# The Clinical Significance and Potential Molecular Mechanisms of Synaptojanin 2 in Lung Adenocarcinoma

**Hui-Ping Lu**
First Affiliated Hospital of Guangxi Medical University
**Shang-Wei Chen**
First Affiliated Hospital of Guangxi Medical University
**Rong-Quan He**
First Affiliated Hospital of Guangxi Medical University
**Zhi-Guang Huang**
First Affiliated Hospital of Guangxi Medical University
**Jin-Liang Kong**
First Affiliated Hospital of Guangxi Medical University
**Jiang-Hui Zeng**
The Third Affiliated Hospital of Guangxi Medical University
**Jie Yang**
Guangxi Medical University
**Da-Ping Yang**
The Eighth Affiliated Hospital of Guangxi Medical University
**Jun-Xian Mo**
The Seventh Affiliated Hospital of Guangxi Medical University
**Zhong-Qing Tang**
The Seventh Affiliated Hospital of Guangxi Medical University
**Chang-Bo Li**
The Seventh Affiliated Hospital of Guangxi Medical University
**Fu-Chao Ma**
First Affiliated Hospital of Guangxi Medical University
**HuaFu Zhou** ( zhouhuafu_gxmu@163.com )
First Affiliated Hostipal of Guangxi Medical University https://orcid.org/0000-0002-2267-3164
**Gang Chen** ( chengang@gxmu.edu.cn )
First Affiliated Hospital of Guangxi Medical University

**Primary research**

**Keywords:** Synaptojanin 2, lung adenocarcinoma, tissue microarrays, immunohistochemistry, RNA sequencing

**Posted Date:** June 16th, 2020

**DOI:** https://doi.org/10.21203/rs.3.rs-35107/v1

# Abstract

Background Lung adenocarcinoma (LUAD) is the main pathological type of pulmonary malignant tumors. Synaptojanin 2 (SYNJ2), a member of the synaptojanin family, is considered as an attractive underlying therapeutic target in several types of cancer. The purpose of the current research was to comprehensively investigate the molecular function and underlying mechanism of SYNJ2 in LUAD with the goal of providing a promising therapeutic target for LUAD. Methods We applied tissue microarrays, immunohistochemistry (IHC), high-throughput RNA sequencing (RNA-seq) data, and gene microarrays of public databases to comprehensively examine the protein level, mRNA expression, clinical significance, and prognosis value of SYNJ2 in LUAD. The standard mean difference (SMD) and summary receiver operating characteristic curves (SROC) were computed to evaluate the comprehensive expression value of SYNJ2 in LUAD. Kaplan-Meier survival curves were plotted to determine the overall survival of SYNJ2 with different expression levels in LUAD. Univariate and multivariate cox regression analyses were implemented to investigate the dependent prognostic ability of SYNJ2 expression and the clinically-relevant traits of LUAD. Moreover, SYNJ2 co-expressed genes were screened for Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) function analysis and protein–protein interactions (PPI) network construction. Results SYNJ2 overexpression in LUAD was confirmed by 2,299 LUAD samples and 1,387 non-tumor tissues obtained from tissue microarrays along with TCGA and public gene microarrays. The pooled SMD was 1.29 (95% CI: 1.09-1.48, I-square=78.6%, P<0.001) and the area under the curve (AUC) of the SROC was 0.88 (95%CI: 0.85-0.91). Compared to the upregulation of SYNJ2, SYNJ2 downregulation could improve OS in LUAD patients to some extent. SYNJ2 may facilitate the proliferation and progression of LUAD by modulating the cell cycle and TGF-beta signaling pathways. The 10 SYNJ2 co-expressed key genes selected from PPI network were all prominently implicated with the prognosis of LUAD patients. Conclusion SYNJ2 may contribute as an attractive and prospective therapeutic target in LUAD.


# Introduction

Lung cancer (LC) is the second-most-common malignant tumor and the first leading cause of malignancy death among males and females around the world, and the 5 year overall survival (OS) rate of LC is merely 19% for all stages combined [1]. In 2020, an estimated 228,820 new cases of LC will occur, and 135,720 individuals will die of this malignancy in the United States [1]. Lung adenocarcinoma (LUAD), which accounts for nearly 45% of LC cases, is the main pathological type of pulmonary malignant tumors [2-3]. Nearly 85% of LUAD cases eventually die of recurrence and/or metastasis within the first 5 years of diagnosis despite the multiple therapeutic strategies (surgery, chemoradiotherapy, molecular targeted therapy, and immunotherapy) that have been applied in clinical for LUAD patients [4-5]. Since the pathogenesis of LUAD is a very complex pathological process, an in-depth study of this process is essential to discover potential therapeutic targets and effectively treat the disease.

Synaptojanins are a family of phosphatidylinositol (PI) phosphatases, and there are two main types of synaptophysin in the human genome [6]. Synaptojanin 2 (SYNJ2), a member of the synaptojanin family,

can specifically combine to the Rho family GTPase Rac1 and is expressed in a wide range of tissues [7-8]. SYJN2 is an effector of Rac1 which is associated with the formation of lamellipodia and invadopodia [9] as well as with tumor cell migration and invasion, making it an attractive underlying therapeutic target in cancer. Chuang et al. [7] found that SYNJ2 played pivotal roles in glioma cell migration and invasion. A previous study also reported an elevated expression of SYNJ2 in breast carcinoma, and it suggested the gene was a potentially-druggable target to inhibit tumor cell migration [8]. However, there have been no reports about the role of SYNJ2 in LUAD until now. Thus, it is necessary for us to comprehensively investigate the molecular function and underlying mechanism of SYNJ2 in LUAD with the goal of providing a promising therapeutic target to LUAD.

Considering the role of SYNJ2 in other malignancies, the purpose of the current research was to integrate a variety of detection methods to comprehensively investigate the expression status and clinical value of SYNJ2 in LUAD. We applied tissue microarrays, immunohistochemistry (IHC), high-throughput RNA sequencing (RNA-seq) data, and gene microarrays of public databases to comprehensively analyze the role of SYNJ2 in the development and progression of LUAD. Furthermore, in order to reveal the underlying molecular mechanism of SYNJ2 in LUAD, SYNJ2 co-expressed genes were screened by gene ontology (GO) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) function analyses. Then, a protein–protein interactions (PPI) network was constructed to screen LUAD co-expressed key genes for further analysis.

# Materials And Methods

### In-house immunohistochemical detection for SYNJ2

Seven tissue microarrays (LUC481, LUC482, LUC483, LUC961, LUC962, LUC1021, and LUC1051) were obtained from Fanpu Biotech, Inc. (Guilin, China), including 90 samples of LUAD tissue and 55 non-cancerous lung tissues. An IHC analysis of these 145 samples was performed to examine the protein expression level of SYNJ2 in LUAD tissues and non-tumor lung tissues. We also explored the correlation between the SYNJ2 protein expression and clinical features. Two senior pathologists (Gang Chen and Hui-Ping Lu) independently assessed the staining intensity and positive proportion of SYNJ2 by using a semi-quantitative scoring system in a double-blind manner. The scores of staining intensity was divided into 0 (negative), 1 (weak), 2 (medium), and 3 (strong), respectively. Positive staining proportion was classified as <10%, 11%-25%, 26%-50%, 51%-75%, and 76%-100% graded as 0, 1, 2, 3, and 4 point, respectively. The ultimate score was obtained in accordance with the product of the positive proportion and the staining intensity [10-11].

### LUAD data source and extraction of SYNJ2 expression

Public high-throughput RNA-seq and microarray data were derived from the Gene Expression Omnibus (GEO), Oncomine, ArrayExpress, SRA, and The Cancer Genome Atlas (TCGA) databases. The search terms we used to screen the database chips from the GEO were (cancer OR carcinoma OR tumour OR tumor OR adenocarcinoma OR malignan* OR neoplas*) AND (lung OR pulmonary OR respiratory OR bronchi OR

bronchioles OR alveoli OR pneumocytes OR “air way”). The search results were filtered by “series” and "Expression profiling by array." The term “Homo sapiens [Organism]” was applied to limit the search range. In the GEO dataset mining process, we utilized the following inclusion criteria: (1) The microarray included LUAD and adjacent or normal samples, and (2) the chip provided data on the SYNJ2 mRNA primitive expression profile in LUAD and the non-tumor control group. Those that met the following criteria were ruled out: The study only involved only LUAD specimens and no controls, and samples were considered insufficient in the LUAD or non-cancer groups (less than three, respectively).

### Integrated analysis of SYNJ2 expression in LUAD

We integrated the tissue microarrays, high-throughput RNA-seq data, and gene microarrays to comprehensively analyze the expression of SYNJ2 in LUAD. The expression of SYNJ2 in LUAD and noncancerous samples was evaluated using the standardized mean difference (SMD). The I-squared value was calculated to assess heterogeneity. When $I^2 < 50\%$ meant low heterogeneity, the fix effect model was chosen; when $I^2 \geq 50\%$ manifested high heterogeneity, a random effect model was used [12-13]. The outcome was visualized by a forest plot. The summary receiver operating characteristic (SROC) curve was established based on the integrated data to evaluate the potential discrimination ability of SYNJ2 in LUAD.

### Detection of the SYNJ2 gene expression and prognosis value analysis

SYNJ2 expression data for different tumor types and different cancer cell lines were retrieved and compared from The Cancer Cell Line Encyclopedia (CCLE) (https://portals.broadinstitute.org/ccle). The ctcRbase (gene expression database of circulating tumor cells/microemboli) is a database to collect expression data of circulating tumor cells (CTCs) and circulating tumor microemboli (CTM) in all kinds of cancer types, including non-small cell lung cancer (NSCLC). We compared the mRNA expression level of SYNJ2 between primary NSCLC and CTCs from ctcRbase (http://www.origin-gene.cn/database/ctcRbase/index.html). SYNJ2 genetic alterations in LUAD were explored via the cBioPortal (https://cbioportal.org/). For data sets that contained survival times and survival statuses, Kaplan-Meier survival curves were plotted to determine the overall survival of SYNJ2 with different expression levels in LUAD by using the survival R package. Furthermore, univariate and multivariate cox regression analyses were implemented to investigate the dependent prognostic ability of SYNJ2 expression and the clinically-relevant traits of LUAD using R software.

