Integrated analysis reveals the prognostic significance of pyrimidine metabolic rate limiting enzymes in patients with lung adenocarcinoma

CURRENT STATUS: POSTED

Haiwei Wang
Fujian Maternity and Child Health Hospital

hwwang@sibs.ac.cn Corresponding Author
ORCiD: https://orcid.org/0000-0002-9675-4039

Xinrui Wang
Fujian Maternity and Child Health Hospital

Liangpu Xu
Fujian Maternity and Child Health Hospital

Ji Zhang
Shanghai Jiao Tong University

Hua Cao
Fujian Maternity and Child Health Hospital

DOI: 10.21203/rs.3.rs-16234/v1

SUBJECT AREAS
   Translational Medicine

KEYWORDS
   lung adenocarcinoma, pyrimidine metabolic rate limiting enzymes, GEO, TCGA
Abstract

Background: Reprogramming of metabolism is illustrated in many types of cancer and is associated with the clinical outcomes of cancer patients. However, the prognostic significance of pyrimidine metabolism signaling pathway in patients with lung adenocarcinoma (LUAD) is unclear.

Methods: Using the Gene Expression Omnibus (GEO) and The Cancer Genome Atlas (TCGA) datasets, we determined the prognostic significance of pyrimidine metabolic rate limiting enzymes in patients with lung cancer. The enrichment of pyrimidine metabolism signaling pathway was analyzed by gene set enrichment analysis (GSEA). The expression profiles of pyrimidine metabolic rate limiting enzymes were identified from Affymetrix gene expression data and RNA-seq data. The Kaplan-Meier Plotter was used to identify the prognostic significance of the pyrimidine metabolic rate limiting enzymes. Spearman’s correlation and multivariate cox regression were used to determine the correlation efficiency of pyrimidine metabolic rate limiting enzymes.

Results: The pyrimidine metabolism signaling pathway is significantly enriched in LUAD tissues, but not in lung squamous cell carcinoma (LUSC) tissues. Compared with normal lung tissues, the pyrimidine metabolic rate limiting enzymes are highly expressed in lung tumor tissues. Furthermore, the high expression levels of pyrimidine metabolic rate limiting enzymes are associated with unfavorable prognosis in patients with LUAD. Moreover, we demonstrate that the hypo-DNA methylation, DNA amplification and TP53 mutation are contributing to the high expression levels of pyrimidine metabolic rate limiting enzymes in cancer cells. Finally, we reveal the significant connections between the pyrimidine metabolic rate limiting enzymes. Comprehensively, the pyrimidine metabolic rate limiting enzymes are highly expressed in patients with bladder cancer, breast cancer, colon cancer, liver cancer and stomach cancer. And the high expression levels of pyrimidine metabolic rate limiting enzymes are associated with unfavorable prognosis in patients with liver cancer.

Conclusions: Pyrimidine metabolism signaling pathway is associated with the development of lung adenocarcinoma. Pyrimidine metabolic rate limiting enzymes CAD, DTYMK, RRM1, RRM2, TK1, TYMS, UCK2, NR5C2 and TK2 have significant prognostic effects in patients with lung adenocarcinoma.
Background

Lung cancer is one of most commonly diagnosed cancer and the leading cause of cancer related mortality [1-3]. Although some improvements of treatment of lung cancer have been achieved in the past few decades, the 5-year survival rate of patients with lung cancer is still low [4, 5]. Lung cancer is a heterogeneous disease, including small cell lung cancer and non-small cell lung cancer (NSCLC) [6]. NSCLC accounts for the 85% of lung cancer cases and could be further divided into 3 major pathologic subtypes: lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC) and large cell carcinoma [7]. Each subtype of NSCLC demonstrates different molecular profiles and different drug response [8, 9]. Although, gene alterations [10, 11], mRNA expression signature [12, 13], microRNA profiles [14, 15], long non-coding RNAs [16, 17], immune signature [18] and tumor microenvironment [19] are used for the prognosis of patients with NSCLC, more candidate biomarkers are needed to be identified.

Reprogramming of cell metabolism is a hallmark of cancer [20]. Cancer cells increase glucose uptake and utilize aerobic glycolysis to facilitate the uncontrolled cell proliferation [21]. Glycolysis-related gene signature is associated with the overall survival of lung adenocarcinoma patients [22]. Besides the mis-regulation of glucose metabolism, the normal pyrimidine metabolism is also disrupted during the development of cancer [23]. The disruption of the pyrimidine metabolism is reflected by the malfunctions of the pyrimidine metabolic rate limiting enzymes. The high expression levels of pyrimidine metabolic rate limiting enzymes CAD, CTPS, CTPS2, DHODH, DTYMK, NT5C2, NT5C3, RRM1, RRM2, TK1, TK2, TYMS, UCK2 and UCKL1 are illustrated in poorly differentiated liver cancer patients and correlated poor clinical outcomes [24]. Inhibition of pyrimidine synthesis by targeting pyrimidine metabolic rate limiting enzymes DHODH and CAD could accentuate the molecular therapy response in patients with glioblastoma [25]. Also inhibition of pyrimidine synthesis sensitizes triple negative breast cancer cells to chemotherapy agents [26]. However, the prognostic significance of the pyrimidine metabolism singling pathway in patients with lung adenocarcinoma is unclear. In order to reveal the prognostic significance of pyrimidine metabolic rate limiting enzymes in patients with lung adenocarcinoma, we used large cohorts of lung cancer patients derived from Gene
Expression Omnibus (GEO) datasets and The Cancer Genome Atlas (TCGA) datasets. Overall, the analysis of GEO and TCGA datasets in the present study allowed an improved understanding of the functions of the pyrimidine metabolic rate limiting enzymes. The results also indicated the potential biomarkers of the pyrimidine metabolic rate limiting enzymes for further clinical studies.

