAD tumor tissue and normal tissue differed, but patients with low expression always had a better prognosis. However, this study did not find correlation between clinical characteristics and the high and low expressions of METTL5. This may be explained by the grouping and statistical methods applied in this study. Our future study will collect more clinical samples for systematic comparative analysis.

Our GSEA results indicated that the high expression of METTL5 was related to epithelial-mesenchymal transition, inflammatory response and KRAS signal in LUAD. The above biological processes are the driving factors for the malignant transformation of LUAD. EMT induces the occurrence, metastasis, invasion and drug resistance of cancer [16,17]. Studies have pointed out that EMT and inflammatory response in tumor microenvironment and immune examination molecules are related to the increase of TDA, and blocking EMT can be used as a biomarker for LUAD immunotherapy [18]. On the other hand, KRAS mutation is the most common gene mutation in LUAD, and it is also an important cause of the death to LUAD patients. Signal activation of KRAS, which is often correlated with a strong inflammatory response in tracheal epithelium, may target and inhibit genes related to KRAS signaling. This may help improve the clinical outcome of KRAS-driven LUAD patients [19,20], and points to the possibility that METTL5 may be a molecular target with great potential in the clinical treatment of LUAD.

A comprehensive knowledge of genomic methylation pattern is of great significance to healthy cells and organs. Methylation defects can lead to immune deficiency, cancer and other diseases [21]. m6A, which is a common RNA modification in mammals, can affect RNA processing, translation, decay and other aspects. m6A methylation facilitates cancer diagnosis and treatment through various mechanisms. m6ARNA methylation is similar to DNA. Moreover, histone methylation is also a form of abnormal supervision and is involved in mRNA processing, nucleation, translation, decay, mammalian development, stress response, tumorigenesis and other biological processes [22,23]. Evidence indicated that m6A methylation leads to certain changes in tumor specificity, accelerates tumor occurrence and promotes tumor progression, moreover, downregulation of m6A can promote the disease progression of a variety of malignant tumors [22,24].

m6A is widely present in mRNA, rRNA, and ncRNA. Studies demonstrated that METTL5 protein can catalyze m6A at position A1832 in 18s rRNA. Loss of METTL5 expression could impair rRNA production, reduce the number of ribosomes, and decrease global translation efficiency, thereby affecting the differentiation of embryonic stem cells and pluripotency [25]. METTL5, which is an active member of the methyltransferase family, is mainly located at the nucleolus and is a main

methyltransferase in human 18srRNA methylation. It has been found that the expression of METTL5 and its cofactor TRMT112 is significantly up-regulated in cancer, which maintain 18s rRNA methylation and regulate the translation efficiency of stress response-related factors (such as ATF4) to overcome tumor pressure [26,27]. However, the role of METTL5 in lung adenocarcinoma is less understood and requires further investigation. Due to the limitations of experimental conditions, this study failed to carry out molecular function experiments and construct experimental animal models. Thus, the specific molecular mechanism of METTL5 in LUAD will become the main direction of our later research. We also plan to carry out multi-center randomized controlled trials to examine the role of METTL5 as a diagnostic and prognostic biomarker.

A total of 9550 DEmRNAs in the mRNA expression profile of LUAD in TCGA was divided into 32 gene co-expression modules by WGCNA, and METTL5 as a hub gene in the blue module, which was related to tumor staging, was identified. This study found for the first time that METTL5 was high-expressed in LUAD. Moreover, a high expression of METTL5 was found to be related to the poor prognosis of LUAD patients. The current findings provide a new understanding of the molecular mechanism of LUAD.

Funds

This study was supported by the National Natural Science Foundation of China (No. 81670822, 81370990 and 81800805), and Qingdao Key Research Project (No. 17-3-3-10-nsh and 19-6-1-3-nsh).

Conflict of interest

None of the authors declare any potential conflict of interest.

References

1. F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, A. Jemal, Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, *CA Cancer J Clin.*, **68** (2018), 394–424. doi: 10.3322/caac.21492.
2. J. A. Barta, C. A. Powell, J. P. Wisnivesky, Global Epidemiology of Lung Cancer, *Ann. Glob. Health*, **85** (2019), 8. doi:10.5334/aogh.2419
3. M. C. S. Wong, X. Q. Lao, K. F. Ho, W. B. Goggins, S. L. A. Tse, Incidence and mortality of lung cancer:global trends and association with socioeconomic status, *Sci. Rep.*, **7** (2017), 1–9.
4. X. Hua, W. Zhao, A. C. Pesatori, D. Consonni, N. E. Caporaso, T. Zhang, et al., Genetic and epigenetic intratumor heterogeneity impacts prognosis of lung adenocarcinoma, *Nat. Commun.*, **11** (2020), 1–11.
5. T. V. Denisenko, I. N. Budkevich, B. Zhivotovsky, Cell death-based treatment of lung adenocarcinoma, *Cell Death Dis.*, **9** (2018), 1–14.
6. R. Liu, W. Zhang, Z. Q. Liu, H. H. Zhou, Associating transcriptional modules with colon cancer survival through weighted gene co-expression network analysis, *BMC Genom.*, **18** (2017), 361.
7. B. Zhang, S. Horvath, A general framework for weighted gene co-expression network analysis, *Stat. Appl. Genet. Mol. Biol.*, **4** (2005), 17. doi: 10.2202/1544-6115.1128.
8. K. Tomczak, P. Czerwińska, M. Wiznerowicz, The Cancer Genome Atlas (TCGA): An immeasurable source of knowledge, *Contemp. Oncol.*, **19** (2015), A68.