### Acquisition of differentially-expressed and co-expressed genes

The limma R package was applied to identify differentially-expressed genes (DEGs) between LUAD and non-tumor specimens. False discovery rates (FDR) less than 0.05 and |log2 fold change (FC)| > 1 were the standard requests for DEGs. Then, we overlapped the DEGs obtained from each data set included our study by using VENNY 2.1.0 (https://bioinfogp.cnb.csic.es/tools/venny/), and the genes that appeared over seven times in all datasets were considered as the final DEGs of LUAD for next analysis. The Pearson correlation coefficients between the SYNJ2 and protein-coding genes (PCGs) in LUAD from gene

microarrays, found in the TCGA and CCLE databases, were calculated to define the co-expressed relationship of SYNJ2 with PCGs. The genes that met the |Pearson correlation coefficient| greater than 0.5 and P-value lower than 0.001 and appeared more than two times in all datasets were selected as the co-expressed genes of SYNJ2 in LUAD.

### GO and KEGG enrichment analyses

GO and KEGG enrichment analyses of the overlapping genes from SYNJ2 co-expressed genes and DEGs were visualized using a clusterProfiler R function package. A P-adjust value < 0.05 was defined as a significance enrichment.

### Construction of a PPI network and hub genes selection

A PPI network was constructed by the overlapping genes that derived from the co-expressed genes of SYNJ2 and DEGs in The Search Tool for the Retrieval of Interacting Genes (STRING v11.0) (https://string-db.org/). Then, we utilized the Molecular Complex Detection (MCODE) in Cytoscape v3.7.1 to select central modules from the PPI network with following threshold: degree cut-off is 2, node score cut-off is 0.2, max depth is 100, and k-score is 2. Additionally, cytoHubba in Cytoscape v3.7.1 was utilized to select the top 10 hub genes ranked by degree.

### The correlation and prognosis analysis of hub genes

The pheatmap and vioplot R packages were utilized to plot heatmaps and violin maps separately in order to visualize the 10 hub genes expression between LUAD and noncancer tissues. Correlations between SYNJ2 and the 10 hub genes were tested and shown using the corrplot R package. We compared the OS rates between the different expression levels of the 10 hub genes using the survival and survminer R packages.

### Statistical analysis

All primitive data of the RNA-seq expressions were converted to log2. Statistical analyses were applied using SPSS (v23.0, IBM Corp., Armonk, NY, USA), Stata (v12.0, Stata Corp LP, College Station, TX, USA), GraphPad Prism (v8.0, GraphPad Software Inc., La Jolla, CA, USA), and R v3.6.1 software function packages. SPSS 23.0 was used to compute the mean value and standard deviation (SD). The paired/unpaired student’s t-test or the Mann-Whitney test were applied to compare SYNJ2 expression between two groups. The Kruskal-Wallis test was implemented to analyze the distribution difference of SYNJ2 in greater than two groups of clinicopathological traits. Stata 12.0 was executed to plot the SROC curve and to calculate the SMD value. The receiver operating characteristic (ROC) curves and scatterplots in each dataset were constructed by the GraphPad Prism 8.0. Pearson correlation coefficients were utilized to analyze the correlation between SYNJ2 expression and PCGs using R 3.6.1 software.

## Results

### SYNJ2 expression level based on in-house tissue microarray

Compared with non-tumor lung tissues, the protein expression of SYNJ2 in LUAD is significantly greater based on in-house IHC with tissue microarrays (**Figure 1A-1F**). We also explored the relationship between SYNJ2 protein expression and clinical traits based on tissue microarrays. The results implied that the protein expression of SYNJ2 was notably elevated in pathologic T3-4 stage compared with T1-2 stage (**Figure 1G-1H**). However, no differences of SYNJ2 protein expression in LUAD were found in other related clinicopathological features, including age, gender, pathological stage, and N stage (all P > 0.05, **Table S1, Additional file 1**).

### SYNJ2 expression and clinical significance according to public databases

A total of 594 cases of (535 cancer and 59 normal tissues) LUAD patients' data with a differential expression of SYNJ2 were obtained from the TCGA database. The mean ± SD was used to displayed the calculated results. The results illustrated that the expression value of SYNJ2 in LUAD was notably greater than that in non-tumor tissues (3.09±0.58 vs 1.59±0.67, P<0.001, **Figure 2A**). The underlying diagnostic capability of SYNJ2 in LUAD was determined using the ROC curve (AUC=0.92, 95% CI: 0.89-0.94, P<0.001, **Figure 2B**). We also investigated the relationship between the SYNJ2 expression and the clinical parameters of LUAD patients. The expression of SYNJ2 in patients aged <60 years was clearly upregulated (P=0.023, **Figure 2C**), and the AUC of the ROC curve was 0.758 (95% CI: 0.651-0.865, P<0.001, **Figure 2D**). No distinct differences were found in SYNJ2 expression with other relevant clinicopathological traits (all P>0.05, **Table S2, Additional file 2**). In GEO, Oncomine, ArrayExpress, and SRA databases, we screened out 46 chips related to LUAD RNA-seq, namely: E-MTAB-5321, GSE1037, GSE1987, GSE2088, GSE4824 (GPL96), GSE4824 (GPL97), GSE7670, GSE10072, GSE10799, GSE11117, GSE11969, GSE19188, GSE21933, GSE27262, GSE27489, GSE28835, GSE29249, GSE30219, GSE31210, GSE31552, GSE32036 (GPL6684), GSE32665, GSE32863, GSE33532, GSE37759, GSE37764, GSE40275, GSE40419, GSE40791, GSE43458, GSE46539 (GPL6883), GSE46539 (GPL14951), GSE51852, GSE62113, GSE62949, GSE63459, GSE74706, GSE75037, GSE83213, GSE85716, GSE85841, GSE87340, GSE89593, GSE103512, GSE118370, and GSE130779. The expression of SYNJ2 in 14 gene microarrays (including E-MTAB-5321, GSE1987, GSE2088, GSE4824 [GPL96], GSE4824 [GPL97], GSE11117, GSE27489, GSE28835, GSE29249, GSE37759, GSE40275, GSE89593, GSE103512, and GSE118370) contained no significant difference in LUAD when compared to para-tumor tissues; however, the SYNJ2 expression in 32 other gene microarrays was clearly higher in LUAD than non-cancer tissues (**Supplementary Figure 1, Additional file 3**). The AUC value of the ROC curve in each dataset was calculated respectively to assess the diagnostic power of SYNJ2 in LUAD (**Supplementary Figure 2, Additional file 4**). The AUC value ranged from 0.52 to 0.97 in all gene microarrays, indicating that SYNJ2 had a good ability to discriminate LUAD and non-tumor tissues.

In addition, we also explored the relationship between SYNJ2 expression and clinicopathological characters of LUAD from gene microarrays which involved clinicopathological parameters. We discovered that the SYNJ2 expression level was remarkably higher in female patients than in males in LUAD from

the GSE33532 and GSE40791 datasets (**Supplementary Figure 3A-3D, Additional file 5**). The expression of SYNJ2 in patients aged <60 years was also clearly overexpressed in GSE40419 (**Supplementary Figure 3E-3F, Additional file 5**), which was consistent with that in the TCGA dataset. According to the result from GSE31210, the expression of SYNJ2 is significantly elevated in smokers compared with non-smokers (**Supplementary Figure 3G-3H, Additional file 5**).

### Validation of SYNJ2 expression in LUAD with an integrated analysis

In order to evaluate the expression status of SYNJ2 in LUAD, we comprehensively analyzed the expression data of SYNJ2 acquired from high-throughput RNA-seq, tissue microarrays, and all gene microarrays. Because the square value of I was 0.786>0.5, a stochastic effects model was carried out in the current analysis. The forest plot showed that the SYNJ2 expression was clearly increased in LUAD compared to that of non-tumor tissues according to random effects analysis (I-square=78.6%, P<0.001, SMD=1.29, 95%CI: 1.09-1.48, **Figure 3A**). The Begg's and Egger's tests suggested that no obvious publication bias was present (Begg's P=0.689>0.05, Egger's P=0.166>0.05, **Figure 3B**). The AUC value of the SROC curve was 0.88 (95%CI: 0.85-0.91), sensitivity was 0.79 (95%CI: 0.74-0.83), and specificity was 0.83 (95%CI: 0.79-0.87), which illustrated that SYNJ2 has a high potential diagnostic efficiency in LUAD (**Figure 3C**).

### Verification of SYNJ2 expression and prognosis value analysis

The intermediate expression value of SYNJ2 in certain LC cell lines was 2.58, based on CCLE, which is basically consistent with that in the LUAD tissues (**Figure 4A**). Regrettably, no expression levels of SYNJ2 in normal lung epithelial cells were available for the comparison. After mining the OncoPrint data, it was found that SYNJ2 altered in 16/518 (3%) of the sequencing samples, including one amplification, 11 deep deletions, one missense mutation, one mRNA high expression, and two mRNA down-regulations (**Figure 4B**). The median expression value of SYNJ2 in primary NSCLC and CTCs is 1.96 and 2.62, respectively (**Figure 4C**). The result in ctcRbase implied that SYNJ2 has potential value in the non-invasive monitoring of advanced NSCLC, as well as early detection of NSCLC metastases. The Kaplan-Meier survival curve demonstrated no significant difference in OS between a high and a low expression of SYNJ2 based on the TCGA, GSE11969, and GSE87340 datasets (all P>0.05, **Figure 5A-5C**). However, the patients with low expression of SYNJ2 in LUAD had better prognoses than that with high expression according to the GSE31210 microarray (**Figure 5D**). The univariate cox analysis revealed that the expression of SYNJ2 and the pathological stage in GSE31210 were all clearly associated with prognosis of LUAD (**Figure 6A**). In the TCGA dataset, the pathological stage, the T and N stages were also notably correlated with the prognosis on the basis of the cox univariate regression analysis (**Figure 6B**). However, the multivariate cox analysis demonstrated that only the pathological stage could be an independent prognostic marker in LUAD patients in accordance with the TCGA and GSE31210 datasets (**Figure 6C-6D**).

### Identification of DEGs and SYNJ2 co-expressed genes

The limma R package was applied to perform a differential expression analysis in all of the high-throughput RNA-seq datasets. The genes that not only met FDR less than 0.05 and |log2 FC| > 1 but also appeared more than seven times in all of the datasets were considered as DEGs for LUAD. Finally, we got 2,552 DEGs, including 935 upregulation and 1617 downregulation genes. With the criterion of |Pearson correlation coefficient| greater than 0.5 and P-value less than 0.001, we screened the co-expressed genes of SYNJ2 from CCLE and all the high-throughput RNA-seq datasets using R software. Then, we overlapped the results from each dataset to acquire a total of 2,056 SYNJ2 co-expressed genes, which appeared more than two times in all of the datasets. Finally, we overlapped the 2,552 DEGs and 2,056 SYNJ2 co-expressed genes using the Venn diagram package of R software to acquire 694 genes, which were not only LUAD DEGs but also co-expressed genes of SYNJ2 (**Supplementary Figure 4A, Additional file 6**).