Methods

Data collection

The TCGA LUAD and LUSC gene expression, DNA mutation and DNA methylation, along with the clinical datasets were downloaded from the TCGA hub (https://tcga.xenahubs.net). The LUAD and LUSC gene expression data was generated from RNA-seq and the DNA methylation data was generated from Illumina HumanMethylation450 Bead Chip. Gene expression data derived from bladder urothelial carcinoma (BLCA), breast invasive carcinoma (BRCA), colon adenocarcinoma (COAD), esophageal carcinoma (ESCA), head and neck squamous cell carcinoma (HNSC), kidney renal clear cell carcinoma (KIRC), kidney renal papillary cell carcinoma (KIRP), liver hepatocellular carcinoma (LIHC), stomach adenocarcinoma (STAD) and thyroid cancer (THCA) patients were also downloaded from TCGA hub.

The gene expression series matrix of normal and cancerous lung tissues was downloaded from the GEO website (www.ncbi.nlm.nih.gov/geo) and included GSE7670, GSE10072, GSE18842, GSE19188, GSE27262, GSE30219, GSE31210, GSE31908, GSE33532 and GSE75324 datasets. The DNA methylation data of patients with lung adenocarcinoma was downloaded from the GEO datasets with GEO number GSE32867 and GSE62948. All the GEO expression datasets were based on Affymetrix Human Genome microarray. The detailed description of the collected data used in this study was illustrated in Fig. 1A.

GEO data processing

The GEO expression datasets were processed using R software (version 3.5.0, https://www.r-project.org/). The matrix file of each dataset was annotated with corresponding platform. When multiple probes corresponded to the same gene symbol, the expression values were averaged using ‘plyr’ package (version 1.8.5) in R software. Plyr package includes multiple tools for splitting, applying
and combining data and could be downloaded from bioconductor (https://cran.r-project.org/web/packages/plyr/index.html). The different gene expression between normal and lung cancer samples was determined using paired Student’s t test. The different DNA methylation intensity between normal and lung cancer samples was also determined using paired Student’s t test.

**Gene set enrichment analysis (GSEA)**

The metabolic singling pathways enriched in lung cancer gene expression profiling were determined using GSEA software (version 2.0) [27]. The GSEA software and the signaling pathways gene sets were downloaded from the GSEA web site (www.broad.mit.edu/gsea/index.html). Genes were ranked by the signal-to-noise ratio, and statistical significance was determined by 1,000 gene set permutations. The results of significance should meet the criteria of nominal P value less than 0.05.

**Heatmap presentation**

Heatmaps were created by R software ‘pheatmap’ package (version 1.0.12). The ‘pheatmap’ package and the basic usage were downloaded from bioconductor (https://cran.r-project.org/web/packages/pheatmap/). The clustering scale was determined by the ‘average’ method. The clustering distance was determined by the ‘correlation’ method. Other parameters were provided in the usage of the ‘pheatmap’.

**Survival analysis using GEO datasets**

The Kaplan-Meier Plotter (https://kmplot.com/analysis/) [28, 29] was used to identify the association between the expression levels of the pyrimidine metabolic rate limiting enzymes and overall survival in patients with lung cancer derived from GEO datasets. The Kaplan-Meier Plotter is an online survival analysis tool to rapidly assess the prognostic effects of genes using GEO microarray data. The patients were divided by the auto select best cutoff using the expression of the pyrimidine metabolic rate limiting enzymes. P values were determined using Log-rank test.

**Survival analysis using TCGA datasets**

R statistics software ‘survival’ package (version 3.1-8) was used to identify the clinical influence of pyrimidine metabolic rate limiting enzymes on overall survival in patients derived from TCGA LUAD, LUSC, LIHC, BRCA and STAD datasets. The ‘survival’ package and the basic usage were downloaded
from bioconductor (https://cran.r-project.org/web/packages/survival/index.html). The patients were divided into two clusters based on the mean expression levels of the pyrimidine metabolic rate limiting enzymes. Kaplan-Meier estimator was applied to determine the clinical outcomes in patients with high expression levels and low expression levels of pyrimidine metabolic rate limiting enzymes. P values were determined using Log-rank test.

**Oncoprints of the pyrimidine metabolic rate limiting enzymes**

The genomic alterations of pyrimidine metabolic rate limiting enzymes in LUAD patients were downloaded from cbioportal (version 3.2.0) based on the TCGA datasets (http://www.cbioportal.org/index.do).

**Correlation plots of the pyrimidine metabolic rate limiting enzymes**

Correlation plots of the pyrimidine metabolic rate limiting enzymes were created using the ‘corrplot’ package (version 0.84) in R. The ‘corrplot’ package and the basic usage were downloaded from bioconductor (cran.r-project.org/web/packages/corrplot/index.html). The Spearman’s correlation test was used to demonstrate the correlation efficiency.

**Multivariate cox regression**

Multivariate cox regression was analyzed by ‘coxph’ method in R software ‘survival’ package (version 3.1-8). The ‘survival’ package and the basic usage were downloaded from bioconductor (https://cran.r-project.org/web/packages/survival/index.html). Log-rank test was used to calculate the P values.

**Statistical analysis**

The box plots were generated from prims5.0. Statistical analysis was performed using the paired Student’s t test. P value less than 0.05 was chosen to be significantly different.

**Results**

Pyrimidine metabolism signaling pathway is highly enriched in lung tumor samples across different datasets.

In order to reveal the metabolism related transcriptional profiling of lung cancer, we analyzed lung
cancer patients with expression data from previously published GEO datasets. Totally, 1290 samples were collected from ten previously published datasets based on Affymetrix gene microarray platforms, including 336 normal lung samples and 954 lung tumor samples. Most of the lung cancer patients belonged to lung adenocarcinoma subtype. A detailed description of the collected data used in this study was illustrated in Fig. 1A.

We then identified the enriched metabolic signaling pathways in patients with lung cancer using the GSEA assay. Among all the enriched metabolic signaling pathways, pyrimidine metabolism signaling pathway was significantly enriched in seven out of ten datasets, including, GSE10072, GSE18842, GSE19188, GSE27262, GSE30219, GSE31210 and GSE75324 datasets, representing the most frequently enriched metabolic signaling pathway (Fig. 1B). Only in GSE7670, GSE31908 and GSE33532 three datasets, the pyrimidine metabolism signaling pathway was not significantly correlated with the transcriptional profiling of lung cancer (Fig. 1B).

