Identification of RRM2 as Novel Potential Sero-Diagnostic Marker for Non-Small Cell Lung Cancer

Xiuming Zhai

The Third Affiliated Hospital of Chongqing Medical University (Gener Hospital)

Dandan Zhou (✉ zhoudandancq@163.com)

Chongqing Key Laboratory of Sichuan-Chongqing Co-construction for Diagnosis and Treatment of Infectious Diseases Integrated Traditional Chinese and Western Medicine

Zihe Dong

Chongqing Hospital of Traditional Chinese Medicine

Ruixue Zhang

Chongqing Medical University

Research Article

Keywords:

Posted Date: March 2nd, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1366273/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background. Non-small cell lung cancer (NSCLC) is one of the cardinal trigger of cancer-related death worldwide. Most cases are diagnosed at an advanced stage using current tumor markers. Here, our study intended to identify the potential biomarkers to NSCLC diagnosis.

Material/Methods. Four independent datasets (GSE18842, GSE19188, GSE30219 and GSE40791) were downloaded from the Gene Expression Omnibus. In order to explore the hub genes, various bioinformatics tools including the DAVID, STRING, GEPIA database, and R, Cytoscape software were used to analyze the dataset. Serum levels of RRM2 were verified by ELISA. The levels of CEA, CYFRA21-1, NSE and ProGRP were analyzed by electrochemiluminescence immunoassay.

Results: A total of 258 genes were co-expression, of which 91 were up-regulated and 167 were down-regulated. A PPI network was constructed and 10 hub genes were screened, including RRM2, CDK1, UBE2C, MAD2L1, BUB1B, CCNA2, KIF20A, BUB1, KIF11, and CCNB2. The results of ELISA analysis showed that RRM2 is not only significantly up-regulated in NSCLC patient compared with the healthy control, but also the serum RRM2 level of NSCLC patient is related to distant metastasis, TNM stage and histological type ($P < 0.05$), but not to tumor size, lymph node metastasis ($P > 0.05$). ROC curve analysis showed that the combined detection of RRM2, CEA and NSE has the highest efficacy in diagnosing NSCLC.

Conclusions. The comprehensive analysis showed that RRM2 could be used as a potential biomarker for diagnosis of NSCLC.

Introduction

Lung cancer is highly invasive and metastatic. It has become one of the leading cause of malignant tumor-related death in the world [1]. Lung cancer is broadly divided into non-small cell cancer (NSCLC, approx. 85% cases) and small cell lung cancer (SCLC, approx. 15% cases) [2]. Despite the current efficacy of NSCLC has been extremely improved, the 5-year survival rate of NSCLC patients is still less than 18%, part of the reason is that most cases are diagnosed at an advanced stage[3]. Consequently, it is critical to identify new biomarkers that can diagnose NSCLC early so that patients can receive the best treatment as soon as possible. With the rapid development of high-throughput technology, tumor genome sequences can be re-sequenced to analyze tumor genome copy number, gene correlation and polymorphism [4, 5]. It provides the possibility for bioinformatics to mine cancer-related gene expression profiles. This study aims to explore the potential biomarkers in NSCLC, which may provide new ways for patients' diagnosis, treatment and prognosis.

Materials And Methods

Data Resources.

Series matrix files of GSE18842 [6], GSE19188 [7], GSE30219 [8] and GSE40791 [9] were downloaded from Gene Expression Omnibus (GEO) database (<http://www.ncbi.nlm.nih.gov/geo/>) [10]. The platforms were all based on GPL570 (Affymetrix Human Genome U133 Plus 2.0 Array). A total of 224 normal samples and 380 NSCLC samples were included in our study, of which 45 normal samples and 46 tumor samples were in GSE18842, 65 normal samples and 91 tumor samples in GSE19188, 14 normal samples and 149 tumor samples in GSE30219, and 100 normal samples and 94 tumor samples in GSE40791. All of the data were available online.