### Go and KEGG enrichment analyses

The 694 overlapping genes were utilized for the GO and KEGG function analyses to explore their potential molecular biological functions and signaling pathways in LUAD. The enrichment of the top-10-most-significant biological processes (BP), cellular components (CC), and molecular functions (MF) of these 694 overlapped genes are displayed in **Supplementary Figure 4B (Additional file 6)**. Extracellular structure organization, extracellular matrix (ECM) organization, and cell-substrate adhesion were the most significant in BP. Regarding CC, the overlapping genes were mainly enriched in the ECM, condensed chromosome, kinetochore, and adherens junction. As for MF, the ECM structural constituent, amide binding, growth factor binding, and peptide binding were the most significant enrichment categories. Based on the KEGG pathway analysis, the TGF-beta signaling pathway and the cell cycle were tightly implicated to a malignant tumor (**Supplementary Figure 4C**, **Additional file 6**).

### PPI network and hub genes analysis.

The MCODE in Cytoscape was applied to select clusters in the PPI network which was acquired from the STRING website. A total of four clusters were calculated according to k-core = 2. Cluster 1, which included 33 nodes and 489 edges, had the highest score of 30.56 (**Supplementary Figure 5A, Additional file 7**). Then, cytoHubba in Cytoscape was applied to screen the top 10 hub genes ranked by degree from cluster 1 (**Supplementary Figure 5B, Additional file 7**). Ultimately, we obtained 10 genes (TOP2A, CCNB1, CCNA2, CEP55, KIF20A, NDC80, TPX2, MELK, KIF2C, and KIF11) as SYNJ2 co-expressed key genes. In order to investigate the expression status of these 10 SYNJ2 co-expressed hub genes in LUAD, we analyzed the genes' expression data between LUAD tissues and non-cancer samples from the TCGA, GSE30219, GSE33532, and GSE75037 datasets. The result indicated that the 10 SYNJ2 co-expressed genes were all overexpressed in LUAD when compared to normal lung tissues (**Figure 7A-7H**). The correlations between SYNJ2 and the 10 co-expressed key genes in the TCGA, GSE30219, GSE33532, and GSE75037 datasets were presented in **Figure 8A-8D**, respectively. The results illustrated that there is a positive correlation between SYNJ2 and its 10 co-expressed core genes (all Pearson correlation coefficient >0.5), and the correlation between the 10 SYNJ2 co-expressed hub genes is very positive. We also analyzed the

prognosis significance of 10 co-expressed key genes in the TCGA and GSE31210 datasets. The results indicated that 10 SYNJ2 co-expressed key genes were all prominently implicated with the prognosis of LUAD patients in the two datasets (**Additional file 8: Supplementary Figure 6 and Additional file 9: Supplementary Figure 8**).

# Discussion

As far as we know, there is no relevant research report on the role and mechanism of SYNJ2 in LUAD until now, and the current research is the first to thoroughly explore the expression status, clinical significance, and potential molecular mechanism of SYNJ2 in LUAD. In our study, we applied multiple methods and collected a large number of cases and data to comprehensively examine the mRNA expression, protein level, gene alternation, and prognosis value of SYNJ2 in LUAD. The discoveries of the present study will provide a potential, new direction for future molecular studies on LUAD.

SYNJ2 overexpression in LUAD was confirmed by 2,299 LUAD samples and 1,387 non-tumor tissues obtained from tissue microarrays along with TCGA and public gene microarrays, ensuring the reliability of our results. The expression of SYNJ2 was prominently elevated in LUAD, suggesting that SYNJ2 may play pivotal roles in LUAD development and progression. In the present study, the different expressions of SYNJ2 in females aged <60 years—as well as smokers in some gene microarrays—indicates that the gender, age, and smoking status may be important factors for aberrant expression of SYNJ2 in LUAD. Moreover, the difference in the expression of SYNJ2 between T stages in our tissue microarray suggests that the overexpression of SYNJ2 might stimulate the proliferation of LUAD tumor cells. Compared to the upregulation of SYNJ2, SYNJ2 downregulation could improve OS in LUAD patients to some extent. This observation revealed that the upregulation of SYNJ2 could promote the progression of LUAD. Clearly, the hypotheses above require additional research in the future.

A single gene analysis has some deficiency in predicting the pathogenesis of diseases, as it ignores the interactions between other genes. Therefore, to further reveal the potential molecular mechanism of the SYNJ2 act on LUAD, we performed a functional analysis on SYNJ2 and its 694 co-expressed genes.

According to the GO enrichment analysis, these were mainly enriched in extracellular structure organization, ECM organization, ECM, and ECM structural constituent. The ECM is a dynamic compartment that can modulate cell functions such as migration, proliferation, and differentiation [14]. More and more studies show that ECM plays a pivotal role in the pathogenesis of multiple neoplasms. For example, dynamic imbalance in the remodeling and homeostasis of ECM directly deregulates the conduct of tumor cells and cancer associated fibroblasts (CAF) which leads to cancer progressions [15-17]. Paolillo M et al. [18] pointed out that almost all cellular components of the tumor microenvironment promote the invasion of metastatic cells through different mechanisms, these focused on ECM modifications. Previous research has reported that SYJN2 is an effector of Rac1 which is involved with the formation of lamellipodia and invadopodia; these seem to be the major cellular structures that are responsible for the degradation of extracellular matrixes by cancer cells [9, 19-20]. Therefore, we

speculated that SYNJ2, acting as oncogene, could promote invadopodia formation in LUAD, thus promoting ECM degradation and facilitating the migration and invasion of LUAD cells. Certainly, this hypothesis needs to be verified further by future research.

Two pathways (the cell cycle and TGF-beta signaling pathway) were discovered to be closely correlated with human cancers based on a KEGG pathway analysis of SYNJ2 and its co-expressed genes. Relevant studies have confirmed that TGF-beta overexpression in LC [21-23], glioma [24], gastric carcinoma [25], prostate cancer [26], and colorectal cancer [27], which associates with these tumors progression and clinical prognosis. Furthermore, TGF-beta signaling facilitates epithelial to mesenchymal transition (EMT); this is a trait of invasive and metastatic cells [28-29], and it has been confirmed that the systemic inhibition of TGF-β can inhibit metastasis [30-31]. Considering the role of TGF-beta signaling pathways in tumors, it is inferred that SYNJ2 and its co-expressed genes can promote LUAD cells' infiltration and migration by TGF-beta signaling pathways. Since cancer is characterized by the deregulation of cell cycle activity that lead to aberrant cell proliferation [32-33], many authors have concluded that cancer can be regarded as a disease of the cell cycle [34]. Previous studies reported that many genes can regulate cell cycles to promote LC cell proliferations [35-38]. Disappointingly, there is no relevant study about the molecular mechanism of SYNJ2 in LUAD. Hence, we speculated that one of the oncogenic mechanisms of SYNJ2 in LUAD is promoting the proliferation and growth of LUAD cells by regulating the cell cycle pathway.

To further investigate the interactions between SYNJ2 and its co-expressed genes, we constructed a PPI network using the STRING database and screened out 10 SYNJ2 co-expressed hub genes (TOP2A, CCNB1, CCNA2, CEP55, KIF20A, NDC80, TPX2, MELK, KIF2C, and KIF11) in Cytoscape. These 10 co-expressed hub genes were all associated with LUAD prognosis. TOP2A has been reported in the upregulation of various malignant tumors, including LUAD. For instance, Guo et al. [39] found that TOP2A overexpression was correlated with poor prognosis of LUAD patients. Wang et al. [40] revealed that the higher the CCNB1 expression level, the lower the OS rate and disease-free survival rate in LC patients. Our study demonstrated that TOP2A and CCNB1 expression was all elevated in LUAD, implying a dismal outcome in OS that was consistent with previous research. It has been proven that CCNA2 is prominently related to cell cycles and cell proliferation [41-42]. Therefore, we have reason to believe that SYNJ2 and CCNA2 both participate in the process of the cell cycle to promote LUAD cells' proliferation and progression. It has been reported that CEP55 can facilitate the proliferation, migration, and infiltration of multiple neoplasms, including glioma [43], breast cancer [44], gastric carcinoma [45], and prostate cancer [46]. Wu et al. [47] discovered that CEP55 is implicated with LUAD prognosis, which is also consistent with our results. Therefore, we argue that SYNJ2 may act with CEP55 to promote the proliferation, migration, and invasion of LUAD. KIF20A also has been detected as being overexpressed in numerous types of human malignancies, including LUAD, and its overexpression can stimulate LUAD cells' proliferation [48]. Matching our results, Sun et al. [49] also reported that NDC80 expression was strikingly elevated in LUAD and related to prognosis. Interestingly, relevant studies about TPX2, MELK, KIF2C, and KIF11 also have been reported in LUAD, and they are all implicated with the progression and poor prognosis of LUAD [50-53]. In summary, all the 10 SYNJ2 co-expressed key genes were overexpressed in

LUAD and were clearly correlated with the prognosis of LUAD. We have reason to believe that SYNJ2 might act cooperatively with these 10 co-expressed gens to promote the proliferation, migration, and progression of LUAD. Of course, the mechanism of interaction between these 10 genes and SYNJ2 in LUAD requires further study.

There are some limitations of the current research. First, retrospective, selective, and confounding biases were inevitable. Second, the protein and mRNA expression levels of SYNJ2 could be further validated using a Western blot analysis, enzyme-linked immunosorbent assay (ELISA), and quantitative real-time PCR (RT-qPCR). Third, the interactions between SYNJ2 and its co-expressed genes should be further confirmed by an experiment such as RNA Binding Protein Immunoprecipitation.

# Conclusion

The expression level of SYNJ2 was significantly elevated in LUAD, and the upregulation of SYNJ2 may facilitate the proliferation and progression of LUAD by modulating the cell cycle and TGF-beta signaling pathways. SYNJ2 may contribute as an attractive and prospective therapeutic target in LUAD.

# Abbreviations

LC: Lung cancer; OS: overall survival; LUAD: Lung adenocarcinoma; SYNJ2: Synaptojanin 2; PI: Phosphatidylinositol; IHC: Immunohistochemistry; RNA-seq: RNA sequencing; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; PPI: Protein–protein interaction; GEO: Gene Expression Omnibus; TCGA: The Cancer Genome Atlas; SMD: Standard mean difference; SROC: Summary receiver operating characteristic; CCLE: Cancer Cell Line Encyclopedia; CTCs: Circulating tumor cells; CTM: Circulating tumor microemboli; NSCLC: Non-small cell lung cancer; DEGs: Differentially-expressed genes; FDR: False discovery rates; FC: Fold change; PCGs: Protein-coding genes; STRING: Search Tool for the Retrieval of Interacting Genes; MOCDE: Molecular Complex Detection; SD: standard deviation; ROC: Receiver operating characteristic; BP: Biological processes; CC: Cellular components; MF: Molecular functions; ECM: Extracellular matrix; CAF: Cancer associated fibroblasts; EMT: Epithelial to mesenchymal transition; ELISA: Enzyme-linked immunosorbent assay; RT-qPCR: Quantitative real-time PCR.

