doi: 10.4149/gpb_2022009

Biomarkers associated with metastasis and prognosis of lung adenocarcinoma based on microarray data

Jiang Du[1], Han Qiao[2] and Da-Long Xie[3]

[1] *Department of Thoracic Surgery, Chinese Medical University Affiliated No. 1 Hospital, Shenyang, Liaoning, China*
[2] *Clinical Medicine, Chinese Medical University, Shenyang, Liaoning, China*
[3] *Department of Anatomy, College of Basic Medicine, China Medical University, Shenyang, Liaoning, China*

Abstract. We attempted to discover the biomarker associated with metastasis and prognosis of lung adenocarcinoma. The mRNA/lncRNA expression profiles (GSE101836) were downloaded from the publicly available database, which included three highly metastatic and three weakly metastatic samples. The differentially expressed genes and lncRNAs were analyzed and survival analysis were performed based on the TCGA database. The prognosis-associated PPI network and mRNA-lncRNA coexpression network were constructed followed by the function and pathway enrichment analysis. The expression levels of key genes were validated in other datasets. Difference in gender was analyzed. Total 256 differentially expressed genes and 2 lncRNAs were found to be closely related with prognosis. PPI network was constructed with 222 nodes and 1464 edges. Two modules were divided from PPI network. Genes in module A were significantly enriched in cell cycle checkpoint, chromosome segregation, and mitotic cell cycle checkpoint. The module B was closely related with pyridine nucleotide metabolic process, nicotinamide nucleotide metabolic process and carbon metabolism. Coexpression network revealed lncRNA H19 and lncRNA SNHG12 were significant nodes. SNHG12 was closely related with GO:0006260~DNA replication, GO:0055114~oxidation-reduction process and hsa00010: Glycolysis/Gluconeogenesis. H19 was enriched in GO:0006555~methionine metabolic process, and GO:0046655~folic acid metabolic process. The expression levels of *TTK* and *CCNB1* were confirmed in other datasets. The expression of *TTK* and *CCNB1* was significantly higher in the male group than in the female group. *TTK*, *CCNB1* and lncRNA SNHG12 may be the biomarker associated with metastasis and prognosis of lung adenocarcinoma.

Key words: Gene expression profile — lncRNA expression profile — Lung adenocarcinoma — Survival analysis — Prognosis — Biomarker

Abbreviations: AUC, area under the curve; BP, biological process; CCNB1, cyclin B1; GEO, Gene expression omnibus; GO, Gene ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; lncRNA, long noncoding RNA; NSCLC, non-small-cell lung carcinoma; PPI, protein-protein interaction; PRDX6, peroxiredoxin 6; ROC, receiver operating characteristic; TCGA, The Cancer Genome Atlas; TTK, threonine and tyrosine kinase.

Electronic supplementary material. The online version of this article (doi: 10.4149/gpb_2022009) contains Supplementary material.

Correspondence to: Da-Long Xie, Department of Anatomy, College of Basic Medicine, China Medical University, No. 77 Puhe Road, Shenyang North New Area, Shenyang, Liaoning, 110122, China
E-mail: xike00794@163.com

Introduction

Lung cancer is originated from the uncontrolled cell proliferation in lung tissues which can spread to other tissues by the process of metastasis. This cancer can be classified into two types such as small-cell lung carcinoma and non-small-cell lung carcinoma (NSCLC) (Maxwell et al. 2019). NSCLC accounts for approximately 85–90% of lung cancer cases, which includes two major histological subtypes: squamous cell carcinoma and lung adenocarcinoma (Chang et al. 2015). Smoking is one of the risk factors for lung cancer. Although the incidence of lung cancer is declining in developed countries, with the increasing rate of smoking in developing countries, the incidence of lung cancer is expected to increase especially for China and India (Behera and Balamugesh 2004; Zhang et al. 2011). Lung adenocarcinoma, as the most common histological subtype of NSCLC is closely related with high mortality and metastasis rate (Travis 2014). For the poor prognosis, the treatment for lung adenocarcinoma has been widely investigated.

Gene alteration is implicated in the progression of cancers and studies about that drive the discovery of biomarkers for the diagnosis and treatment of lung cancer. Recently, Lathwal et al. (2020) performed an integrative analysis by screening for survival-associated genes of NSCLC from TCGA (The Cancer Genome Atlas) dataset. They constructed two prognostic gene signatures for squamous cell carcinoma and lung adenocarcinoma, respectively. Additionally, based on the GEO (Gene expression omnibus) datasets, Ni et al. (2018) found that up-regulated TOP2, CCNB1, CCNA2, UBE2C and KIF20A in NSCLC tissues were associated with poor prognosis of NSCLC. In addition to mRNA, long noncoding RNAs (lncRNAs) are also found to be dysregulated in lung cancers and play a regulatory role in metastasis process of tumor cells (Jen et al. 2017; Xiang et al. 2017). It is reported that lncRNA MALAT1 (metastasis associated lung adenocarcinoma transcript 1) is overexpresseed in lung cancer cells and is the biomarker for predicting metastasis and prognosis of NSCLC (Ji et al. 2003). lncRNA ANRIL is reported to be overexpressed in lung cancer tissues and served as the marker for predicting prognosis of NSCLC patients (Lin et al. 2015). Recent evidence suggests that lncRNA CAR10 (chromatin-associated RNA 10) is up-regulated in lung tumor tissues and promotes the metastasis of lung adenocarcinoma cells (Ge et al. 2019). Despite the advances in exploring the pathogenesis of lung cancer, biomarkers predicting the prognosis of lung adenocarcinoma have not been fully understood.

Therefore, in this paper, we downloaded the lncRNA/mRNA microarray data GSE101836 (Yu et al. 2017) associated with lung adenocarcinoma cell metastasis and re-analyzed the differentially expressed mRNAs and lncRNAs between highly metastatic and weekly metastatic lung adenocarcinoma cells. Different from that study, we also downloaded the mRNA/lncRNA expression profiles of lung adenocarcinoma and predicted the prognosis-associated genes and lncRNAs through survival analysis. The prognosis-associated gene-lncRNA network was constructed. We aimed to explore the biomarkers associated with the metastasis and prognosis of lung adenocarcinoma.

Materials and Methods

Microarray data acquiring

The lncRNA/mRNA expression profiles (GSE101836) produced from metastatic NSCLC cells were retrieved from the public GEO repository (Ashburner et al. 2000) (https://www.ncbi.nlm.nih.gov/gds/?term=) at NCBI (National Center for Biotechnology Information) database. There were total 6 samples, including 3 SPC-A-1sci (highly metastatic) and 3 SPC-A-1 (weakly metastatic) cells. The platform was GPL17843 Agilent-042818 Human lncRNA Micorarray 8_24_v2.

Data preprocessing and differential expression analysis

The raw data were downloaded and preprocessed by Limma package (Smyth 2005), including background correction, normalization, and concentration prediction. The expression values of mRNA and lncRNA were calculated based on the annotation information of probes. Differential expression analysis between SPC-A-1sci and SPC-A-1 samples was performed with the application of limma package. Classic Bayes test and Benjamini/Hochberg method were used for correction. lncRNA or mRNA with adjust p value < 0.05 and $|\log_2 \text{FC (fold change)}| > 2$ were considered as differentially expressed (dif) gene. Heatmap of dif-mRNAs and lncRNAs were visualized by pheatmap (Kolde and Kolde 2015) (version: 1.0.10, https://cran.r-project.org/web/packages/pheatmap/index.html) in R package.

Survival analysis

The data of lung adenocarcinoma samples with prognostic information were downloaded from TCGA database. The expression values and clinical information of dif-lncRNA and dif-mRNA (gene) were captured. The samples were classified into high expression and low expression groups based on the median expression value of a given lncRNA or gene. The Kaplan-Meier survival curves were analyzed by Survival analysis (Therneau and Lumley 2011) (version: 2.42-6, https://cran.r-project.org/web/packages/survival/index.html) in R package with $p < 0.05$.

Functional analysis

The dif-mRNAs and dif-lncRNAs that significantly associated with prognosis were subjected to enrichment analysis of GO (Gene ontology) biologial process (BP) terms and KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway (Kanehisa and Goto 2000) with the application of clusterprofiler package (Yu et al. 2012) (version 2.4.3, http://bioconductor.org/packages/3.2/bioc/html/clusterProfiler.html) in R package. The GO terms or pathways with $p < 0.05$ and count ≥ 2 were considered to be significant.