Next, using the TCGA lung cancer dataset, we found that the pyrimidine metabolism signaling pathway was positively associated with the transcriptional profiling of lung cancer in LUAD dataset (Fig. 1B). However, in another subtype of lung cancer LUSC, pyrimidine metabolism signaling pathway was not highly enriched in the tumor patients (Fig. 1B).

Pyrimidine metabolic rate limiting enzymes are up-regulated in lung cancer cells across different datasets.

The pyrimidine metabolism signaling pathway was involving multiple genes. Previous results suggested that pyrimidine metabolism was highly controlled by pyrimidine metabolic rate limiting enzymes [24]. CAD, CTPS, CTPS2, DHODH, DTYMK, NT5C2, NT5C3, RRM1, RRM2, TK1, TK2, TYMS, UCK2 and UCKL1 were reported pyrimidine metabolic rate limiting enzymes [24]. The expression levels of those pyrimidine metabolic rate limiting enzymes in lung normal tissues and corresponding lung tumor tissues were investigated in GSE7670, GSE10072, GSE18842, GSE19188, GSE27262, GSE31908, GSE33532 and GSE75324 datasets. As illustrated in the heatmaps, pyrimidine metabolic rate limiting enzymes CAD, CTPS, CTPS2, DHODH, DTYMK, NT5C3, RRM1, RRM2, TK2, TYMS, UCK2 and UCKL1 was up-regulated in lung cancer tissues (Fig. 2). However, TK2 and NT5C2 were relatively
down-regulated in lung cancer tissues, compared with the normal lung tissues (Fig. 2).

Similar results were derived from TCGA LUAD dataset. Compared with the normal lung tissues, pyrimidine metabolic rate limiting enzymes CAD, CTPS, CTPS2, DTYMK, NT5C3, RRM1, RRM2, TK2, TYMS, UCK2 and UCKL1 was all highly expressed in lung cancer tissues in TCGA LUAD dataset. However, TK2 and NT5C2 were down regulated in lung cancer tissues (Fig. 2).

Expression levels of pyrimidine metabolic rate limiting enzymes are associated with the tumor overall survival in patients with lung cancer: analysis from GEO datasets.

Since the pyrimidine metabolic rate limiting enzymes were activated in lung tumor tissues, the present study next assessed the prognostic effects of pyrimidine metabolic rate limiting enzymes in lung cancer patients. The Kaplan-Meier Plotter is an online survival analysis tool to rapidly assess the prognostic effects of genes using the integrated GEO microarray data derived from 1926 lung cancer patients [28, 29]. Using Kaplan-Meier Plotter, the present study showed that, high expression levels of the pyrimidine metabolic rate limiting enzymes CAD, CTPS, DHODH, DTYMK, RRM1, RRM2, TK1, TYMS and UCK2 were unfavorable prognostic markers in patients with lung cancer (Fig. 3). Lung cancer patients with higher expression levels of CAD, CTPS, DHODH, DTYMK, RRM1, RRM2, TK1, TYMS and UCK2 had worse clinical outcomes than patients with low expression levels of those genes (Fig. 3).

However, consistent with the decreased expression levels of NR5C2 and TK2 in lung cancer tissues, patients with higher expression levels of NR5C2 and TK2 had better prognosis than patients with low expression levels of those genes (Fig. 3).

Expression levels of pyrimidine metabolic rate limiting enzymes are associated with the tumor overall survival in patients with lung adenocarcinoma: analysis from TCGA LUAD dataset.

Furthermore, using TCGA LUAD dataset, we confirmed the prognostic effects of pyrimidine metabolic rate limiting enzymes in patients with lung adenocarcinoma. Similarly, the Kaplan-Meier survival analysis revealed that the pyrimidine metabolic rate limiting enzymes DTYMK, NT5C3, RRM1, RRM2, TK1, TYMS and UCK2 were all associated with worse prognosis in the patients with lung cancer using TCGA LUAD dataset (Fig. 4). Patients with high expression levels of DTYMK, NT5C3, RRM1, RRM2, TK1, TYMS and UCK2 were with low overall survival. However, we found that the CAD, CTPS, DHODH,
NR5C2 and TK2 had no prognostic effect in LUAD dataset (Fig. 4).

The prognostic significance of pyrimidine metabolic rate limiting enzymes in patients with LUSC was also tested using TCGA LUSC dataset. However, unlike LUAD, pyrimidine metabolic rate limiting enzymes CAD, CTPS, DHODH, DTMYK, NT5C3, RRM1, RRM2, TK1, TYMS, UCK2 and NT5C2 had no prognostic effect in patients with LUSC (Fig. S1). Only high expression levels of TK2 were associated with the worse prognostic outcomes in LUSC dataset (Fig. S1).

Increased expression levels of the pyrimidine metabolic rate limiting enzymes in lung cancer cells are induced by DNA hypo-methylation.

Next, we tried to determine the mechanisms that induced the activation of the pyrimidine metabolic rate limiting enzymes in lung cancer. The high expression levels of oncogenes are usually mediated by hypo-DNA methylation, DNA amplification and gene mutation [30]. Using the DNA methylation data deposited in GSE32867 and GSE62948 datasets, we analyzed the DNA methylation intensity of the pyrimidine metabolic rate limiting enzymes in normal lung tissues and lung cancer tissues. The detailed description of the GSE32867 and GSE62948 datasets was illustrated in Fig. 1A. Compared with the lung normal tissues, the pyrimidine metabolic rate limiting enzymes CAD, RRM2 and TK1 were with hypo-DNA methylation patterns in lung cancer tissues derived from GSE32867 dataset (Fig. 5A). Similar results were obtained in GSE62948 dataset that the DNA methylation intensity of CAD, RRM2 and TK1 was lower in lung cancer tissues, compared with normal lung tissues (Fig. 5B). Also, in TCGA LUAD dataset, pyrimidine metabolic rate limiting enzymes RRM2, TK1, CAD, UCK2, TYMS and CTPS genes exhibited hypo-DNA methylation in LUAD tumor tissues (Fig. 5C). These observations suggested that DNA methylation was partially contributing to the activation of pyrimidine metabolic rate limiting enzymes in the tumor cells.