# Declarations

### Ethics approval and consent to participate

The study protocol was approved by the Ethics Committee of the First Affiliated Hospital of Guangxi Medical University, with informed consent signed by all participants.

### Consent for publication

Not applicable.

### Availability of data and materials

All data generated or analyzed during this study are included in this article.

### Competing interests

The authors declare that they have no competing interests


### Funding

This study was supported by Natural Science Foundation of Guangxi, China (2017GXNSFAA198016); Innovation Project of Guangxi Graduate Education (YCBZ2020045); Guangxi Degree and Postgraduate Education Reform and Development Research Projects, China (JGY2019050); Guangxi Medical University Training Program for Distinguished Young Scholars; Medical Excellence Award Funded by the Creative Research Development Grant from the First Affiliated Hospital of Guangxi Medical University.


### Authors' contributions

All authors have contributed to this study for submission. GC, HFZ, HPL, and SWC conception and design. HPL and SWC performed the data extraction and statistical analysis and drafted the paper. RQH, ZGH, JLK, JHZ, JY, DPY, and JXM contributed to supervise all experiments and correct the paper. HPL, SWC, ZQT, CBL, and FCM collected, extracted and analyzed the data. GC, HFZ, HPL, and SWC critically revised the manuscript. All authors have contributed to, read and approved the final manuscript for submission.


### Acknowledgements

Not applicable.


## References


[1] Siegel R L, Miller K D, Jemal A. Cancer statistics, 2020. CA: a cancer journal for clinicians. 2020;70:7-30.

[2] Wang X, Yao S, Xiao Z, Gong J, Liu Z, Han B, et al. Development and validation of a survival model for lung adenocarcinoma based on autophagy-associated genes. J Transl Med. 2020;18:149.

[3] Li J P, Hsieh M J, Chou Y E, Chao Y H, Tsao T C, Yang S F. CD44 Gene Polymorphisms as a Risk Factor for Susceptibility and Their Effect on the Clinicopathological Characteristics of Lung Adenocarcinoma in Male Patients. Int J Environ Res Public Health. 2020;17.

[4] Bray F, Ferlay J, Soerjomataram I, Siegel R L, Torre L A, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68:394-424.

[5] Miller K D, Nogueira L, Mariotto A B, Rowland J H, Yabroff K R, Alfano C M, et al. Cancer treatment and survivorship statistics, 2019. CA Cancer J Clin. 2019;69:363-385.

[6] Chuang Y Y, Tran N L, Rusk N, Nakada M, Berens M E, Symons M. Role of synaptojanin 2 in glioma cell migration and invasion. Cancer Res. 2004;64:8271-5.

[7] Malecz N, McCabe P C, Spaargaren C, Qiu R, Chuang Y, Symons M. Synaptojanin 2, a novel Rac1 effector that regulates clathrin-mediated endocytosis. Curr Biol. 2000;10:1383-6.

[8] Ben-Chetrit N, Chetrit D, Russell R, Körner C, Mancini M, Abdul-Hai A, et al. Synaptojanin 2 is a druggable mediator of metastasis and the gene is overexpressed and amplified in breast cancer. Sci Signal. 2015;8:ra7.

[9] Nakahara H, Otani T, Sasaki T, Miura Y, Takai Y, Kogo M. Involvement of Cdc42 and Rac small G proteins in invadopodia formation of RPMI7951 cells. Genes Cells. 2003;8:1019-27.

[10] Zhang Y, Li Z Y, Hou X X, Wang X, Luo Y H, Ying Y P, et al. Clinical significance and effect of AEG-1 on the proliferation, invasion, and migration of NSCLC: a study based on immunohistochemistry, TCGA, bioinformatics, in vitro and in vivo verification. Oncotarget. 2017;8:16531-16552.

[11] Yang X, Mo W, Fang Y, Wei G, Wei M, Dang Y, et al. Up-regulation of Polo-like Kinase 1 in nasopharyngeal carcinoma tissues: a comprehensive investigation based on RNA-sequencing, gene chips, and in-house tissue arrays. Am J Transl Res. 2018;10:3924-3940.

[12] Ioannidis J P, Patsopoulos N A, Evangelou E. Uncertainty in heterogeneity estimates in meta-analyses. BMJ. 2007;335:914-6.

[13] Higgins J P, Thompson S G, Deeks J J, Altman D G. Measuring inconsistency in meta-analyses. BMJ. 2003;327:557-60.

[14] Bonnans C, Chou J, Werb Z. Remodelling the extracellular matrix in development and disease. Nat Rev Mol Cell Biol. 2014;15:786-801.

[15] Ando A, Hashimoto N, Sakamoto K, Omote N, Miyazaki S, Nakahara Y, et al. Repressive role of stabilized hypoxia inducible factor 1α expression on transforming growth factor β-induced extracellular matrix production in lung cancer cells. Cancer Sci. 2019;110:1959-1973.

[16] Cox T R, Erler J T. Remodeling and homeostasis of the extracellular matrix: implications for fibrotic diseases and cancer. Dis Model Mech. 2011;4:165-78.

[17] Lu P, Weaver V M, Werb Z. The extracellular matrix: a dynamic niche in cancer progression. J Cell Biol. 2012;196:395-406.

[18] Paolillo M, Schinelli S. Extracellular Matrix Alterations in Metastatic Processes. Int J Mol Sci. 2019;20.

[19] Buccione R, Orth J D, McNiven M A. Foot and mouth: podosomes, invadopodia and circular dorsal ruffles. Nat Rev Mol Cell Biol. 2004;5:647-57.

[20] Chen W T. Proteolytic activity of specialized surface protrusions formed at rosette contact sites of transformed cells. J Exp Zool. 1989;251:167-85.

[21] Domagała-Kulawik J, Hoser G, Safianowska A, Grubek-Jaworska H, Chazan R. Elevated TGF-beta1 concentration in bronchoalveolar lavage fluid from patients with primary lung cancer. Arch Immunol Ther Exp (Warsz). 2006;54:143-7.

[22] Barthelemy-Brichant N, David J L, Bosquée L, Bury T, Seidel L, Albert A, et al. Increased TGFbeta1 plasma level in patients with lung cancer: potential mechanisms. Eur J Clin Invest. 2002;32:193-8.

[23] Hasegawa Y, Takanashi S, Kanehira Y, Tsushima T, Imai T, Okumura K. Transforming growth factor-beta1 level correlates with angiogenesis, tumor progression, and prognosis in patients with nonsmall cell lung carcinoma. Cancer. 2001;91:964-71.

[24] Bruna A, Darken R S, Rojo F, Ocaña A, Peñuelas S, Arias A, et al. High TGFbeta-Smad activity confers poor prognosis in glioma patients and promotes cell proliferation depending on the methylation of the PDGF-B gene. Cancer Cell. 2007;11:147-60.

[25] Saito H, Tsujitani S, Oka S, Kondo A, Ikeguchi M, Maeta M, et al. The expression of transforming growth factor-beta1 is significantly correlated with the expression of vascular endothelial growth factor and poor prognosis of patients with advanced gastric carcinoma. Cancer. 1999;86:1455-62.

[26] Wikström P, Stattin P, Franck-Lissbrant I, Damber J E, Bergh A. Transforming growth factor beta1 is associated with angiogenesis, metastasis, and poor clinical outcome in prostate cancer. Prostate. 1998;37:19-29.

[27] Tsushima H, Kawata S, Tamura S, Ito N, Shirai Y, Kiso S, et al. High levels of transforming growth factor beta 1 in patients with colorectal cancer: association with disease progression. Gastroenterology. 1996;110:375-82.

[28] Oft M, Heider K H, Beug H. TGFbeta signaling is necessary for carcinoma cell invasiveness and metastasis. Curr Biol. 1998;8:1243-52.

[29] Deckers M, van Dinther M, Buijs J, Que I, Löwik C, van der Pluijm G, et al. The tumor suppressor Smad4 is required for transforming growth factor beta-induced epithelial to mesenchymal transition and bone metastasis of breast cancer cells. Cancer Res. 2006;66:2202-9.

[30] Bandyopadhyay A, Agyin J K, Wang L, Tang Y, Lei X, Story B M, et al. Inhibition of pulmonary and skeletal metastasis by a transforming growth factor-beta type I receptor kinase inhibitor. Cancer Res. 2006;66:6714-21.

[31] Biswas S, Guix M, Rinehart C, Dugger T C, Chytil A, Moses H L, et al. Inhibition of TGF-β with neutralizing antibodies prevents radiation-induced acceleration of metastatic cancer progression. J Clin Invest. 2017;127:1116.

[32] Otto T, Sicinski P. Cell cycle proteins as promising targets in cancer therapy. Nat Rev Cancer. 2017;17:93-115.

[33] Cai Z, Liu Q. Cell Cycle Regulation in Treatment of Breast Cancer. Adv Exp Med Biol. 2017;1026:251-270.

[34] Diaz-Moralli S, Tarrado-Castellarnau M, Miranda A, Cascante M. Targeting cell cycle regulation in cancer therapy. Pharmacol Ther. 2013;138:255-71.

[35] Wang H, Ye L, Xing Z, Li H, Lv T, Liu H, et al. CDCA7 promotes lung adenocarcinoma proliferation via regulating the cell cycle. Pathol Res Pract. 2019;215:152559.

[36] Ge H, Cheng N, Xu X, Yang Z, Hoffman R M, Zhu J. AMMECR1 Inhibits Apoptosis and Promotes Cell-cycle Progression and Proliferation of the A549 Human Lung Cancer Cell Line. Anticancer Res. 2019;39:4637-4642.

[37] Cheng L, Yang Q, Li C, Dai L, Yang Y, Wang Q, et al. DDA1, a novel oncogene, promotes lung cancer progression through regulation of cell cycle. J Cell Mol Med. 2017;21:1532-1544.

[38] Luo J, Liu K, Yao Y, Sun Q, Zheng X, Zhu B, et al. DMBX1 promotes tumor proliferation and regulates cell cycle progression via repressing OTX2-mediated transcription of p21 in lung adenocarcinoma cell. Cancer Lett. 2019;453:45-56.