Prognosis associated PPI (protein-protein interaction) network analysis

The interactions between the dif-mRNA encoding proteins were predicted by STRING analysis (Szklarczyk et al. 2010) (version 10.0, http://www.string-db.org/) with PPI score (median confidence) ≥ 0.4. Cytoscape software (Shannon et al. 2003) (version 3.2.0, http://www.cytoscape.org/) was used for PPI network construction. The significantly clustered modules were divided by MCODE plugin to Cytoscape (Bandettini et al. 2012) (version 1.4.2, http://apps.cytoscape.org/apps/MCODE). The modules with score ≥ 5 were captured. Then, the function and pathway related with module genes were analyzed with $p < 0.05$ and count ≥ 2.

Coexpression analysis of prognosis associated lncRNA and mRNA analysis

The Pearson correlation coefficient between prognosis-associated lncRNA and mRNA was calculated. lncRNA-mRNA coexpression pairs with $r > 0.99$ and $p < 0.05$ were collected. The lncRNA-mRNA coexpression network was constructed. The lncRNA associated function and pathways were further analyzed.

Data validation

UALCAN (http://ualcan.path.uab.edu/) is a comprehensive, and interactive web resource for analyzing cancer OMICS data (Chandrashekar et al. 2017). In this study, the expressions of several key genes in different grades (N0: no regional lymph node metastasis; N1: metastases in 1 to 3 axillary lymph nodes; N2: metastases in 4 to 9 axillary lymph nodes; and N3: metastases in 10 or more axillary lymph nodes) were analyzed in UALCAN tool based on the lung adenocarcinoma samples from TCGA database.

Additionally, GSE126548 and GSE17475 were included to validate the expression levels of key genes between metastatic and non-metastatic samples. In GSE126548, there were 6 NSCLC patients (3 with bone metastases (BM+) and 3 without metastasis (BM−)). GSE17475 included 28 patients with lung adenocarcinoma, of whom 12 had no metastasis and 16 had lymph node metastasis.

Furthermore, we combined the OS prognostic information of TCGA lung adenocarcinoma patients and performed univariate Cox regression analysis for age, gender, pathologic M, pathologic N, pathologic T, tumor stage, and the key genes. Then the prognostic correlation factors were further performed multivariate Cox regression.

Difference in gender

There has been a growing interest in the possible differences in the presentation of lung cancer in women compared with men. Thus, in this study, we also analyzed the survival differences and the expression differences of key genes between male and female based on the TCGA lung adenocarcinoma datasets. Survival Kaplan-Meier curve in R3.6.1 was used to evaluate the prognostic correlation between male and female.

Results

RNAs with differential expression

Based on the threshold value, 1457 dif-mRNAs (64 up-regulated and 1393 down-regulated ones) and 119 dif-lncRNAs (19 up-regulated and 100 down-regulated lncRNAs) were obtained. The heatmap illustrated that the SPC-A-1sci and SPC-A-1 samples were clearly distinguished based on the expression profile of dif-mRNAs (Fig. 1A) and dif-lncRNAs (Fig. 1B).

Survival analysis

In order to explore the mRNAs or lncRNAs closely associated with prognosis, the Kaplan-Meier survival curve analysis was performed. The expression signature of 256 mRNAs were associated with prognosis including 11 up-regulated mRNAs and 245 down-regulated mRNAs. Two down-regulated lncRNAs (H19 and SNHG12) were identified to be associated with prognosis.

Significant GO function and pathways for prognosis-associated genes

The up-regulated mRNAs were significantly enriched in 5 GO-BP terms, such as regulation of apoptotic signaling pathway (GO:2001233), extrinsic apoptotic signaling pathway (GO:0097191) and positive regulation of apoptotic signaling pathway (GO:2001235). No pathways were significantly overrepresented by up-regulated mRNAs.

The down-regulated mRNAs were closely related with 34 KEGG pathways and 558 GO-BP terms, such as glyoxylate

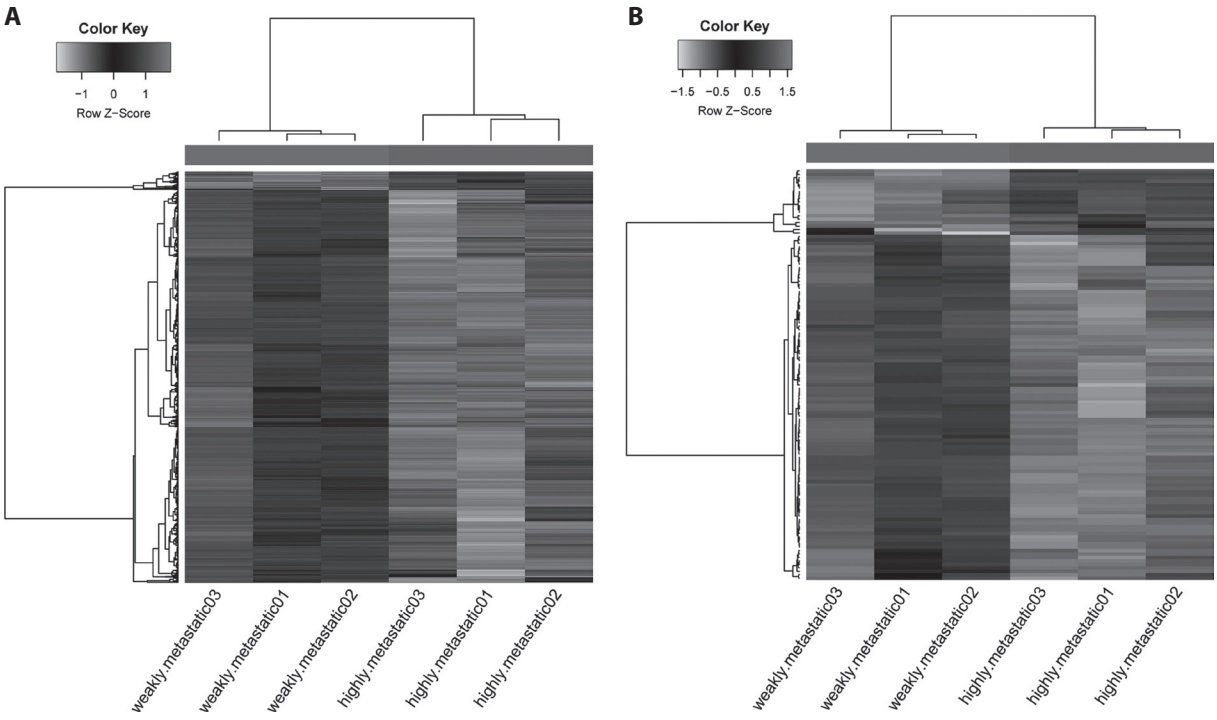

Figure 1. The heatmap for differentially expressed genes and differentially expressed lncRNAs. The samples between groups are distinguished clearly based on the expression profile of differentially expressed genes (**A**) and differentially expressed lncRNAs (**B**). For color figure see online version of the manuscript.

and dicarboxylate metabolism (hsa00630), pentose phosphate pathway (hsa00030), organelle fission (GO:0048285) and nuclear division (GO:0000280). The top 10 GO terms and pathways were displayed in Figure 2.

PPI network and modules

The PPI network consisted of 1464 interactions connecting 222 nodes (Fig. 3A). With the application of MCODE plugin, two modules with score ≥ 5 were captured as module A (score = 30.848) and module B (score = 6.5). Module A was comprised of 34 nodes and 509 edges (Fig. 3B) and module B contained 13 nodes and 39 interaction pairs (Fig. 3C). CCNB1 (cyclin B1) had the highest degree among the upregulated genes in module A and TTK (threonine and tyrosine kinase) had the highest degree among the downregulated genes in module A. PRDX6 (peroxiredoxin 6) had the highest degree in module B. The module A genes were primarily enriched in cell cycle, DNA replication related pathways and chromosome segregation, cell cycle checkpoint related GO-BP terms (Fig. 4A and B). The genes in module B were closely related with the pathways of glycolysis/gluconeogenesis, central carbon metabolism in cancer and GO functions of pyridine nucleotide metabolic process and nicotinamide nucleotide metabolic process (Fig. 4A and B).

lncRNA-mRNA coexpressed network

All the prognosis-associated genes and lncRNAs were subjected to coexpression analysis. Total 129 lncRNA-mRNA interaction pairs were obtained and the coexpressed network was constructed with 101 nodes containing 2 lncRNAs and 99 genes (Fig. 5). lncRNA H19 and lncRNA SNHG12 were the significant nodes in coexpression network. GO function and pathway analysis showed that target genes of SNHG12 was significantly enriched in 5 KEGG pathways and 7 GO-BP terms such as GO:0006260~DNA replication, GO:0000082~G1/S transition of mitotic cell cycle and hsa00010:glycolysis/Gluconeogenesis pathway. H19 was closely related with GO:0006555~methionine metabolic process, GO:0046655~folic acid metabolic process, hsa01130: Biosynthesis of antibiotics and hsa04115: p53 signaling pathway (Fig. 6).