9. J. Liu, T. Lichtenberg, K. A. Hoadley, L. M. Poisson, A. J. Lazar, A. D. Cherniack, et al., An integrated TCGA pan-cancer clinical data resource to drive high-quality survival outcome analytics, *Cell*, **173** (2018), 400–416. e11.
10. G. P. Kalemkerian, N. Narula, E. B. Kennedy, W. A. Biermann, J. Donington, N. B. Leighl, et al., Molecular testing guideline for the selection of patients with lung cancer for treatment with targeted tyrosine kinase inhibitors: American society of clinical oncology endorsement of the college of american pathologists/international association for the study of lung cancer/association for molecular pathology clinical practice guideline update, *J. Clin. Oncol.,* **36** (2018), 911.
11. D. J. Myers, J. M Wallen, Lung adenocarcinoma, *StatPearls [Internet]*, 2020.
12. J. Zhang, Q. Dai, D. Park, X. Deng, Targeting DNA replication stress for cancer therapy, *Genes,* **7** (2016), 51.
13. H. Kitao, M. Iimori, Y. Kataoka, T. Wakasa, E. Tokunaga, H. Saeki, et al., DNA replication stress and cancer chemotherapy, *Cancer Sci.*, **109** (2018), 264–271.
14. J. V. Forment, M. J. O'Connor, Targeting the replication stress response in cancer, *Pharmacol. Ther*., **188** (2018), 155–167.
15. S. Sun, K. Fei, G. Zhang, J. Wang, Y. Yang, W. Guo, et al., Construction and comprehensive analyses of a METTL5-associated prognostic signature with immune implication in lung adenocarcinomas, *Front. Genet.*, **11** (2021), 1801.
16. M. K. Jolly, C. Ward, M. S. Eapen, S. Myers, O. Hallgren, H. Levine, et al., Epithelial--mesenchymal transition, a spectrum of states: Role in lung development, homeostasis, and disease, *Dev. Dyn.*, **247** (2018), 346–358.
17. K. Li, D. Sun, Q. Gou, X. Ke, Y. Gong, Y. Zuo, et al., Long non-coding RNA linc00460 promotes epithelial-mesenchymal transition and cell migration in lung cancer cells, *Cancer Lett.,* **420** (2018), 80–90.
18. Y. Lou, L. Diao, E. R. P. Cuentas, W. L. Denning, L. Chen, Y. H. Fan, et al., Epithelial--mesenchymal transition is associated with a distinct tumor microenvironment including elevation of inflammatory signals and multiple immune checkpoints in lung adenocarcinoma, *Clin. Cancer Res.*, **22** (2018), 3630–3642.
19. H. Ji, A. M. Houghton, T. J. Mariani, S. Perera, C. B. Kim, R. Padera, et al., K-ras activation generates an inflammatory response in lung tumors, *Oncogene*, **25** (2006), 2105–2112.
20. M. Serresi, B. Siteur, D. Hulsman, C. Company, M. J. Schmitt, C. Lieftink, et al., Ezh2 inhibition in Kras-driven lung cancer amplifies inflammation and associated vulnerabilities, *J. Exp. Med.*, **215** (2018), 3115–3135.
21. J. F. Costello, C. Plass, Methylation matters, *J. Med. Genet.*, **38** (2001), 285–303.
22. T. Sun, R. Wu, L. Ming, The role of m6A RNA methylation in cancer, *Biomed. Pharmacother.*, **112** (2019),108613.
23. Y. Pan, P. Ma, Y. Liu, W. Li, Y. Shu, Multiple functions of m^6A RNA methylation in cancer, *J. Hematol. Oncol.*, **11** (2018), 48.
24. Q. Cui, H. Shi, P. Ye, L. Li, Q. Qu, G. Sun, et al., m^6A RNA methylation regulates the self-renewal and tumorigenesis of glioblastoma stem cells, *Cell Rep.*, **18** (2017), 2622–2634.
25. V. V. Ignatova, P. Stolz, S. Kaiser, T. H. Gustafsson, P. R. Lastres, A. Sanz-Moreno, et al., The rRNA m6A methyltransferase METTL5 is involved in pluripotency and developmental programs, *Genes Dev.*, **34** (2020), 715–729. doi:10.1101/gad.333369.119.
26. H. Chen, Q. Liu, D. Yu, S. K. Natchiar, C. Zhou, C. H. Hsu, et al., METTL5, an 18S

rRNA-specific m6A methyltransferase, modulates expression of stress response genes, *Bio. Rxiv.*, doi: (2020)10.1101/2020.0427.064162.

27. V. V. Ignatova, P. Stolz, S. Kaiser, T. H. Gustafsson, P. R. Lastres, A. Sanz-Moreno, et al., The rRNA m6A methyltransferase METTL5 is involved in pluripotency and developmental programs, *Genes Dev.*, **34** (2020), 715–729.

AIMS Press