Increased expression levels of the pyrimidine metabolic rate limiting enzymes in lung cancer cells are induced by DNA amplification and TP53 mutation.

Another factor determining the activation of pyrimidine metabolic rate limiting enzymes in lung cancer cells was genomic aberration, particularly DNA amplification. It was revealed that 6% lung cancer patients were with UCK2 gene amplification and 5% lung cancer patients were with UCKL1
gene amplification (Fig. 6A). Also TK1 gene amplification was occurred in 2.2% lung cancer patients (Fig. 6A). However, other pyrimidine metabolic rate limiting enzymes were without DNA amplification in lung cancer tissues (Fig. 6A).

TP53 is a critical regulator of multiple metabolism signaling pathways in lung cancer cells [31–33]. Loss of TP53 functions induces uncontrolled pyrimidine synthesis [34]. The present study assessed whether TP53 regulated the expression levels of the pyrimidine metabolic rate limiting enzymes. It was revealed that, pyrimidine metabolic rate limiting enzymes CAD, CTPS, DTYMK, RRM1, RRM2, TYMS, UCK2 and TK1 were all highly expressed in patients with TP53 mutant lung cancer patients, compared with lung cancer patients with wild type TP53 in GSE72094 dataset (Fig. 6B). Interestingly, TK2 which was down regulated in lung tumor tissues was highly expressed in lung cancer patients with wild type TP53 (Fig. 6B).

Those results were further validated in the TCGA LUAD dataset. The expression levels of pyrimidine metabolic rate limiting enzymes CAD, CTPS, DTYMK, RRM1, RRM2, TYMS, UCK2 and TK1 were particularly higher in patients with TP53 mutant lung cancer patients (Fig. 6C). And the expression levels of TK2 were lower in TP53 mutant lung cancer patients (Fig. 6C). Overall, our results suggested that hypo-DNA methylation, DNA amplification and TP53 mutation were combined contributing to the high expression levels of pyrimidine metabolic rate limiting enzymes in lung cancer cells.

Correlation between pyrimidine metabolic rate limiting enzymes is identified in patients with lung cancer.

Next, we tried to determine the connections between those pyrimidine metabolic rate limiting enzymes. Spearman correlation demonstrated high correlation of those genes. Particularly, RRM2 was highly associated with the expression of TK1, RRM1, TYMS and DTYMK genes in GSE30219 lung cancer expression dataset (Fig. 7A). However, NT5C2 and TK2 were negatively correlated with other pyrimidine metabolic rate limiting enzymes (Fig. 7A). Similar results were obtained from TCGA LUAD expression dataset. RRM2 was positively correlated with other pyrimidine metabolic rate limiting enzymes, while, TK2 was negatively correlated with other pyrimidine metabolic rate limiting enzymes (Fig. 7A).
Furthermore, we used multivariate cox regression analysis to determine the connections between the pyrimidine metabolic rate limiting enzymes. It was revealed that RRM2 was an independent prognostic marker in patients with lung cancer in GSE30219 dataset (Fig. 7B). In LUAD dataset, all pyrimidine metabolic rate limiting enzymes were interconnected with each other and those genes were not independent prognostic markers (Fig. 7B).

Combined prognostic effects of pyrimidine metabolic rate limiting enzymes are identified in patients with lung cancer.

In both GSE30219 and TCGA LUAD datasets, the pyrimidine metabolic rate limiting enzymes were interconnected with each other and were not independent prognostic markers. So, we tested the combined prognostic effects of pyrimidine metabolic rate limiting enzymes in patients with lung cancer. Lung cancer patients were divided into two clusters based on the unsupervised clustering of the expression levels of pyrimidine metabolic rate limiting enzymes in GSE30219 dataset (Fig. 8A). The cluster1 lung patients were with lower expression levels of CAD, CTPS, RRM1, RRM2, DTYMK, TK1, TYMS and UCK2 (Fig. 8A). Correspondingly, lung cancer patients in cluster1 were with longer overall survival time, compared with lung cancer patients in cluster2 (Fig. 8B).

Similarly, the patients were divided into two clusters by the unsupervised clustering of the pyrimidine metabolic rate limiting enzymes in TCGA LUAD dataset (Fig. 8C). CAD, CTPS, RRM1, RRM2, DTYMK, TK1, TYMS and UCK2 were down regulated in cluster1 lung cancer patients (Fig. 8C). Correspondingly, lung cancer patients in cluster1 demonstrated better prognostic outcomes, compared with lung cancer patients in cluster2 in TCGA LUAD dataset (Fig. 8D).

Pyrimidine metabolic rate limiting enzymes are up-regulated in multiple types of tumor.

Using TCGA database, we comprehensively investigated the expression levels of pyrimidine metabolic rate limiting enzymes across different types of cancer. The expression levels of the pyrimidine metabolic rate limiting enzymes in normal tissues and corresponding tumor tissues were investigated in bladder urothelial carcinoma (BLCA), breast invasive carcinoma (BRCA), colon adenocarcinoma (COAD), esophageal carcinoma (ESCA), kidney renal papillary cell carcinoma (KIRP), liver hepatocellular carcinoma (LIHC), lung squamous cell carcinoma (LUSC), stomach adenocarcinoma
(STAD) and thyroid cancer (THCA) tissues (Fig. 9). As illustrated in the heatmaps, pyrimidine metabolic rate limiting enzymes CAD, CTPS, CTPS2, DHODH, DTYMK, NT5C3, RRM1, RRM2, TK2, TYMS, UCK2 and UCKL1 was highly expressed in tumor tissues, compared with normal tissues (Fig. 9). However, TK2 and NT5C2 were not significantly over-expressed in tumor tissues (Fig. 9). These results indicated the universal importance of pyrimidine metabolic rate limiting enzymes in the development of cancer.