[39] Guo W, Sun S, Guo L, Song P, Xue X, Zhang H, et al. Elevated TOP2A and UBE2C expressions correlate with poor prognosis in patients with surgically resected lung adenocarcinoma: a study based on immunohistochemical analysis and bioinformatics. J Cancer Res Clin Oncol. 2020;146:821-841.

[40] Wang F, Chen X, Yu X, Lin Q. Degradation of CCNB1 mediated by APC11 through UBA52 ubiquitination promotes cell cycle progression and proliferation of non-small cell lung cancer cells. Am J Transl Res. 2019;11:7166-7185.

[41] Ma Q. MiR-219-5p suppresses cell proliferation and cell cycle progression in esophageal squamous cell carcinoma by targeting CCNA2. Cell Mol Biol Lett. 2019;24:4.

[42] Gan Y, Li Y, Li T, Shu G, Yin G. CCNA2 acts as a novel biomarker in regulating the growth and apoptosis of colorectal cancer. Cancer Manag Res. 2018;10:5113-5124.

[43] Li F, Jin D, Guan L, Zhang C C, Wu T, Wang Y J, et al. CEP55 promoted the migration, invasion and neuroshpere formation of the glioma cell line U251. Neurosci Lett. 2019;705:80-86.

[44] Yin Y, Cai J, Meng F, Sui C, Jiang Y. MiR-144 suppresses proliferation, invasion, and migration of breast cancer cells through inhibiting CEP55. Cancer Biol Ther. 2018;19:306-315.

[45] Tao J, Zhi X, Tian Y, Li Z, Zhu Y, Wang W, et al. CEP55 contributes to human gastric carcinoma by regulating cell proliferation. Tumour Biol. 2014;35:4389-99.

[46] You B, Zhang K C. MicroRNA-144-3p inhibits cell proliferation and promotes apoptosis in castration-resistant prostate cancer by targeting CEP55. Eur Rev Med Pharmacol Sci. 2018;22:7660-7670.

[47] Wu S, Wu D, Pan Y, Liu H, Shao Z, Wang M. Correlation between EZH2 and CEP55 and lung adenocarcinoma prognosis. Pathol Res Pract. 2019;215:292-301.

[48] Zhao X, Zhou L L, Li X, Ni J, Chen P, Ma R, et al. Overexpression of KIF20A confers malignant phenotype of lung adenocarcinoma by promoting cell proliferation and inhibiting apoptosis. Cancer Med. 2018;7:4678-4689.

[49] Sun Z Y, Wang W, Gao H, Chen Q F. Potential therapeutic targets of the nuclear division cycle 80 (NDC80) complexes genes in lung adenocarcinoma. J Cancer. 2020;11:2921-2934.

[50] Zhang M Y, Liu X X, Li H, Li R, Liu X, Qu Y Q. Elevated mRNA Levels of AURKA, CDC20 and TPX2 are associated with poor prognosis of smoking related lung adenocarcinoma using bioinformatics analysis. Int J Med Sci. 2018;15:1676-1685.

[51] Zang X, Qian C, Ruan Y, Xie J, Luo T, Xu B, et al. Higher maternal embryonic leucine zipper kinase mRNA expression level is a poor prognostic factor in non-small-cell lung carcinoma patients. Biomark Med. 2019;13:1349-1361.

[52] Bai Y, Xiong L, Zhu M, Yang Z, Zhao J, Tang H. Co-expression network analysis identified KIF2C in association with progression and prognosis in lung adenocarcinoma. Cancer Biomark. 2019;24:371-382.

[53] Li S, Xuan Y, Gao B, Sun X, Miao S, Lu T, et al. Identification of an eight-gene prognostic signature for lung adenocarcinoma. Cancer Manag Res. 2018;10:3383-3392.

# Additional Files

**Additional file 1: Table S1.** Relationship between SYNJ2 expression and clinicopathological parameters of LUAD from tissue microarray.

**Additional file 2: Table S2.** Relationship between SYNJ2 expression and clinicopathological parameters of LUAD from TCGA.

**Additional file 3: Supplementary Figure 1. SYNJ2 expression in LUAD and non-tumor tissues based on gene microarrays**. The scatter plots display the differential expression levels of SYNJ2 in LUAD and non-cancer tissues for each of the included microarray datasets. Data are expressed as means ± SD, and P<0.05 indicates a statistically significant difference when compared to the normal control.

**Additional file 4: Supplementary Figure 2. ROC curves based on gene microarray datasets**. A panel of ROC curves shows the diagnostic ability of SYNJ2 for LUSC in each of the included gene microarray datasets. AUC: 0.5–0.7 (low), 0.7–0.9 (moderate), and 0.9–1.0 (high). P<0.05 indicates a statistically significant difference. ROC, receiver operating characteristic; AUC, area under the curve; CI, confidence interval.

**Additional file 5: Supplementary Figure 3. The relationship between SYNJ2 expression and related clinical features in different datasets**. **A-B.** The expression levels of SYNJ2 between female and male patients based on GSE33532 dataset. **C-D**. SYNJ2 expression level in female and male patients according to GSE40791 dataset. **E-F**. Expression levels of SYNJ2 between patients of different ages based on GSE40419 dataset. **G-H.** SYNJ2 expression in patients with different smoking statuses in accordance with GSE31210 dataset**.**

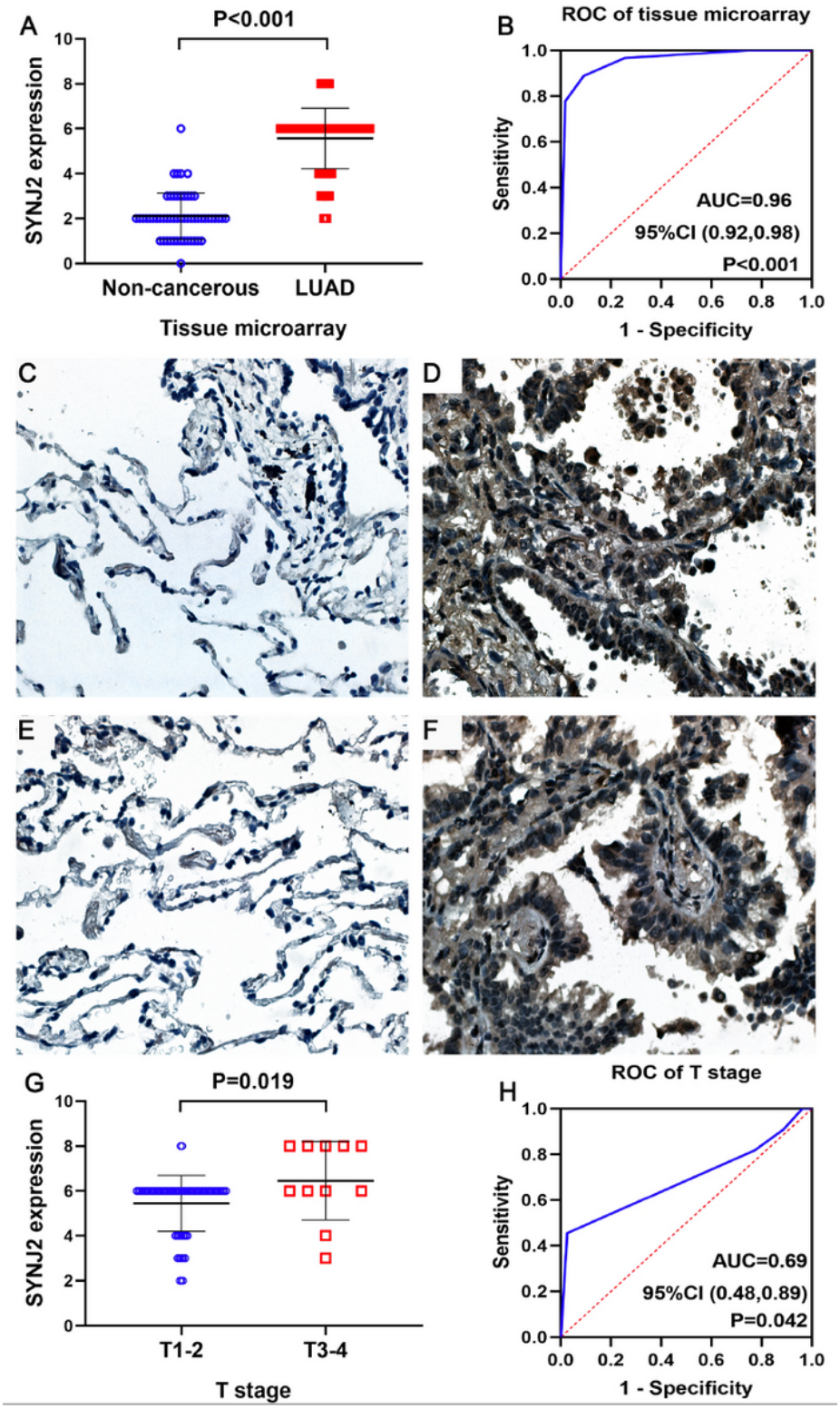
**Additional file 6: Supplementary Figure 4. GO and KEGG analyses of the overlapping genes from SYNJ2 co-expressed genes and DEGs**. **A**. Venn diagram of overlapping genes from the intersection of two independent datasets. **B.** GO enrichment analysis bar diagram. **C**. KEGG pathway enrichment analysis bar diagram. The x-axis represents the numbers of involved genes, and the y-axis represents the GO and KEGG terms. Each bar represents a term. The length of the bar indicates the number of involved genes. The color hue indicates the P-adjust value. Red indicates higher degrees of significance of gene enrichment analysis than blue. DEGs, differentially expressed genes; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; BP, biological process; CC, cellular component; MF, molecular function.

**Additional file 7: Supplementary Figure 5**. **Module analysis and hub genes selection**. **A**. Module rank 1 obtained from PPI network using Molecular Complex Detection (MCODE) in Cytoscape. This cluster consists of 33 nodes and 489 edges and has the highest score of the clusters. **B.** Ten SYNJ2 co-expressed hub genes screened from module cluster 1 using cytoHubba in Cytoscape.

**Additional file 8: Supplementary Figure 6. Survival analysis of 10 SYNJ2 co-expressed hub genes based on TCGA dataset**. The result indicates that they are all related to the prognosis of patients with LUAD. P<0.05 indicates a statistically significant difference.