Expression levels of key genes in different grades

The expression levels of *TTK*, *CCNB1* and lncRNA SNHG12 in lung adenocarcinoma based on nodal metastasis status are shown in Figure 7A. The results showed that *TTK* had the highest expression in N3. For SNHG12, higher expression was found in N0 compared with N1 and N2. For *CCNB1*,

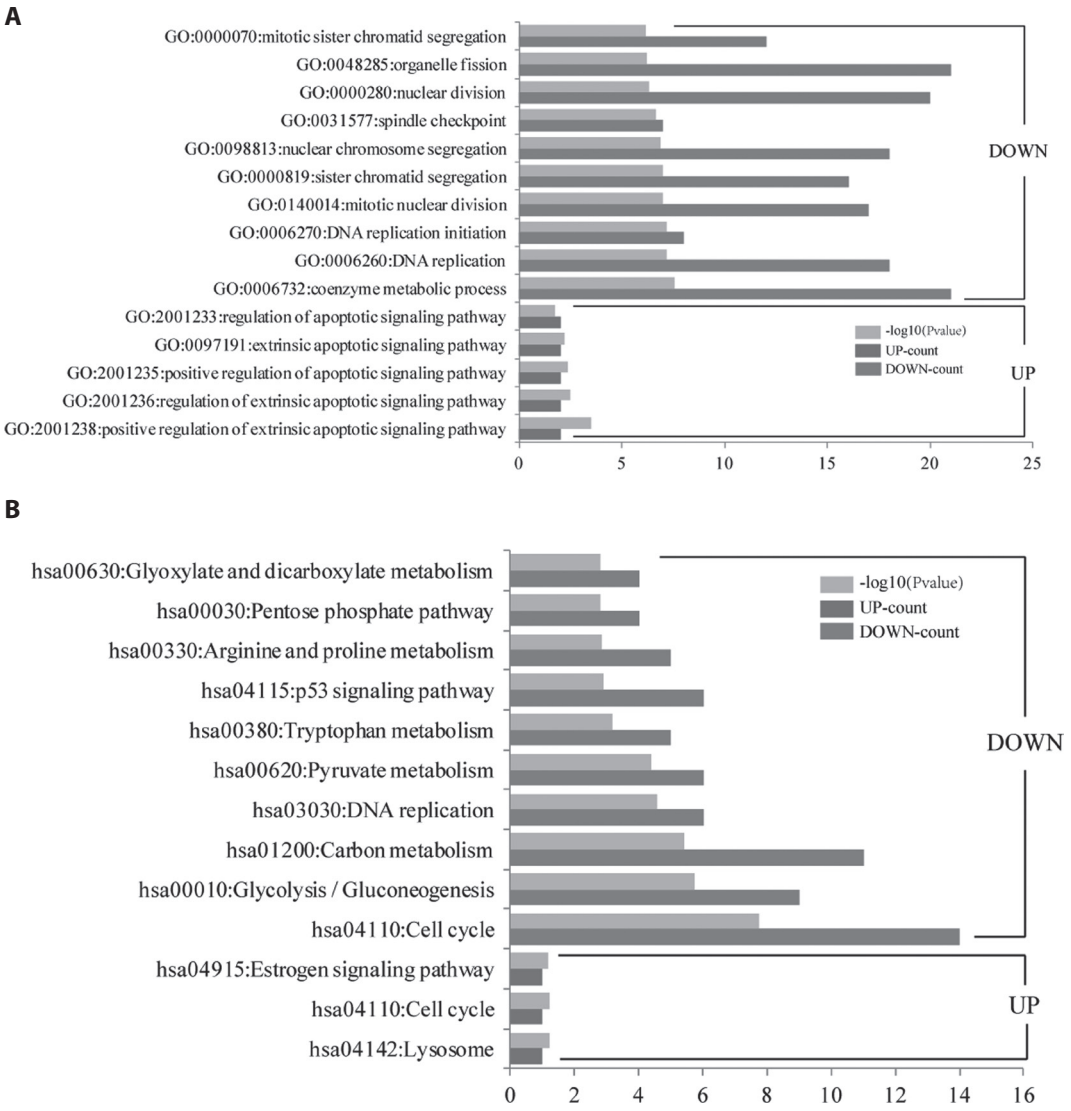

Figure 2. Top 10 GO function (**A**) and pathways (**B**) for prognosis associated genes.

higher expression was found in N2 compared with N3. Thus, the results were in consistent with the analysis above.

Expression levels of key genes between metastatic and non-metastatic samples

The expression differences of *TTK*, *CCNB1* and lncRNA SNHG12 in the metastatic *vs*. non-metastatic groups were detected, and a box diagram was drawn (Fig. 7B). It can be seen that in both datasets, the expression of *TTK* in the bone metastasis or lymph node metastasis group showed an upward trend compared with the non-metastasis group, while the expression of CCNB1 in the bone metastasis group showed a downward trend compared with the non-metastasis group, consistent with the results of the original analysis.

Independent prognostic factors screening

Univariate Cox regression analysis revealed that all of pathologic M, pathologic N, pathologic T, tumor stage, *TTK*, *CCNB1* and lncRNA SNHG12 were prognostic factors. We further included these factors into the multivariate Cox regression analysis. As shown in Figure 8, pathologic M, pathologic N and *CCNB1* were independent prognostic factors.

TTK, CCNB1 and SNHG12 were taken as explanatory variables, and OS status was taken as response variable to construct the support vector machine (SVM) model. Then the ROC (receiver operating characteristic) curve was plotted, as shown in Figure 8. The AUC (area under the curve) reached 0.647, showing a good prognostic prediction ability.