However, using GSEA assay, we found that pyrimidine metabolism signaling pathway was only significantly enriched in BRCA and THCA tumors (Fig. S2). Although, pyrimidine metabolic rate limiting enzymes were over-expressed in BLCA, COAD, ESCA, LIHC and STAD tissues, the pyrimidine metabolism signaling pathway was not significantly enriched (Fig. S2).

The association between the expression levels of pyrimidine metabolic rate limiting enzymes and the tumor overall survival in patients with liver cancer, breast cancer or stomach cancer: analysis from BRCA, STAD and LIHC datasets.

Like LUAD, pyrimidine metabolic rate limiting enzymes were highly expressed in patients with BRCA, LIHC and STAD. However, in TCGA BRCA dataset, pyrimidine metabolic rate limiting enzymes CAD, CTPS, DHODH, DTYMK, NT5C3, RRM1, RRM2, TK1, TYMS, UCK2, NT5C2 and TK2 demonstrated no prognostic effect in patients with breast cancer (Fig. S3). Similarly, expression levels of pyrimidine metabolic rate limiting enzymes CAD, CTPS, DHODH, DTYMK, NT5C3, RRM1, RRM2, TYMS, UCK2, NT5C2 and TK2 had no clinical relevance in patients with stomach cancer in TCGA STAD dataset (Fig. S4). Only, TK1 was associated with better clinical outcomes in patients with stomach cancer (Fig. S4).

On the contrary, like LUAD patients, high expression levels of pyrimidine metabolic rate limiting enzymes CAD, DTYMK, NT5C3, RRM1, RRM2, TK1, TYMS and UCK2 were all associated with worse clinical outcomes in patients with liver cancer in TCGA LIHC dataset (Fig. 10). Moreover, consistent with the decreased expression levels of TK2 in liver cancer tissues, patients with higher expression levels of TK2 had better prognosis than patients with low expression levels of TK2 (Fig. 10). Those results highlighted the different prognostic effects of pyrimidine metabolic rate limiting enzymes in different tumor types.
Discussion
Metabolic reprogramming is a hallmark of cancer [20] and provides critical information for cancer classification and clinical prognosis [35]. Here, we identified the clinical relevance of pyrimidine metabolic rate limiting enzymes in patients with lung cancer based on their mRNA expression patterns using GEO and TCGA datasets. Ideally, we should use the metabolic activities to determine the prognostic effects of the pyrimidine metabolic rate limiting enzymes. However, previous results suggested that the metabolic reprogramming was caused by gene expression changes [36] and the expression profiles of metabolic pathway genes reflected the actual metabolic activities [37]. So, the expression patterns of metabolic genes had potential clinical implications for clinical prognosis.

Lung cancer is a heterogeneous disease, including many different subtypes with different genetic and epigenetic abnormality [6]. Because of the complexity of lung cancer [38], results derived from GEO and TCGA datasets were not always consistent with each other. Also, the difference in treatment protocol and microarray platform in individual study limited the further applications of these findings [39]. For example, CAD, CTPS and DHODH had no prognostic effect in LUAD dataset, but associated with the clinical outcomes in lung cancer patients derived from GEO datasets. To address this problem, we collected and studied multiple lung cancer GEO datasets and TCGA datasets. Our results suggested that the converged pyrimidine metabolism signaling pathway was generally altered in many gene expression datasets. And the pyrimidine metabolic rate limiting enzymes DTYMK, NT5C3, RRM1, RRM2, TK1, TYMS and UCK2 had particular values in lung cancer prognosis.

Some of the pyrimidine metabolic rate limiting enzymes, such as TK1 [40], UCK2 [41] and RRM1 [42] were reported to be associated with the poor outcomes of lung cancer in systematic review and meta-analysis. Functional studies identified pyrimidine metabolic rate limiting enzymes DHODH [43] and DTYMK [44] as therapeutic targets in lung cancer. Using the GEO and TCGA datasets, we confirmed the prognostic significance of pyrimidine metabolic rate limiting enzymes TK1, UCK2 and RRM1 in patients with lung adenocarcinoma. Furthermore, we found that pyrimidine metabolic rate limiting enzymes CAD, RRM2, DTYMK, TYMS, TK2 and NR5C2 were all associated with the clinical outcomes of lung cancer patients and liver cancer patients. Comprehensively, we showed that pyrimidine
metabolic rate limiting enzymes had no or little prognostic effect in patients with breast cancer, stomach cancer or lung squamous cell carcinoma, suggested that, compared with other tumor types, pyrimidine metabolic rate limiting enzymes were more important in the development of LUAD and LIHC.

The results of the present study indicated the biological functions of pyrimidine metabolism signaling pathway in the development of lung cancer and provided potential biomarkers for further clinical studies for patients with lung cancer. However, there were some limitations in this study. First, although, the expression profiles of metabolic pathway genes reflected the actual metabolic activities, the metabolic activities of pyrimidine metabolic rate limiting enzyme should be further tested. Second, clinical validations and functional studies were needed to reveal the inner mechanisms of how pyrimidine metabolic rate limiting enzymes correlated with the clinical outcomes of lung cancer patients. Our results suggested the different prognostic effects of pyrimidine metabolic rate limiting enzymes in patients with LUAD and LUSC. So, the clinical relevance the pyrimidine metabolic rate limiting enzymes in different subtypes of lung cancer patients should also be further illustrated. In our further studies, we will address those limitations and provided a more precise and reliable prognostic signature based on the metabolic activities of the pyrimidine metabolic rate limiting enzymes.

Conclusions
Pyrimidine metabolism signaling pathway is associated with the development of lung adenocarcinoma. Pyrimidine metabolic rate limiting enzymes TK1, UCK2, CAD, RRM1, RRM2, DTYMK, TYMS, TK2 and NR5C2 have significant prognostic effects in patients with lung adenocarcinoma.