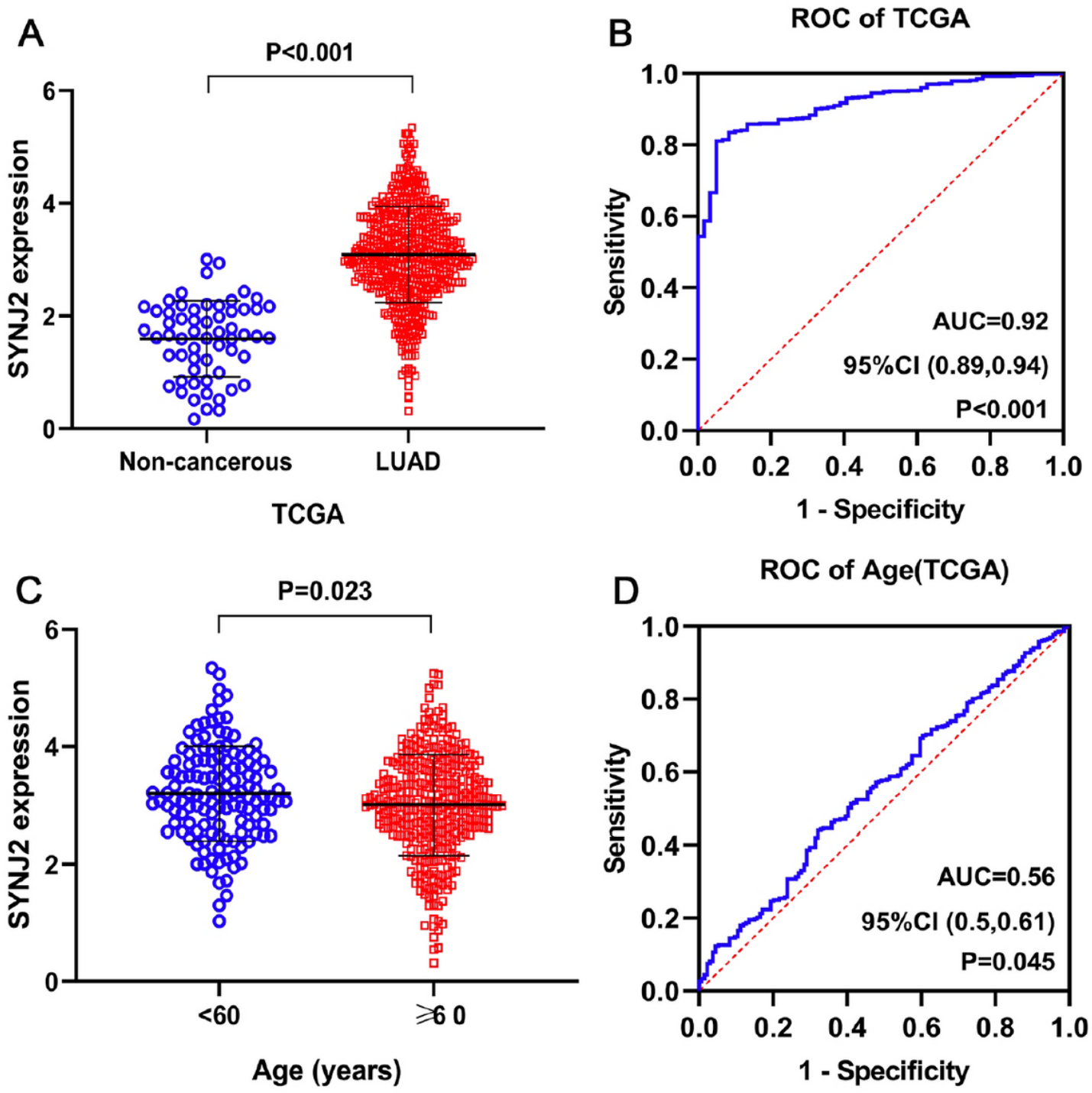
**Additional file 9: Supplementary Figure 7**. **Survival analysis of 10 SYNJ2 co-expressed hub genes based on GSE31210 dataset**. The result demonstrates that they are all related to the prognosis of patients with LUAD. P<0.05 indicates a statistically significant difference.

# Figures



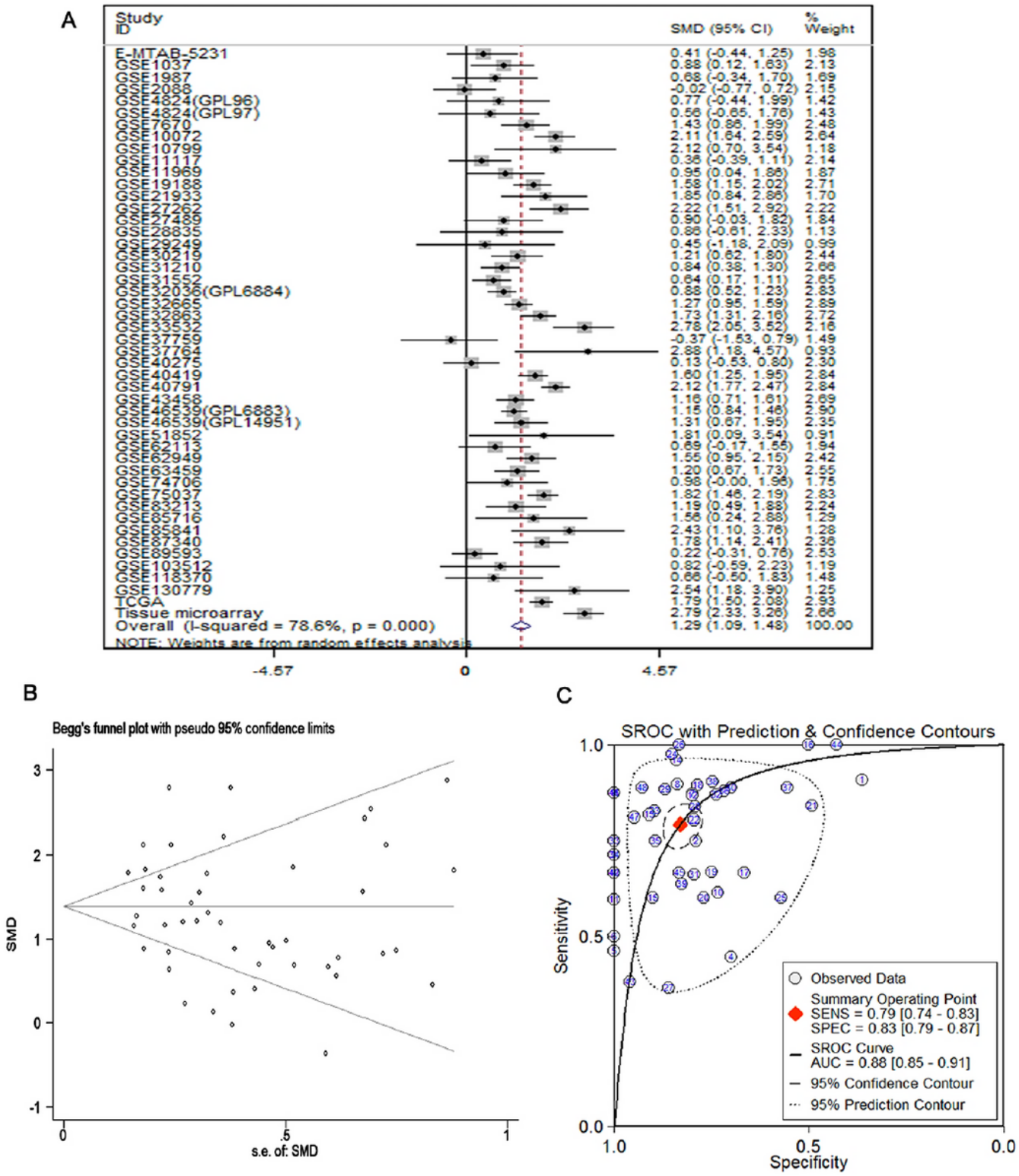

## Figure 1

SYNJ2 protein overexpression in lung adenocarcinoma (LUAD) base on in-house tissue microarrays. A. The protein expression of SYNJ2 between LUAD and non-cancerous tissues. B. ROC curve of SYNJ2 protein expression in LUAD. C, E. Normal lung tissues (400×). D, F. LUAD tissues (400×). G-H. Expression level of SYNJ2 protein between patients in T1, T2 and T3, T4 stages. LUAD, lung adenocarcinoma; ROC, receiver operating characteristic; AUC, area under the curve; CI, confidence interval.




## Figure 2

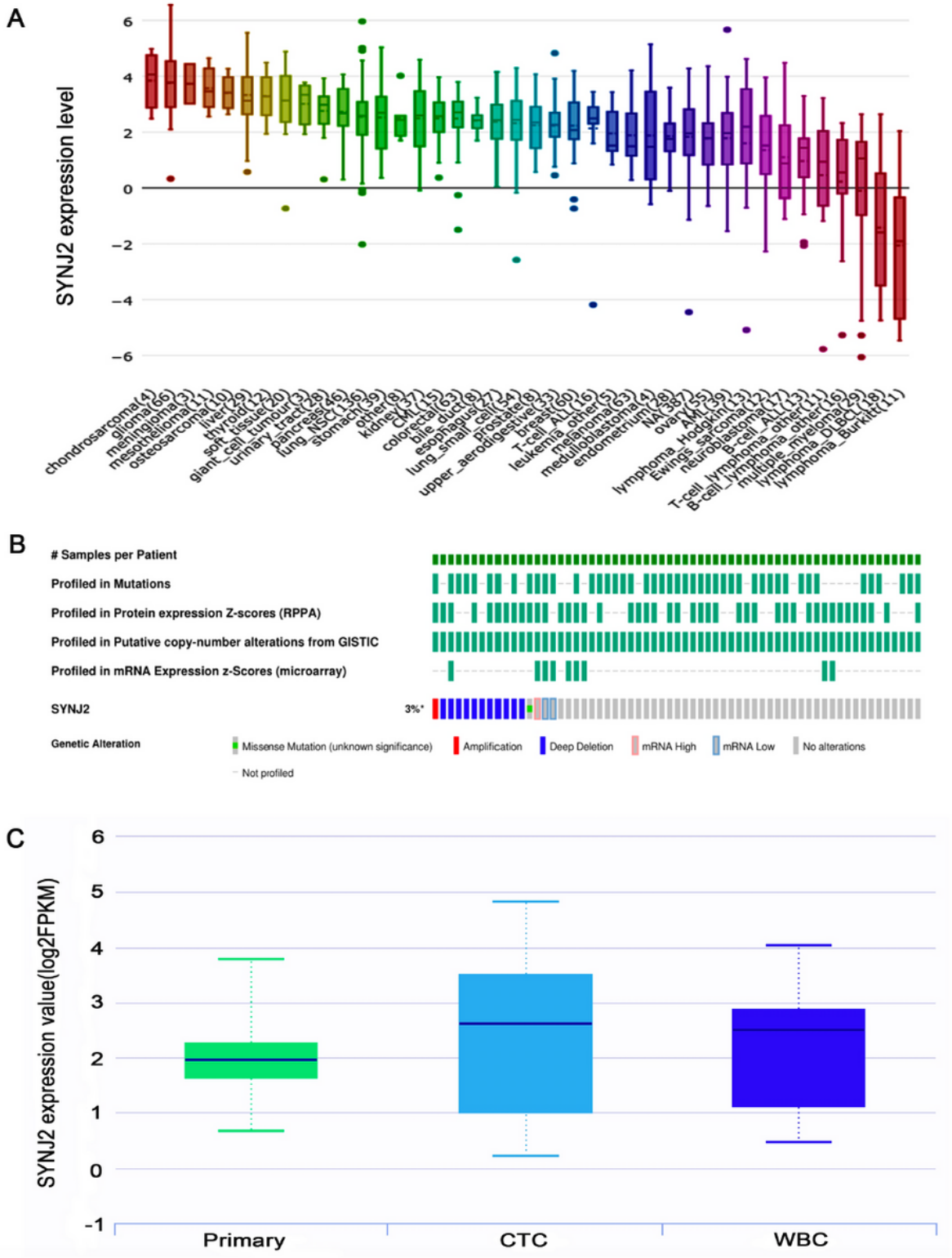
Expression of SYNJ2 in LUAD and its association with clinicopathological parameters from TCGA. A. The expression of SYNJ2 in LUAD and non-tumor tissues. B. The ROC curve was generated to assess the diagnostic ability of SYNJ2 in LUAD and non-cancerous lung tissues (AUC=0.92, 95%CI: 0.89-0.94, P<0.001). C. Expression of SYNJ2 for age of LUAD patients. D. ROC curve of SYNJ2 for age of LUAD (AUC=0.56, 95%CI: 0.50-0.61, P=0.045). LUAD, Lung Adenocarcinoma; TCGA, The Cancer Genome Atlas; ROC, receiver operating characteristic; AUC, area under the curve; CI, confidence interval.