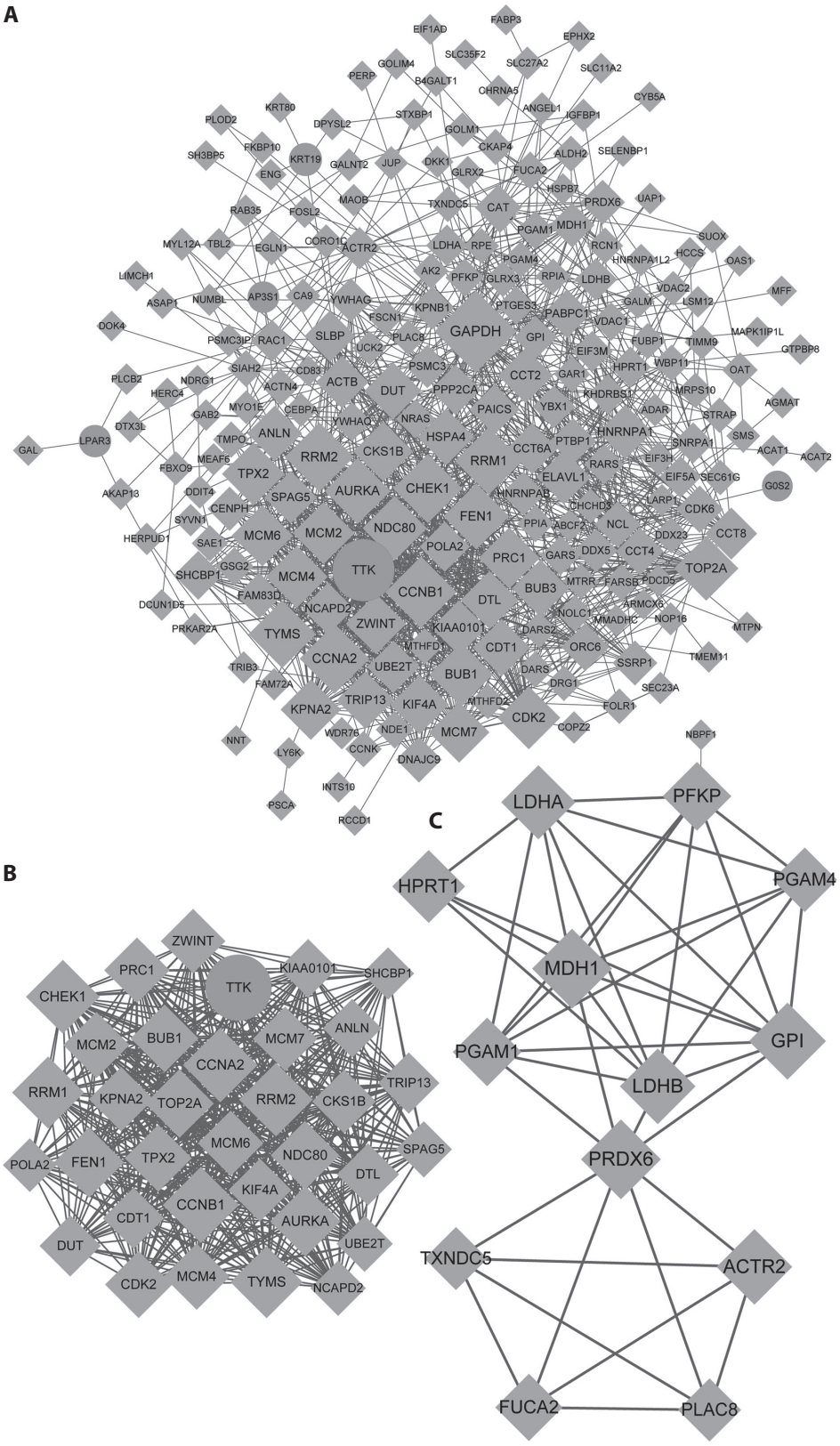

Figure 3. Protein-protein interaction network (**A**), module A (**B**) and module B (**C**). Circle, up-regulated genes; rhombus, down-regulated genes; node size represents the node degree.

Difference in gender

There was no significant difference in survival prognostic information between males and females (Fig. S1A in Supplementary material) in lung adenocarcinoma. The correlation between *TTK*, *CCNB1* and SNHG12 and gender was further detected. As shown in Figure S1B, the expression of *TTK* and *CCNB1* was significantly higher in the male group than in the female group. There was no significant difference between males and females for SNHG12.

Discussion

Lung adenocarcinoma is the most common type of lung cancer with 80% incidence of all lung cancers (Lu et al.

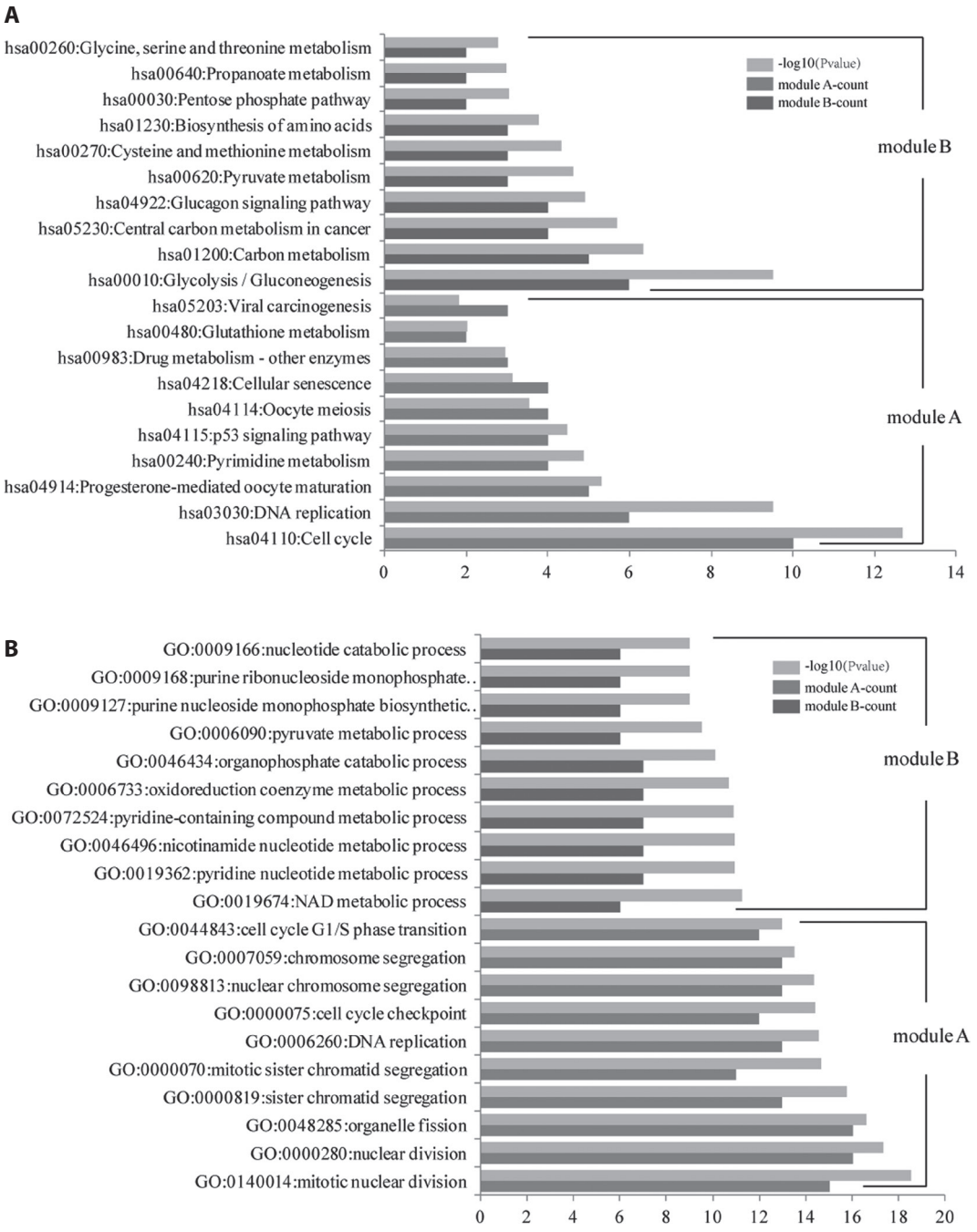

Figure 4. The significant pathways (**A**) and GO terms (**B**) for modules.

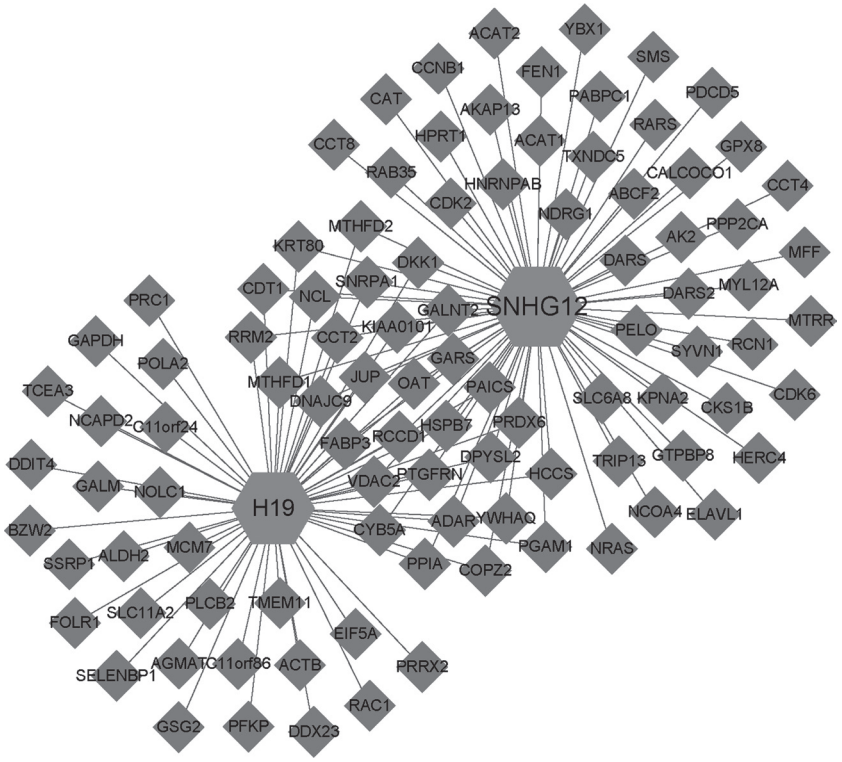

Figure 5. LncRNA-mRNA interaction network. Rhombus represents gene and hexagon represents lncRNA.