Abbreviations
GEO: Gene Expression Omnibus; TCGA: The Cancer Genome Atlas; NSCLC: non–small cell lung cancer; LUAD: lung adenocarcinoma; LUSC: lung squamous cell carcinoma; BLCA: bladder urothelial carcinoma; BRCA: breast invasive carcinoma; COAD: colon adenocarcinoma; ESCA: esophageal carcinoma; KIRP: kidney renal papillary cell carcinoma; LIHC: liver hepatocellular carcinoma; STAD, stomach adenocarcinoma; THCA: thyroid cancer; NES: normalized enrichment score.

Declarations
Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Availability of data and materials
The datasets generated and/or analyzed during the current study are available in TCGA (tcga.xenahubs.net) and GEO (www.ncbi.nlm.nih.gov/geo) repositories.

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

Funding
The present study was supported by grants from the Fujian Provincial Maternity and the Children's Hospital (grant nos. YCXB 18-10 and YCXM 19-04).

Authors' contributions
HW designed the study and wrote the manuscript. HW, XW and LX performed the data analysis. HC and JZ designed the study and supervised the work.

Acknowledgements
Not applicable.

References
1. Siegel RL, Miller KD, Jemal A: Cancer statistics, 2018. CA Cancer J Clin 2018, 68:7-30.
2. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, 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.
3. Ferlay J, Colombet M, Soerjomataram I, Mathers C, Parkin DM, Pineros M, Znaor A, Bray F: Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. Int J Cancer 2019, 144:1941-1953.
4. Miller VA, Hirsh V, Cadranel J, Chen YM, Park K, Kim SW, Zhou C, Su WC, Wang M, Sun Y, et al: Afatinib versus placebo for patients with advanced, metastatic non-small-cell lung cancer after failure of erlotinib, gefitinib, or both, and one or two lines of chemotherapy (LUX-Lung 1): a phase 2b/3 randomised trial. *Lancet Oncol* 2012, 13:528-538.

5. Moll HP, Pranz K, Musteanu M, Grabner B, Hruschka N, Mohrherr J, Aigner P, Stiedl P, Brcic L, Laszlo V, et al: Afatinib restrains K-RAS-driven lung tumorigenesis. *Sci Transl Med* 2018, 10.

6. Herbst RS, Heymach JV, Lippman SM: Lung cancer. *N Engl J Med* 2008, 359:1367-1380.

7. Travis WD: Pathology of lung cancer. *Clin Chest Med* 2011, 32:669-692.

8. Cancer Genome Atlas Research N: Comprehensive molecular profiling of lung adenocarcinoma. *Nature* 2014, 511:543-550.

9. Cancer Genome Atlas Research N: Comprehensive genomic characterization of squamous cell lung cancers. *Nature* 2012, 489:519-525.

10. Wu K, Zhang X, Li F, Xiao D, Hou Y, Zhu S, Liu D, Ye X, Ye M, Yang J, et al: Frequent alterations in cytoskeleton remodelling genes in primary and metastatic lung adenocarcinomas. *Nat Commun* 2015, 6:10131.

11. Jiang J, Gu Y, Liu J, Wu R, Fu L, Zhao J, Guan Y: Coexistence of p16/CDKN2A homozygous deletions and activating EGFR mutations in lung adenocarcinoma patients signifies a poor response to EGFR-TKIs. *Lung Cancer* 2016, 102:101-107.

12. Wang Z, Wang Z, Niu X, Liu J, Wang Z, Chen L, Qin B: Identification of seven-gene signature for prediction of lung squamous cell carcinoma. *Onco Targets Ther* 2019, 12:5979-5988.
13. Li Y, Gu J, Xu F, Zhu Q, Ge D, Lu C: Transcriptomic and functional network features of lung squamous cell carcinoma through integrative analysis of GEO and TCGA data. *Sci Rep* 2018, 8:15834.

14. Li X, Shi Y, Yin Z, Xue X, Zhou B: An eight-miRNA signature as a potential biomarker for predicting survival in lung adenocarcinoma. *J Transl Med* 2014, 12:159.

15. Siriwardhana C, Khadka VS, Chen JJ, Deng Y: Development of a miRNA-seq based prognostic signature in lung adenocarcinoma. *BMC Cancer* 2019, 19:34.

16. Sui J, Yang S, Liu T, Wu W, Xu S, Yin L, Pu Y, Zhang X, Zhang Y, Shen B, Liang G: Molecular characterization of lung adenocarcinoma: A potential four-long noncoding RNA prognostic signature. *J Cell Biochem* 2019, 120:705-714.

17. Lin T, Fu Y, Zhang X, Gu J, Ma X, Miao R, Xiang X, Niu W, Qu K, Liu C, Wu Q: A seven-long noncoding RNA signature predicts overall survival for patients with early stage non-small cell lung cancer. *Aging (Albany NY)* 2018, 10:2356-2366.

18. Song Q, Shang J, Yang Z, Zhang L, Zhang C, Chen J, Wu X: Identification of an immune signature predicting prognosis risk of patients in lung adenocarcinoma. *J Transl Med* 2019, 17:70.

19. Yue C, Ma H, Zhou Y: Identification of prognostic gene signature associated with microenvironment of lung adenocarcinoma. *PeerJ* 2019, 7:e8128.

20. Hanahan D, Weinberg RA: Hallmarks of cancer: the next generation. *Cell* 2011, 144:646-674.

21. Shukla SK, Purohit V, Mehla K, Gunda V, Chaika NV, Vernucci E, King RJ, Abrego J, Goode GD, Dasgupta A, et al: MUC1 and HIF-1alpha Signaling Crosstalk Induces Anabolic Glucose Metabolism to Impart Gemcitabine Resistance to Pancreatic Cancer. *Cancer Cell* 2017, 32:71-87 e77.
22. Zhang L, Zhang Z, Yu Z: Identification of a novel glycolysis-related gene signature for predicting metastasis and survival in patients with lung adenocarcinoma. *J Transl Med* 2019, **17**:423.