## Figure 3

Integrated analysis of the tissue microarray, TCGA dataset, and public gene microarrays. A. Forest plot of SMD to validate the high expression of SYNJ2 in LUAD. B. The Begg's funnel plot of the publication bias. C. SROC curve of SYNJ2 in the diagnosis ability of LUAD data from all involved datasets (AUC=0.88, 95%CI: 0.85-0.91). P<0.05 indicates a statistically significant difference. SMD, standard mean difference;

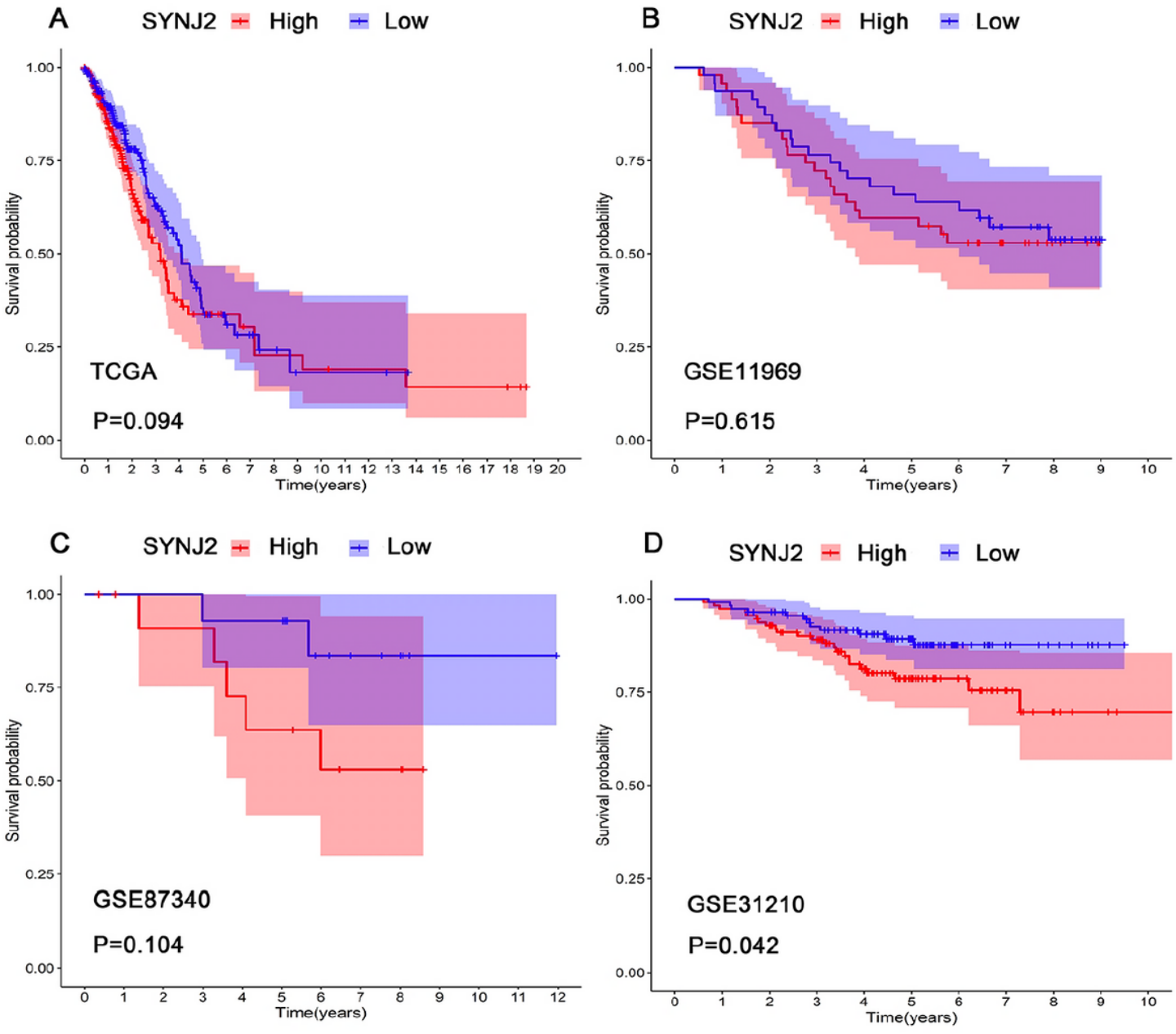
TCGA, The Cancer Genome Atlas; CI, confidence interval; SROC, summary receiver operating characteristic; AUC, area under the curve.




## Figure 4

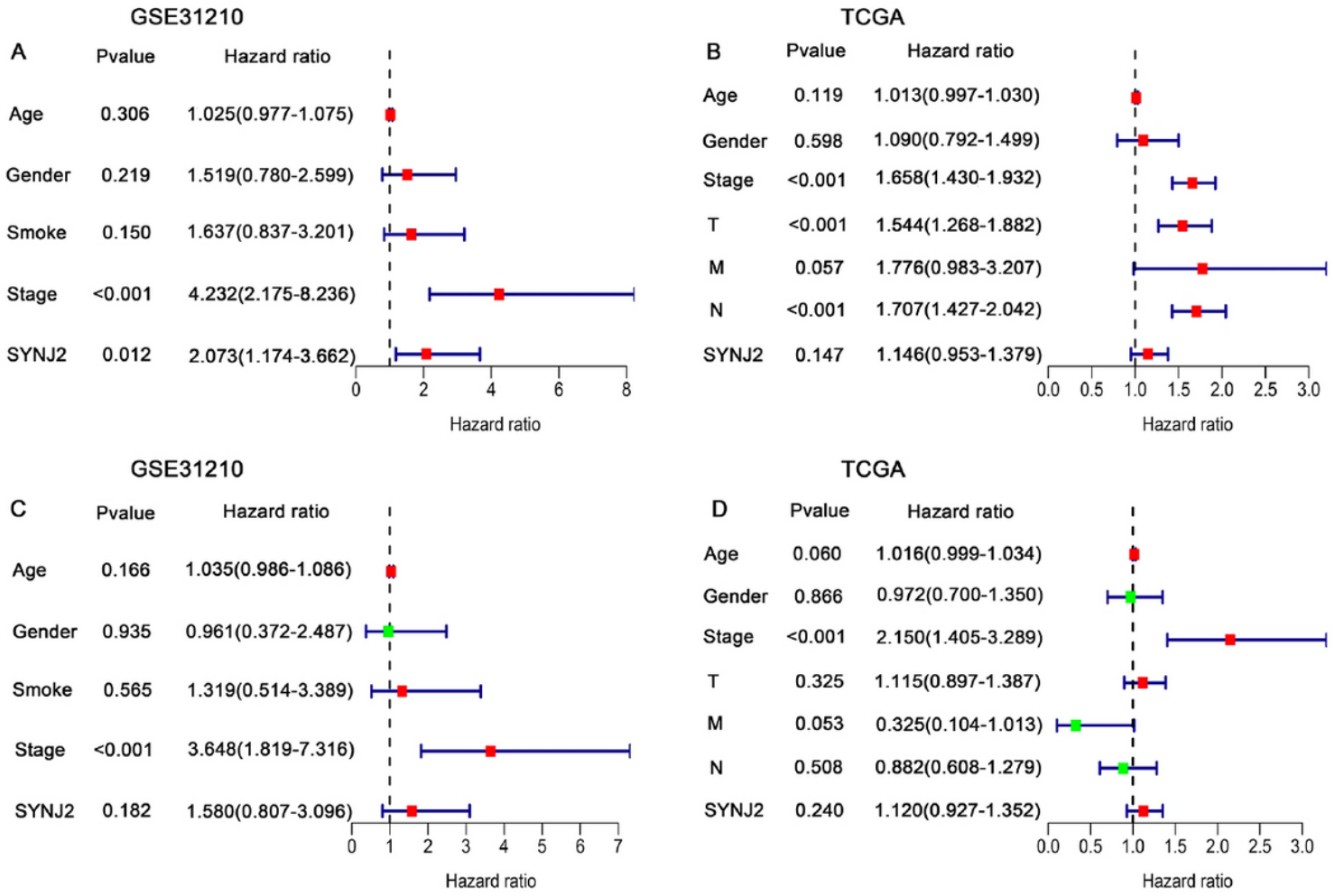
SYNJ2 mRNA expression levels and gene alteration. A. Histogram of SYNJ2 expression levels in LC cells, which were the median between those of several cancer cell lines. The histogram was downloaded from The Cancer Cell Line Encyclopedia. Different colors represent the expression level of SYNJ2 in different

cell types. B. The OncoPrint schematic revealed that the SYNJ2 gene became altered in 16/518 (3%) sequenced samples. C. Boxplot of SYNJ2 expression levels in primary non-small cell LC, circulating tumor cell (CTC), and white blood cells (WBC). The boxplot downloaded from gene expression database of CTCs. LUAD, Lung Adenocarcinoma.