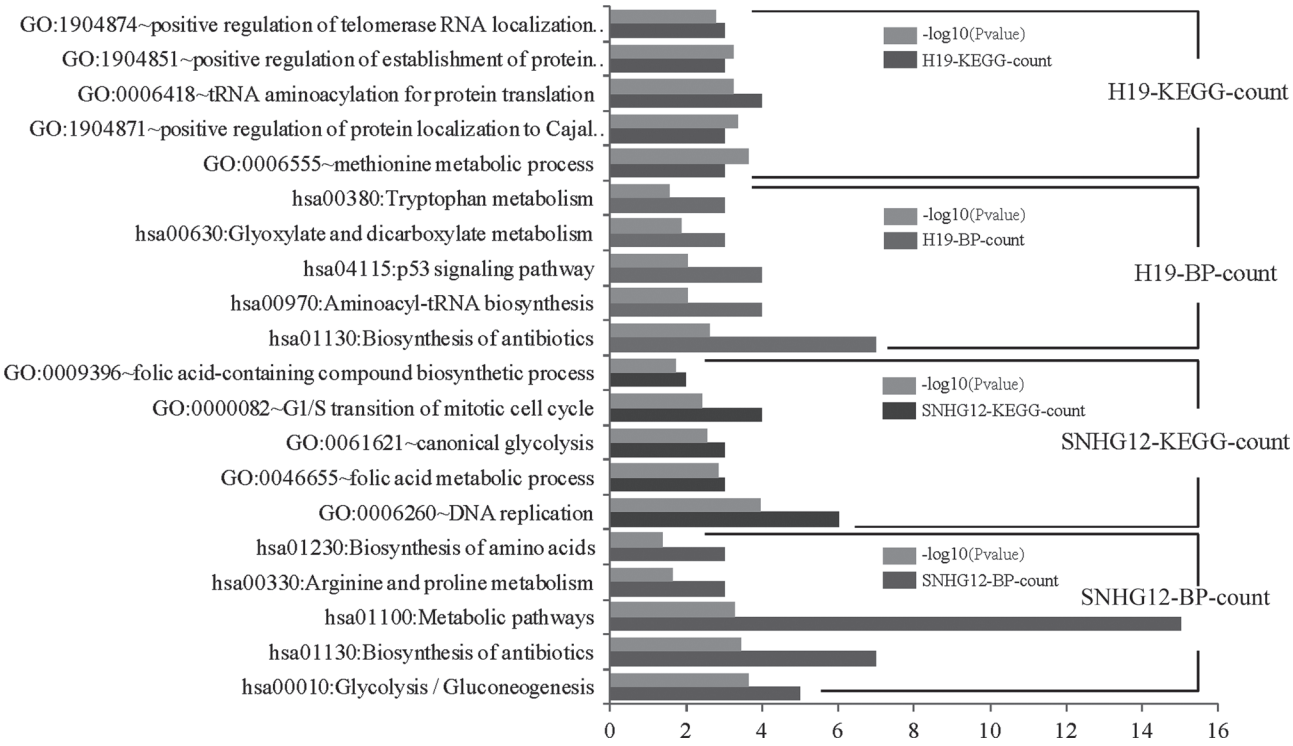

Figure 6. The significant KEGG pathways and GO functions for H19 and SNHG12.

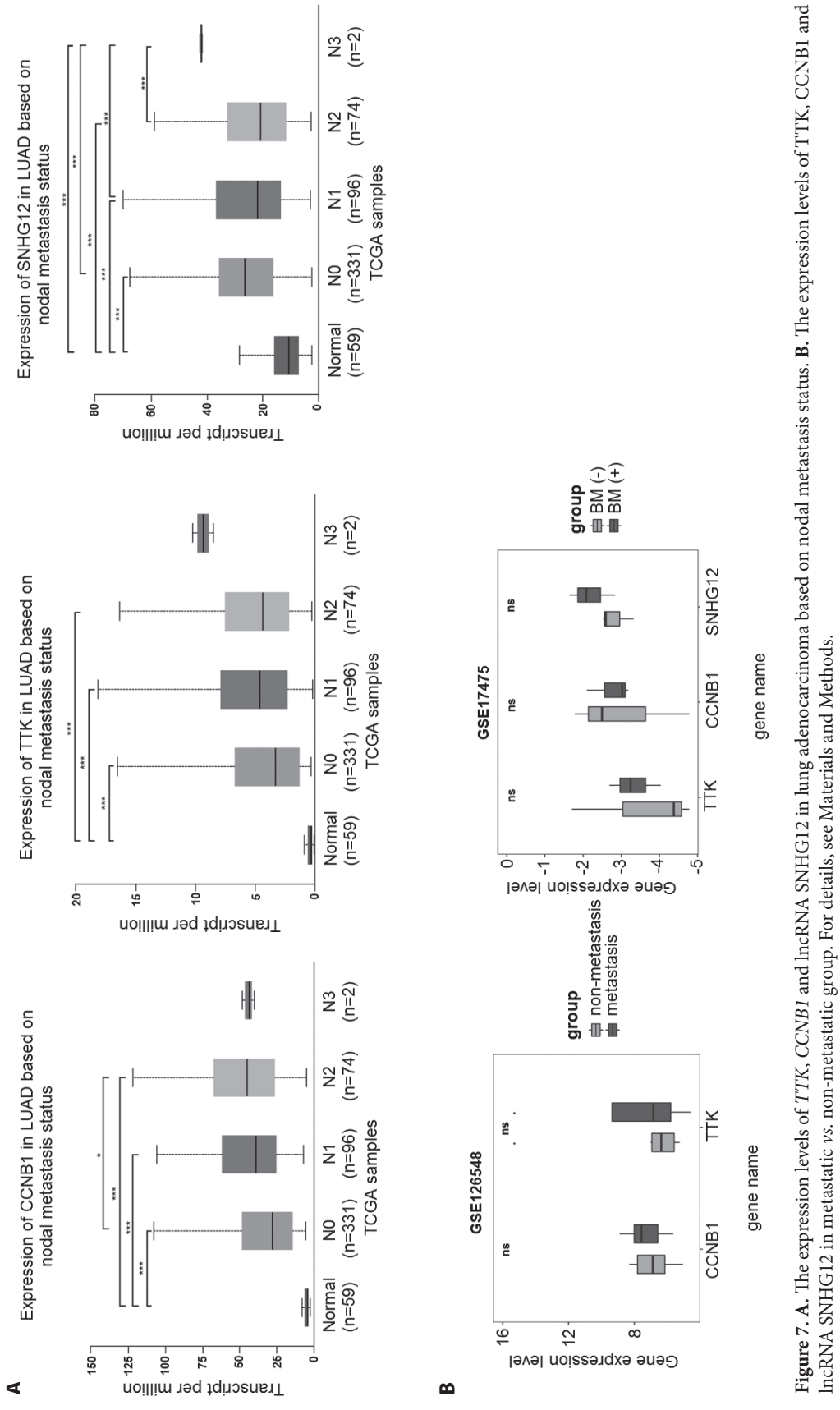

Figure 7. A. The expression levels of *TTK*, *CCNB1* and lncRNA SNHG12 in lung adenocarcinoma based on nodal metastasis status. **B.** The expression levels of TTK, CCNB1 and lncRNA SNHG12 in metastatic *vs.* non-metastatic group. For details, see Materials and Methods.