23. Rabinovich S, Adler L, Yizhak K, Sarver A, Silberman A, Agron S, Stettner N, Sun Q, Brandis A, Helbling D, et al: Diversion of aspartate in ASS1-deficient tumours fosters de novo pyrimidine synthesis. *Nature* 2015, **527**:379-383.

24. Yeh HW, Lee SS, Chang CY, Hu CM, Jou YS: Pyrimidine metabolic rate limiting enzymes in poorly-differentiated hepatocellular carcinoma are signature genes of cancer stemness and associated with poor prognosis. *Oncotarget* 2017, **8**:77734-77751.

25. Wang X, Yang K, Wu Q, Kim LJY, Morton AR, Gimple RC, Prager BC, Shi Y, Zhou W, Bhargava S, et al: Targeting pyrimidine synthesis accentuates molecular therapy response in glioblastoma stem cells. *Sci Transl Med* 2019, **11**.

26. Brown KK, Spinelli JB, Asara JM, Toker A: Adaptive Reprogramming of De Novo Pyrimidine Synthesis Is a Metabolic Vulnerability in Triple-Negative Breast Cancer. *Cancer Discov* 2017, **7**:391-399.

27. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP: Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005, **102**:15545-15550.

28. Gyurffy B, Surowiak P, Budczies J, Lanczky A: Online survival analysis software to assess the prognostic value of biomarkers using transcriptomic data in non-small-cell lung cancer. *PLoS One* 2013, **8**:e82241.

29. Nagy A, Lanczky A, Menyhart O, Gyuroffy B: Validation of miRNA prognostic power in hepatocellular carcinoma using expression data of independent datasets.
30. Sun W, Bunn P, Jin C, Little P, Zhobotynsky V, Perou CM, Hayes DN, Chen M, Lin DY: **The association between copy number aberration, DNA methylation and gene expression in tumor samples.** *Nucleic Acids Res* 2018, **46**:3009-3018.

31. Vousden KH, Ryan KM: **p53 and metabolism.** *Nat Rev Cancer* 2009, **9**:691-700.

32. Gottlieb E, Vousden KH: **p53 regulation of metabolic pathways.** *Cold Spring Harb Perspect Biol* 2010, **2**:a001040.

33. Gottlieb E: **p53 guards the metabolic pathway less travelled.** *Nat Cell Biol* 2011, **13**:195-197.

34. Hubackova S, Davidova E, Boukalova S, Kvarova J, Bajzikova M, Coelho A, Terp MG, Ditzel HJ, Rohlena J, Neuzil J: **Replication and ribosomal stress induced by targeting pyrimidine synthesis and cellular checkpoints suppress p53-deficient tumors.** *Cell Death Dis* 2020, **11**:110.

35. Hu J, Locasale JW, Bielas JH, O'Sullivan J, Sheahan K, Cantley LC, Vander Heiden MG, Vitkup D: **Heterogeneity of tumor-induced gene expression changes in the human metabolic network.** *Nat Biotechnol* 2013, **31**:522-529.

36. Vander Heiden MG, DeBerardinis RJ: **Understanding the Intersections between Metabolism and Cancer Biology.** *Cell* 2017, **168**:657-669.

37. Peng X, Chen Z, Farshidfar F, Xu X, Lorenzi PL, Wang Y, Cheng F, Tan L, Mojumdar K, Du D, et al: **Molecular Characterization and Clinical Relevance of Metabolic Expression Subtypes in Human Cancers.** *Cell Rep* 2018, **23**:255-269 e254.

38. Takeuchi T, Tomida S, Yatabe Y, Kosaka T, Osada H, Yanagisawa K, Mitsudomi T, Takahashi T: **Expression profile-defined classification of lung adenocarcinoma shows close relationship with underlying major genetic changes and clinicopathologic behaviors.** *J Clin Oncol* 2006, **24**:1679-1688.
39. Wang S, Liu F, Wang Y, Fan W, Zhao H, Liu L, Cen C, Jiang X, Sun M, Han P: Integrated analysis of 34 microarray datasets reveals CBX3 as a diagnostic and prognostic biomarker in glioblastoma. J Transl Med 2019, 17:179.

40. Wei YT, Luo YZ, Feng ZQ, Huang QX, Mo AS, Mo SX: TK1 overexpression is associated with the poor outcomes of lung cancer patients: a systematic review and meta-analysis. Biomark Med 2018, 12:403-413.

41. Wu Y, Jamal M, Xie T, Sun J, Song T, Yin Q, Li J, Pan S, Zeng X, Xie S, Zhang Q: Uridine-cytidine kinase 2 (UCK2): A potential diagnostic and prognostic biomarker for lung cancer. Cancer Sci 2019, 110:2734-2747.

42. Zhu CM, Lian XY, Bi YH, Hu CC, Liang YW, Li QS: Prognostic value of ribonucleotide reductase subunit M1 (RRM1) in non-small cell lung cancer: A meta-analysis. Clin Chim Acta 2018, 485:67-73.

43. Li L, Ng SR, Colon CI, Drapkin BJ, Hsu PP, Li Z, Nabel CS, Lewis CA, Romero R, Mercer KL, et al: Identification of DHODH as a therapeutic target in small cell lung cancer. Sci Transl Med 2019, 11.

44. Liu Y, Marks K, Cowley GS, Carretero J, Liu Q, Nieland TJ, Xu C, Cohoon TJ, Gao P, Zhang Y, et al: Metabolic and functional genomic studies identify deoxythymidylate kinase as a target in LKB1-mutant lung cancer. Cancer Discov 2013, 3:870-879.