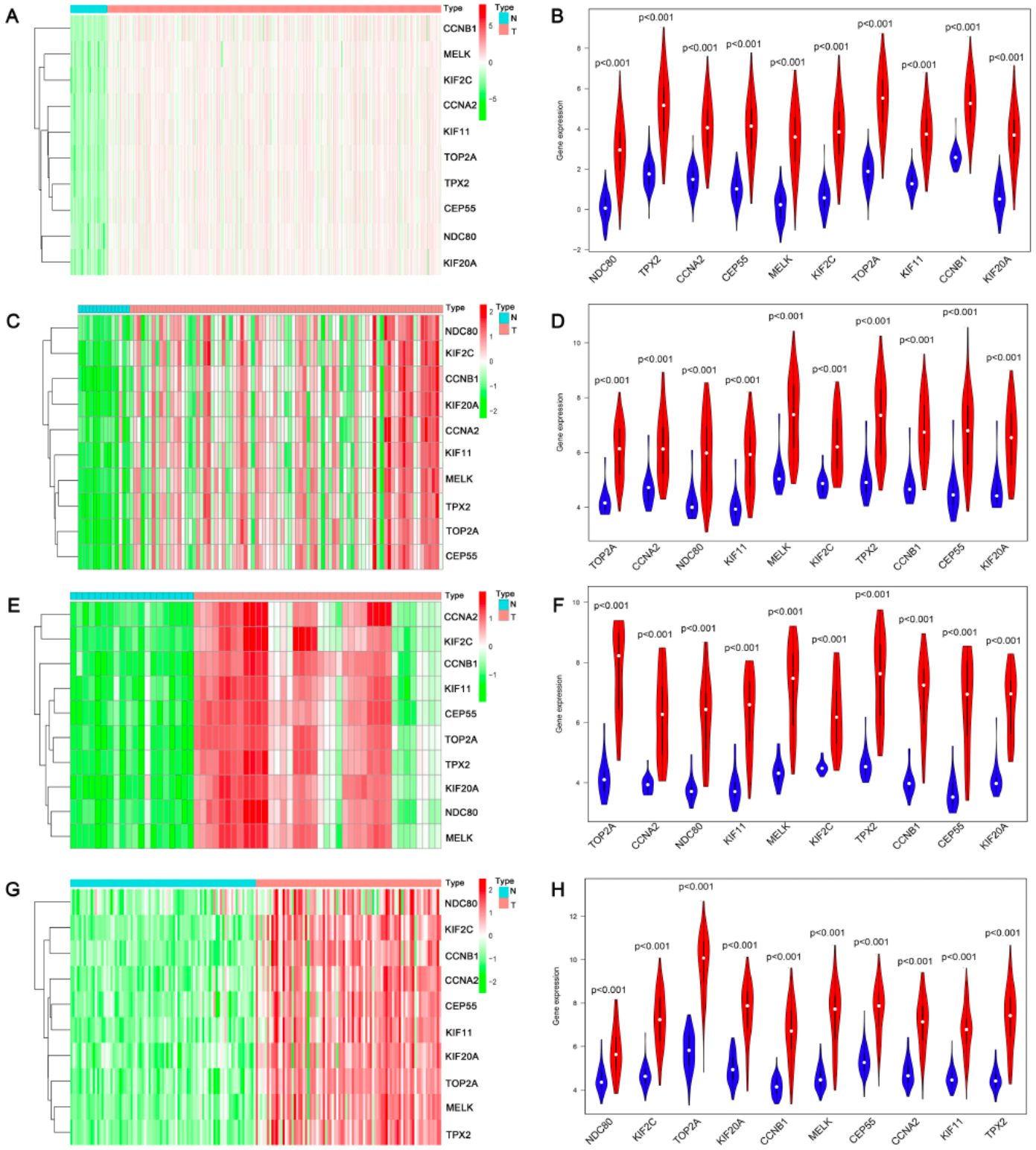

## Figure 5

Kaplan-Meier survival curves of SYNJ2 in different datasets. A-C. The Kaplan-Meier survival curve demonstrated no significance difference in OS between a high and a low expression of SYNJ2 based on the TCGA, GSE11969, and GSE87340 datasets. D. The patients with low expression of SYNJ2 in LUAD had a better prognosis than that with high expression according to the GSE31210 microarray.




## Figure 6

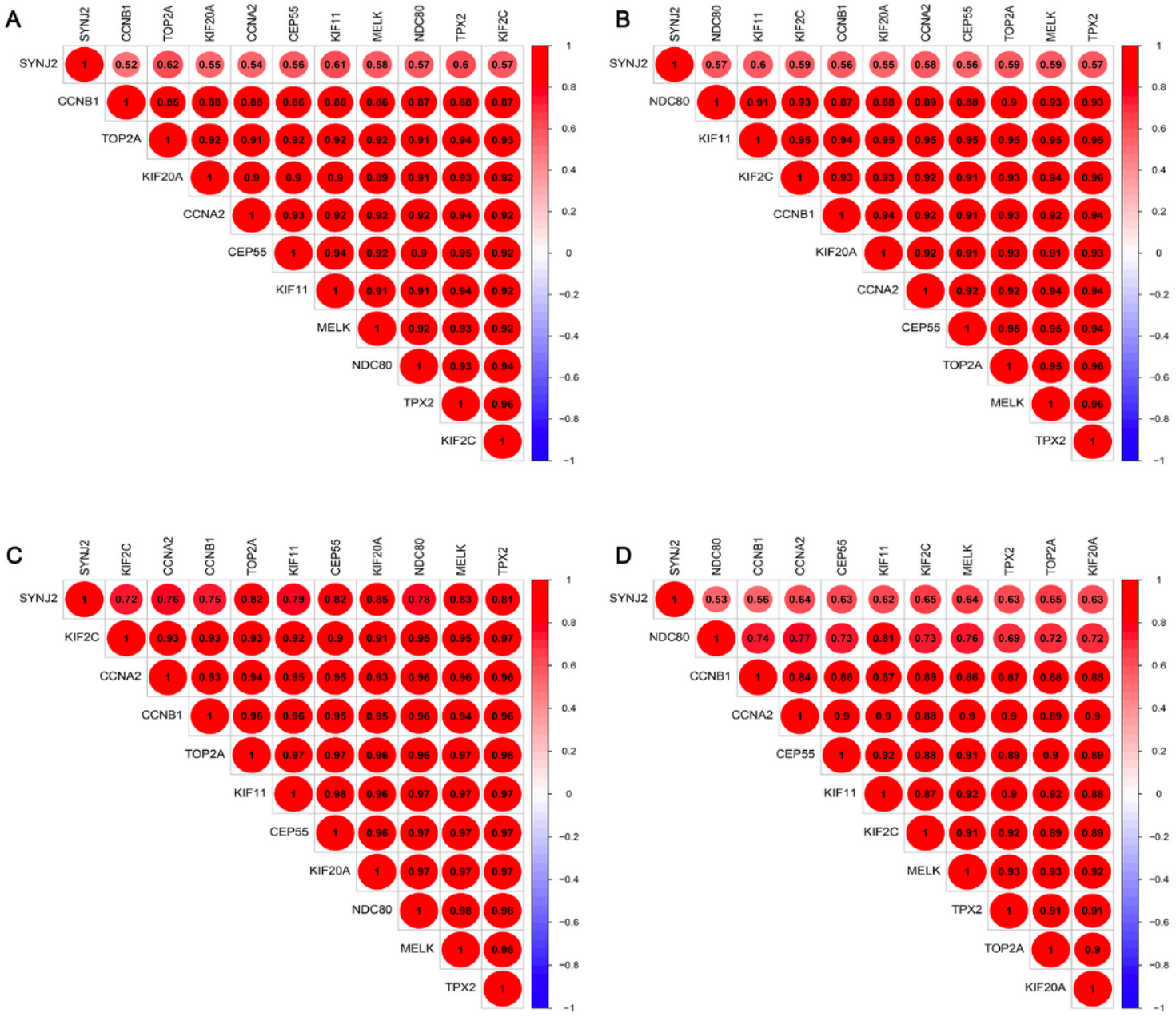
Univariate and multivariate cox regression analyses for the prognostic value of SYNJ2. A-B. Univariate cox regression analysis (A: GSE31210, B: TCGA). C-D. Multivariate cox regression analysis (C: GSE31210, D: TCGA).




## Figure 7

Expression of 10 SYNJ2 co-expressed hub genes in LUAD samples based on different datasets. A-B. Heatmap and violin diagram of the 10 SYNJ2 co-expressed hub genes expression level in LUAD and normal samples based on TCGA dataset. C-D. Heatmap and violin diagram of the 10 SYNJ2 co-expressed hub genes expression level in LUAD and normal samples based on GSE30219 dataset. E-F. Heatmap and violin diagram of the 10 SYNJ2 co-expressed hub genes expression level in LUAD and normal samples

based on GSE33532 dataset. G-H. Heatmap and violin diagram of the 10 SYNJ2 co-expressed hub genes expression level in LUAD and normal samples based on GSE70537 dataset. In heatmap, the depth of red represents the level of high expression, and the depth of green represents the level of low expression. In violin diagram, red represents LUAD, blue represents normal samples, and the positions of the white spots represent the median values of expression. N, normal; T, tumor.




## Figure 8

Correlation between SYNJ2 and its 10 co-expressed hub genes in different datasets. A. Spearman correlation analysis of SYNJ2 and its 10 co-expressed hub genes in the TCGA dataset. B. Spearman correlation analysis of SYNJ2 and its 10 co-expressed hub genes in GSE 30219 dataset. C. Spearman correlation analysis of SYNJ2 and its 10 co-expressed hub genes in GSE33532 dataset. D. Spearman

correlation analysis of SYNJ2 and its 10 co-expressed hub genes in GSE70537 dataset. Red: positive correlation; blue: negative correlation; all P<0.001.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFigure1.pdf
- SupplementaryFigure2.pdf
- SupplementaryFigure3.pdf
- SupplementaryFigure4.pdf
- SupplementaryFigure5.pdf
- SupplementaryFigure6.pdf
- SupplementaryFigure7.pdf
- Supplementarytable1.docx
- Supplementarytable2.docx