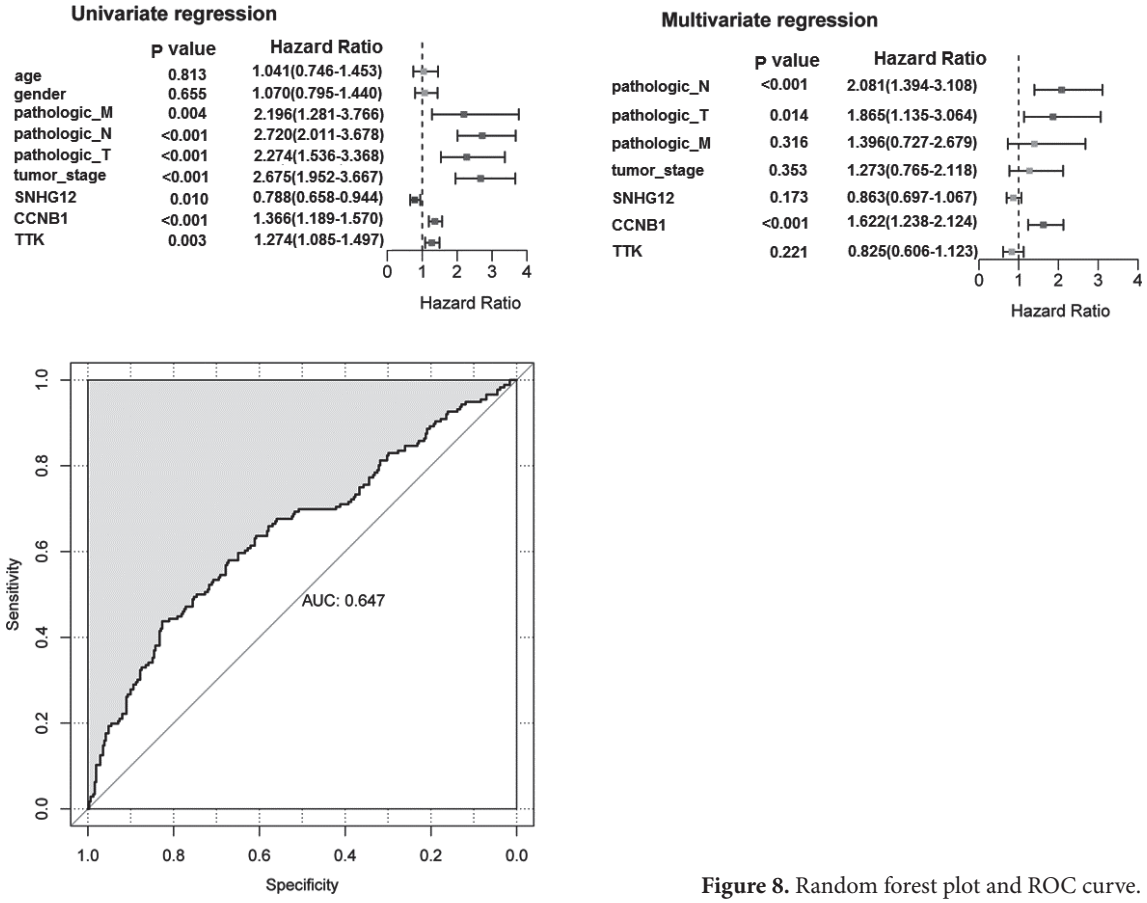

Figure 8. Random forest plot and ROC curve.

2010), which is closely related with metastasis and poor prognosis. Histopathology is necessary for cancer diagnosis, while it is inadequacy for predicting disease progression and prognosis of lung adenocarcinoma (Beer et al. 2002). The gene expression profile analysis based on microarray data facilitates the discovery of biomarkers for patient survival prediction in lung adenocarcinoma. Here, in this paper, we utilized the lncRNA/mRNA expression profiles of NSCLC cells to identify the biomarkers for predicting metastasis and prognosis of lung adenocarcinoma patients.

PPI network was constructed for genes with differential expression. Our data showed that CCNB1 and TTK were the significant nodes in PPI network and both of the genes were clustered in module A. CCNB1 encoding cyclin B1 protein is a member of conserved cyclin family that plays a regulatory role in cell cycle (Milatovich and Francke 1992). The function and pathway analysis showed that CCNB1 was involved in cell cycle related biological processes, such as cell cycle checkpoint and chromosome segregation, which was according to the previous report (Li et al. 2013). The genetic polymorphisms of CCNB1 was found to be related with the susceptibility, progression, and survival of breast cancer in Han Chinese (Li et al. 2013). CCNB1 has been proposed to be the marker for predicting prognosis of patients with ER+ breast cancer (Ding et al. 2014). In addition, the down-regulation of CCNB1 is closely related impaired cell proliferation and tumor growth of colorectal cancer (Fang et al. 2014). Besides, the previous gene expression profiling analysis showed that cell cycle genes such as CCNB1 were altered at the early stage of lung adenocarcinoma (Singhal et al. 2003). Previous evidences have showed that CCNB1 expression is associated with tumor progression and prognosis of lung adenocarcinoma (Park et al. 2018; Liu et al. 2019). Our survival analysis also showed CCNB1 was a prognosis associated gene and was expected to be a biomarker for predicting prognosis of lung adenocarcinoma.

TTK, also known as Mps1 (monopolar spindle1), is the core component of spindle assembly checkpoint, which plays a key role in chromosomes allocation. TTK has been found to be overexpressed in several types of human cancers, such as glioma, breast cancer, and colon cancer (Xie et al. 2017). It is reported that TTK protein was significantly up-regulated in liver tumor tissues compared with adjacent normal hepatic tissues (Miao et al. 2016). TTK expression inhibition by TTK siRNA significantly suppressed the liver tumor growth and

the spread of tumor cells. Targeting TTK has been proposed to be an adjunct therapy for liver cancer. In addition, TTK has been found to be specially overexpressed in triple-negative breast cancer, a aggressive subtype of breast cancer (Maire et al. 2013). The expression of TTK has been determined to evaluate the prognosis of colon and breast cancer (Reinhard et al. 2009; Xu et al. 2016). Although the role of TTK in cancers has been widely reported, the studies about TTK in lung adenocarcinoma are limited. In this paper, TTK was found to be a prognosis-associated gene and prominently enriched in cell cycle checkpoint, chromosome segregation. We suggested that TTK was involved in cell proliferation and may be the biomarker for predicting the prognosis of lung adenocarcinoma.

In addition, lncRNA-mRNA interaction network showed that CCNB1 was a target for lncRNA SNHG12 which was the most significant node in lncRNA-mRNA network. A recent study suggested that SNHG12 mediated oxygen-glucose deprivation and induced reoxygenation damage in neurons underlying ischaemic stroke. Our function enrichment analysis showed that SNHG12 was closely related with GO:0055114~oxidation-reduction process and hsa00010:Glycolysis/Gluconeogenesis pathway. Our findings may explain the role of SNHG12 in ischaemic stroke, which suggested that our findings were significant. Besides, SNHG12 was also involved in GO:0006260~DNA replication and GO:0000082~G1/S transition of mitotic cell cycle, which suggested that SNHG12 could regulate the cell proliferation. Recent evidences show that lncRNA SNHG12 promotes proliferation and metastasis of osteosarcoma and papillary thyroid carcinoma (Ding et al. 2018; Zhou et al. 2018), which was consistent with our findings. The oncogenic role of lncRNA SNHG12 is also determined in cervical cancer (Jin et al. 2019) and prostate cancer (Song et al. 2019). However, the prognostic role in lung adenocarcinoma has been reported rarely. In this paper, survival analysis showed that lncRNA SNHG12 was closely associated with prognosis. Thus, we suggested lncRNA SNHG12 as a biomarker for prognosis of lung adenocarcinoma.

There was no clinical samples to verify our results, which was a limitation of our study. Thus, in order to validated our analysis results, we combined the TCGA dataset and another three GEO dataset to detect the expression levels of TTK, CCNB1 and lncRNA SNHG12 in different grades, between metastatic and non-metastatic groups. The results showed that TTK had the highest expression in N3. For SNHG12, higher expression was found in N0 compared with N1 and N2. For CCNB1, higher expression was found in N2 compared with N3. The results suggested that TTK and CCNB1 presented higher expression in higher tumor grade, while SNHG12 had a lower expression in higher tumor grade. Additionally, the expression of TTK in the bone metastasis or lymph node metastasis group showed an upward trend compared with the non-metastasis group, while the expression of CCNB1 in the bone metastasis group showed a downward trend compared with the non-metastasis group, consistent with the results of the original analysis. Furthermore, pathologic M, pathologic N and CCNB1 were independent prognostic factors of lung adenocarcinoma, while age and gender factors did not affect the prognosis.

In recent years, the impact of gender differences on lung cancer has attracted much attention. It has been reported that there is a narrowing of male–female gap in lung cancer rates over the past five years, with rates falling for men (9.2%) and rising for women (6.0%) (Carioli et al. 2020). In the present study, we found that there was no significant difference in survival prognostic information between males and females. A recent study showed that there was no statistically significant association between sex and overall survival in over 13,000 patients with solid tumors treated by anti-PD-L1 checkpoint inhibitors from 23 randomized clinical trials (Wallis et al. 2019). However, our study did not consider the treatment strategy. Additionally, we found that the expression of TTK and CCNB1 was significantly different between males and females. Recently, Mederos et al. (2020) reviewed the gender differences in lung cancer and they concluded that genetic and biological difference between men and women could explain the disparity in incidence and mortality of lung cancers, but much remains unanswered. In conclusion, the mRNA and lncRNA expression profile revealed the differentially expressed mRNAs and lncRNAs based on microarray data. TTK and CCNB1 were differentially expressed in high metastatic lung adenocarcinoma cells and they were the significant nodes in PPI network. lncRNA SNHG12 was identified to be the differentially expressed lncRNA in lung adenocarcinoma cells and was a core node in coexpression network. *TTK*, *CCNB1* and lncRNA SNHG12 were proposed to be the biomarker for the metastasis and prognosis of lung adenocarcinoma.