Figures
Pyrimidine metabolism signaling pathway is highly enriched in patients with lung cancer. (A)

Table showed the detailed GEO datasets and TCGA datasets used in this study. (B)

Enrichment plots demonstrated the enriched pyrimidine metabolism signaling pathway in GSE7670, GSE10072, GSE18842, GSE19188, GSE27262, GSE30219, GSE31210, GSE31908, GSE33532, GSE75324 and TCGA LUAD, LUSC datasets. Enrichment of normalized enrichment score (NES) and P-values were presented. LUAD: lung adenocarcinoma; LUSC: lung squamous cell carcinoma;
Expression profiling of pyrimidine metabolic rate limiting enzymes in lung normal and tumor tissues
Figure 2

Pyrimidine metabolic rate limiting enzymes are up-regulated in lung cancer cells across different datasets. Heatmaps demonstrated the expression levels of pyrimidine metabolic rate limiting enzymes in normal lung tissues and lung cancer tissues in GSE7670, GSE10072, GSE18842, GSE19188, GSE27262, GSE31908, GSE33532, GSE75324 and TCGA LUAD datasets. Upregulated (red), downregulated (green) and unchanged (black) genes were delineated.
Expression levels of pyrimidine metabolic rate limiting enzymes are associated with the tumor overall survival in patients with lung cancer: analysis from GEO datasets. The Kaplan-Meier Plotters demonstrated the prognostic effects of the expression levels of pyrimidine metabolic rate limiting enzymes TK1, UCK2, CAD, RRM1, RRM2, DTYMK, TYMS, TK2 and NR5C2 in patients with lung cancer using the integrated GEO datasets. The patients were divided by the auto select best cutoff based on the expression of the pyrimidine metabolic rate limiting enzymes. The log-rank test was used to determine the overall survival P-value.
Expression levels of pyrimidine metabolic rate limiting enzymes are associated with the tumor overall survival in patients with lung adenocarcinoma: analysis from TCGA LUAD dataset. The Kaplan-Meier Plotters demonstrated the associations between pyrimidine metabolic rate limiting enzymes and overall survival in patients with lung cancer using the TCGA LUAD dataset. The log-rank test was used to determine the overall survival P-value.
Increased expression levels of the pyrimidine metabolic rate limiting enzymes in lung cancer cells are induced by DNA hypo-methylation. (A) Box plots demonstrated the DNA methylation intensity (β value) of pyrimidine metabolic rate limiting enzymes CAD, RRM2 and TK1 in normal lung tissues and lung adenocarcinoma tissues in GSE32867 dataset. (B) Box plots demonstrated the DNA methylation intensity of CAD, RRM2 and TK1 genes in normal lung tissues and lung adenocarcinoma tissues in GSE62948 dataset. (C) Heatmaps demonstrated the methylation level (β value) of the pyrimidine metabolic rate limiting enzymes in normal and tumor tissues in patients with LUAD. Hypermethylated (red), hypomethylated (green) and unchanged (black) genes were delineated.
Increased expression levels of the pyrimidine metabolic rate limiting enzymes in lung cancer cells are induced by DNA amplification and TP53 mutation. (A) Oncoprints demonstrated the alteration frequency of pyrimidine metabolic rate limiting genes in patients with LUAD. Each line represented one patient. (B) Box plots demonstrated the expression levels of the pyrimidine metabolic rate limiting enzymes in patients with lung cancer. P-values indicated the differences between patients with TP53 mutant and TP53 wild-type. (C) Heatmap demonstrated the expression levels of the pyrimidine metabolic rate limiting enzymes in TP53 mutant and TP53 wild-type LUAD patients.
Figure 7

Correlation between pyrimidine metabolic rate limiting enzymes is identified in patients with lung cancer. (A) Corrplots demonstrated the correlation between pyrimidine metabolic rate limiting enzymes in the GSE30219 and TCGA LUAD dataset. The color and the size of the circle represented the correlation coefficients. (B) Multivariate cox regression was used to test the relationships of pyrimidine metabolic rate limiting enzymes expressions and overall survival in lung cancer patients in the GSE30219 and TCGA LUAD dataset.
Combined prognostic effects of pyrimidine metabolic rate limiting enzymes are identified in patients with lung cancer. (A) Unsupervised clustering heatmap showing the division of two clusters of lung cancer patients by the expression levels of pyrimidine metabolic rate limiting enzymes in GSE30219 dataset. Each line represented one patient. (B) The Kaplan-Meier Plotter demonstrated the different clinical outcomes of those two clusters of lung cancer patients in GSE30219 dataset. The log-rank test was used to determine the overall survival P-value. (C) Unsupervised clustering heatmap showing the division of two clusters of lung cancer patients by the expression levels of pyrimidine metabolic rate limiting enzymes in TCGA LUAD dataset. (D) The Kaplan-Meier Plotter demonstrated the different clinical outcomes of those two clusters of lung cancer patients in TCGA LUAD dataset. The log-rank test was used to determine the overall survival P-value.
Pyrimidine metabolic rate limiting enzymes are up-regulated in multiple types of tumor. Heatmaps demonstrated the expression levels (log2 count) of pyrimidine metabolic rate limiting enzymes in normal and tumor samples in BLCA, BRCA, COAD, ESCA, KIRP, LIHC, LUSC, STAD and THCA. Upregulated (red), downregulated (green) and unchanged (black) genes were delineated. BLCA, bladder urothelial carcinoma; BRCA, breast invasive carcinoma; COAD, colon adenocarcinoma; ESCA, esophageal carcinoma; KIRP, kidney renal papillary cell carcinoma; LIHC, liver hepatocellular carcinoma; LUSC, lung squamous cell carcinoma; STAD, stomach adenocarcinoma; THCA, thyroid cancer.
Expression levels of pyrimidine metabolic rate limiting enzymes are associated with the overall survival in patients with liver cancer: analysis from TCGA LIHC dataset. The Kaplan-Meier Plotters demonstrated the associations between pyrimidine metabolic rate limiting enzymes and overall survival in patients with liver cancer using the TCGA LIHC dataset. The log-rank test was used to determine the overall survival P-value. LIHC: liver hepatocellular carcinoma;

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

Fig. S4.tif
Fig. S3.tif
Fig. S2.tif
Fig. S1.tif