Data availability statement. Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Conflict of interest. The authors report no conflict of interest.

Funding. None.

References

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. (2000): Gene Ontology: tool for the unification of biology. Nat. Genet. **25,** 25-29
https://doi.org/10.1038/75556

Bandettini WP, Kellman P, Mancini C, Booker OJ, Vasu S, Leung SW, Wilson JR, Shanbhag SM, Chen MY, Arai AE

(2012): MultiContrast Delayed Enhancement (MCODE) improves detection of subendocardial myocardial infarction by late gadolinium enhancement cardiovascular magnetic resonance: a clinical validation study. J. Cardiovasc. Magn. Reson. **14,** 83
https://doi.org/10.1186/1532-429X-14-83

Beer DG, Kardia SL, Huang C-C, Giordano TJ, Levin AM, Misek DE, Lin L, Chen G, Gharib TG, Thomas DG (2002): Gene-expression profiles predict survival of patients with lung adenocarcinoma. Nat. Med. **8,** 816-824
https://doi.org/10.1038/nm733

Behera D, Balamugesh T (2004): Lung cancer in India. Indian J. Chest Dis. Allied Sci. **46,** 269-281

Carioli G, Bertuccio P, Boffetta P, Levi F, La Vecchia C, Negri E, Malvezzi M (2020): European cancer mortality predictions for the year 2020 with a focus on prostate cancer. Ann. Oncol. **31,** 650-658
https://doi.org/10.1016/j.annonc.2020.02.009

Chandrashekar DS, Bashel B, Balasubramanya SAH, Creighton CJ, Ponce-Rodriguez I, Chakravarthi BV, Varambally S (2017): UALCAN: a portal for facilitating tumor subgroup gene expression and survival analyses. Neoplasia **19,** 649-658
https://doi.org/10.1016/j.neo.2017.05.002

Chang JT-H, Lee YM, Huang RS (2015): The impact of the Cancer Genome Atlas on lung cancer. Transl. Res. **166,** 568-585
https://doi.org/10.1016/j.trsl.2015.08.001

Ding K, Li W, Zou Z, Zou X, Wang C (2014): CCNB1 is a prognostic biomarker for ER+ breast cancer. Med. Hypotheses **83,** 359-364
https://doi.org/10.1016/j.mehy.2014.06.013

Ding S, Qu W, Jiao Y, Zhang J, Zhang C, Dang S (2018): LncRNA SNHG12 promotes the proliferation and metastasis of papillary thyroid carcinoma cells through regulating wnt/β-catenin signaling pathway. Cancer Biomark. **22,** 1-10
https://doi.org/10.3233/CBM-170777

Fang Y, Yu H, Liang X, Xu J, Cai X (2014): Chk1-induced CCNB1 overexpression promotes cell proliferation and tumor growth in human colorectal cancer. Cancer Biol. Ther. **15,** 1268-1279
https://doi.org/10.4161/cbt.29691

Ge X, Li G, Jiang L, Jia L, Zhang Z, Li X, Wang R, Zhou M, Zhou Y, Zeng Z (2019): Long noncoding RNA CAR10 promotes lung adenocarcinoma metastasis via miR-203/30/SNAI axis. Oncogene **38,** 3061-3076
https://doi.org/10.1038/s41388-018-0645-x

Jen J, Tang Y-A, Lu Y-H, Lin C-C, Lai W-W, Wang Y-C (2017): Oct4 transcriptionally regulates the expression of long non-coding RNAs NEAT1 and MALAT1 to promote lung cancer progression. Mol. Cancer **16,** 104
https://doi.org/10.1186/s12943-017-0674-z

Ji P, Diederichs S, Wang W, Böing S, Metzger R, Schneider PM, Tidow N, Brandt B, Buerger H, Bulk E (2003): MALAT-1, a novel noncoding RNA, and thymosin β4 predict metastasis and survival in early-stage non-small cell lung cancer. Oncogene **22,** 8031
https://doi.org/10.1038/sj.onc.1206928

Jin XJ, Chen XJ, Zhang ZF, Hu WS, Ou RY, Li S, Xue JS, Chen LL, Hu Y, Zhu H (2019): Long noncoding RNA SNHG12 promotes the progression of cervical cancer via modulating miR-125b/STAT3 axis. J. Cell. Physiol. **234,** 6624-6632
https://doi.org/10.1002/jcp.27403

Kanehisa M, Goto S (2000): KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. **28,** 27-30
https://doi.org/10.1093/nar/28.1.27

Kolde R, Kolde MR (2015): Package ‚Pheatmap'. Version 0.7.4
https://doi.org/10.12688/f1000research.6925.1

Lathwal A, Kumar R, Arora C, Raghava GPS (2020): Identification of prognostic biomarkers for major subtypes of non-small-cell lung cancer using genomic and clinical data. J. Cancer Res. Clin. Oncol. **146,** 2743-2752
https://doi.org/10.1007/s00432-020-03318-3

Li Y, Chen Y-L, Xie Y-T, Zheng L-Y, Han J-Y, Wang H, Tian X-X, Fang W-G (2013): Association study of germline variants in ccnb1 and cdk1 with breast cancer susceptibility, progression, and survival among chinese han women. PLoS One **8,** e84489
https://doi.org/10.1371/journal.pone.0084489

Lin L, Gu Z-T, Chen W-H, Cao K-J (2015): Increased expression of the long non-coding RNA ANRIL promotes lung cancer cell metastasis and correlates with poor prognosis. Diagn. Pathol. **10,** 14
https://doi.org/10.1186/s13000-015-0247-7

Liu W, Ouyang S, Zhou Z, Wang M, Wang T, Qi Y, Zhao C, Chen K, Dai L (2019): Identification of genes associated with cancer progression and prognosis in lung adenocarcinoma: Analyses based on microarray from Oncomine and The Cancer Genome Atlas databases. Mol. Gen. Genom. Med. **7,** e00528
https://doi.org/10.1002/mgg3.528

Lu C, Onn A, Vaporciyan A (2010): Cancer of the lung. In: Holland-Frei Cancer Medicine. People's Medical Publishing House

Maire V, Baldeyron C, Richardson M, Tesson B, Vincent-Salomon A, Gravier E, Marty-Prouvost B, De Koning L, Rigaill G, Dumont A, et al. (2013): TTK/hMPS1 is an attractive therapeutic target for triple-negative breast cancer. PloS One **8,** e63712
https://doi.org/10.1371/journal.pone.0063712

Maxwell O, Kayode AA, Olusola YO, Joshua AI, Chinedu OV (2019): Lung cancer: a chronic disease epidemiology prevalence study. Asian J. Adv. Res. Rep. **2019,** 1-7
https://doi.org/10.9734/ajarr/2019/v3i430096

Mederos N, Friedlaender A, Peters S, Addeo A (2020): Gender-specific aspects of epidemiology, molecular genetics and outcome: lung cancer. ESMO Open **5,** e000796
https://doi.org/10.1136/esmoopen-2020-000796

Miao R, Wu Y, Zhang H, Zhou H, Sun X, Csizmadia E, He L, Zhao Y, Jiang C, Miksad RA, Ghaziani T, Robson SC, Zhao H (2016): Utility of the dual-specificity protein kinase TTK as a therapeutic target for intrahepatic spread of liver cancer. Sci. Rep. **6,** 33121
https://doi.org/10.1038/srep33121

Milatovich A, Francke U (1992): Human cyclin B1 gene (CCNB1) assigned to chromosome 5 (q13-qter). Somat. Cell Mol. Genet. **18,** 303-307
https://doi.org/10.1007/BF01233865

Ni M, Liu X, Wu J, Zhang D, Tian J, Wang T, Liu S, Meng Z, Wang K, Duan X (2018): Identification of candidate biomarkers correlated with the pathogenesis and prognosis of non-small cell lung cancer via integrated bioinformatics analysis. Front. Genet. **9,** 469

https://doi.org/10.3389/fgene.2018.00469
Park S-M, Choi E-Y, Bae D-H, Sohn HA, Kim S-Y, Kim Y-J (2018): The LncRNA EPEL promotes lung cancer cell proliferation through E2F target activation. Cell. Physiol. Biochem. **45,** 1270-1283
https://doi.org/10.1159/000487460
Reinhard C, Jefferson AB, Chan VW (2009): Determining prognosis of colon or breast cancer by measuring TTK expression. Google Patents
Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003): Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Res. **13,** 2498-2504
https://doi.org/10.1101/gr.1239303
Singhal S, Amin K, Kruklitis R, DeLong P, Friscia ME, Litzky LA, Putt ME, Kaiser LR, Albelda SM (2003): Alterations in cell cycle genes in early stage lung adenocarcinoma identified by expression profiling. Cancer Biol. Ther. **2,** 291-298
https://doi.org/10.4161/cbt.2.3.399
Smyth GK (2005): Limma: linear models for microarray data. In: Bioinformatics and Computational Biology Solutions Using R and Bioconductor. (Eds. Gentleman R, Carey VJ, Huber W, Irizarry RA, Dudoit S), pp. 397-420, Springer New York, NY
https://doi.org/10.1007/0-387-29362-0_23
Song J, Wu X, Ma R, Miao L, Xiong L, Zhao W (2019): Long non-coding RNA SNHG12 promotes cell proliferation and activates Wnt/β-catenin signaling in prostate cancer through sponging microRNA-195. J. Cell. Biochem. **120,** 13066-13075
https://doi.org/10.1002/jcb.28578
Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguez P, Doerks T, Stark M, Muller J, Bork P, et al. (2010): The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. Nucleic Acids Res. **39,** D561-568
https://doi.org/10.1093/nar/gkq973
Therneau T, Lumley T (2011): Survival: Survival analysis, including penalised likelihood. R package version 2.36-32.2010
Travis W (2014): The 2015 WHO classification of lung tumors. Pathologe **35,** 188-188
https://doi.org/10.1007/s00292-014-1974-3
Wallis CJ, Butaney M, Satkunasivam R, Freedland SJ, Patel SP, Hamid O, Pal SK, Klaassen Z (2019): Association of patient sex with efficacy of immune checkpoint inhibitors and overall survival in advanced cancers: a systematic review and meta-analysis. JAMA Oncology **5,** 529-536
https://doi.org/10.1001/jamaoncol.2018.5904
Xiang Y, Zhang Y, Tang Y, Li Q (2017): MALAT1 modulates TGF-β1-induced endothelial-to-mesenchymal transition through downregulation of miR-145. Cell. Physiol. Biochem. **42,** 357-372
https://doi.org/10.1159/000477479
Xie Y, Wang A, Lin J, Wu L, Zhang H, Yang X, Wan X, Miao R, Sang X, Zhao H (2017): Mps1/TTK: a novel target and biomarker for cancer. J. Drug Target. **25,** 112-118
https://doi.org/10.1080/1061186X.2016.1258568
Xu Q, Xu Y, Pan B, Wu L, Ren X, Zhou Y, Mao F, Lin Y, Guan J, Shen S (2016): TTK is a favorable prognostic biomarker for triple-negative breast cancer survival. Oncotarget **7,** 81815
https://doi.org/10.18632/oncotarget.13245
Yu G, Wang LG, Han Y, He QY (2012): clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS **16,** 284-287
https://doi.org/10.1089/omi.2011.0118
Yu T, Zhao Y, Hu Z, Li J, Chu D, Zhang J, Li Z, Chen B, Zhang X, Pan H (2017): MetaLnc9 facilitates lung cancer metastasis via a PGK1-activated AKT/mTOR pathway. Cancer Res. **77,** 5782-5794
https://doi.org/10.1158/0008-5472.CAN-17-0671
Zhang J, Ou JX, Bai CX (2011): Tobacco smoking in China: prevalence, disease burden, challenges and future strategies. Respirology **16,** 1165-1172
https://doi.org/10.1111/j.1440-1843.2011.02062.x
Zhou S, Yu L, Xiong M, Dai G (2018): LncRNA SNHG12 promotes tumorigenesis and metastasis in osteosarcoma by upregulating Notch2 by sponging miR-195-5p. Biochem. Biophys. Res. Commun. **495,** 1822-1832
https://doi.org/10.1016/j.bbrc.2017.12.047

Received: July 23, 2021
Final version accepted: January 24, 2022

doi: 10.4149/gpb_2022009

Supplementary Material

Biomarkers associated with metastasis and prognosis of lung adenocarcinoma based on microarray data

Jiang Du[1], Han Qiao[2] and Da-Long Xie[3]

[1] *Department of Thoracic Surgery, Chinese Medical University Affiliated No. 1 Hospital, Shenyang, Liaoning, China*
[2] *Clinical Medicine, Chinese Medical University, Shenyang, Liaoning, China*
[3] *Department of Anatomy, College of Basic Medicine, China Medical University, Shenyang, Liaoning, China*

Supplementary Figure

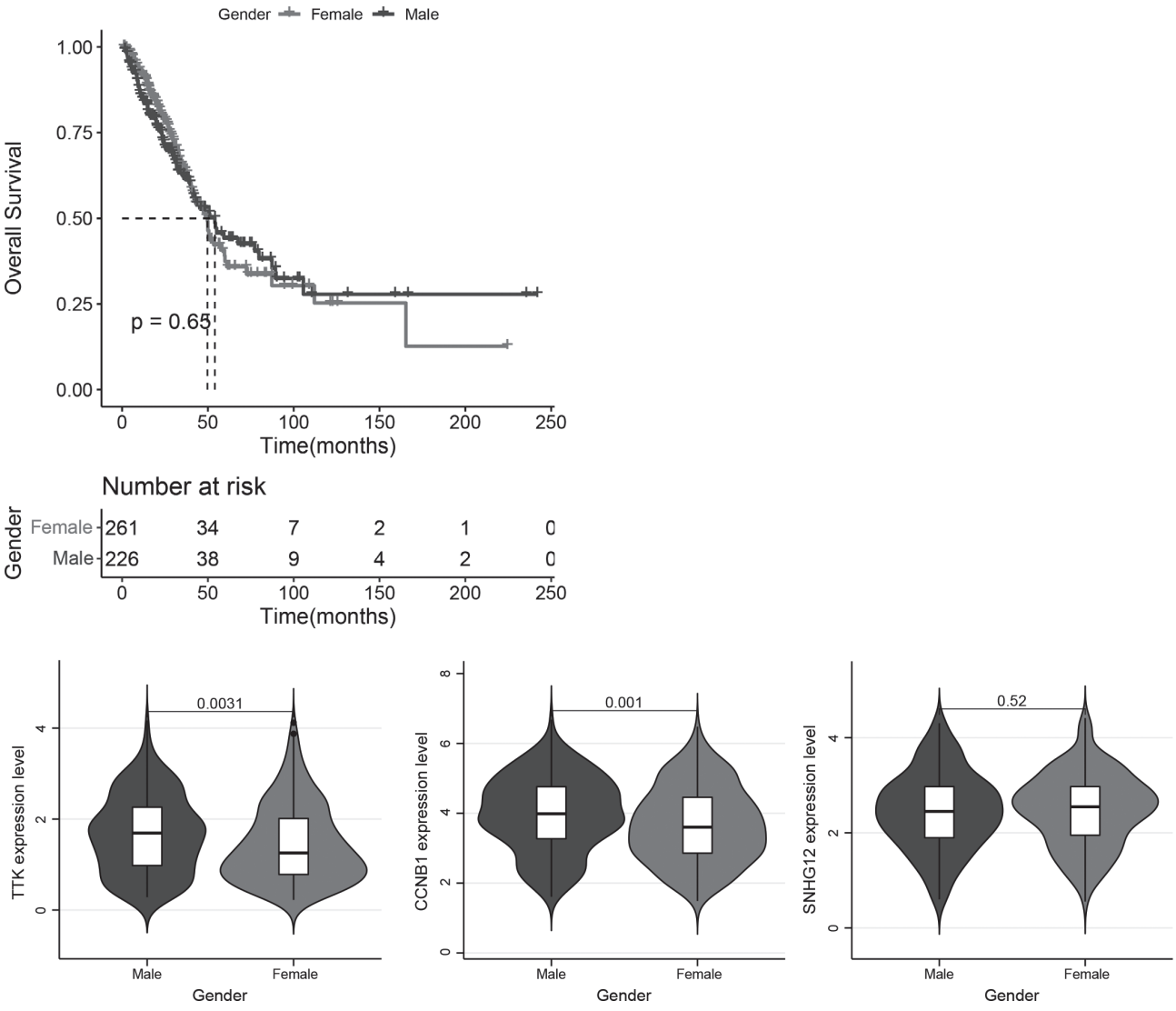

Figure S1. The survival differences (**A**) and the expression differences of key genes (**B**) between male and female based on the TCGA lung adenocarcinoma datasets.