110 NSCLC patients who were treated at Chongqing Hospital of Traditional Chinese Medicine from October 2020 to September 2021 were selected as the NSCLC group. The TNM staging for non-small cell lung cancer is based on the 8th edition AJCC TNM staging system [11]. Exclusion criteria: (1) Patients with incomplete medical records; (2) Patients without clear histopathological or cytological diagnosis; (3) Patients combined with tumors of other systems; (4) Patients with seriously cerebrovascular and cardiovascular diseases, diabetes, and autoimmune diseases; (5) Patients who have received radiotherapy, chemotherapy or immunotherapy. In addition, 50 patients who underwent a physical examination in our hospital during the same period were selected as the healthy control group. There was no statistically significant difference in basic characteristics (gender, age) between NSCLC patients and healthy controls (**Supplemental file 1**).

Screening for DEGs.

Bioinformatics analysis was performed by R software (v3.6.2; <http://www.r-project.org>). The GEOquery, limma and dplyr packages were used for data normalization, the RobustRankAggreg package was used to screen the differentially expressed genes, using adjust *P* value < 0.05 and |log₂ fold change (FC)| ≥ 2 as cut-off criteria. The VennDiagram R package was used to present significant co-expression genes.

Gene Ontology and KEGG Pathway Enrichment Analysis of DEGs.

Gene ontology (GO) [12], a tool for annotating genes from various ontologies by using standard expression terms, currently include three aspects of biological content: cellular components (CC), biological processes (BP) and molecular functions (MF). The Kyoto Encyclopedia of Genes and Genomes (KEGG) [13] is a kind of bioinformatics resource for "understanding the advanced functions and utility resource library of biological systems", KEGG pathway mainly presents intermolecular interactions and intermolecular networks.

The DAVID database (v6.8; <http://david.abcc.ncifcrf.gov>) [14], an online tool, was utilized to perform GO enrichment and KEGG pathway analysis, using "after FDR" (corrected *P*-Value < 0.05, gene count ≥ 10) set as statistically significant. The visualized result of GO functional enrichment was presented by ggplot2 package in R.

PPI Network and Hub Gene Analysis.

The protein-protein interaction (PPI) network of DEGs was obtained via online analysis website STRING (v11.0; <https://string-db.org/>) [15]. Subsequently, Visual analysis of the PPI network graphs were adjusted in Cytoscape (v3.7.2; <https://cytoscape.org>), and hub genes were analyzed with the Cytoscape plugin CytoHubba [16]. The MCC algorithm was used to identify the top 10 hub genes. Hub genes were further performed KEGG pathway analysis using the DAVID database.

Survival Analysis and Validation of Hub Genes.

The Gene Expression Profiling Interactive Analysis (GEPIA) is a web-based tool, which integrates the TCGA and GTEx data for cancer and normal gene expression profiling and interactive analyses, including differential expression analysis, patient survival analysis, profiling plotting, correlation analysis, similar gene detection and dimensionality reduction analysis[17]. To further validate the 10 hub genes, the expression levels were analyzed and the Kaplan-Meier survival curves were mapped using the GEPIA (<http://gepia.cancer-pku.cn/index.html>).

Laboratory analysis.

Approximately 5 ml of peripheral blood samples of test subjects were collected in sterile tubes without anticoagulant prior to any treatment. Then all samples were centrifuged at 4000 r/min for 10 min at the room temperature. Finally, the resultant serum were stored at -20°C until measurement. The level of RRM2 in serum was measured with enzyme linked immunosorbent assay (ELISA) on TECAN Freedom EVOlyzer-2 150 platform (Tecan, Männedorf, Switzerland). RRM2 ELISA kit was purchased from Shanghai Tongwei Biotechnology Co., Ltd (Shanghai, China). According to the manufacturer's instructions, the absorbance of RRM2 was measured at 450 nm, and RRM2 concentrations were calculated using an appropriate calibration curve. RRM2 results were shown as ng/mL. Serum CEA, CYFRA21-1, ProGRP and NSE were analyzed using electrochemiluminescence immunoassay (ECLIA) on Cobas 6000 e601 (Roche Diagnostics, Mannheim, Germany) with the same batch of reagents. The normal reference interval of CEA is less than 5 ng/mL, the normal reference interval of CYFRA21-1 is less than 3.3 ng/mL, the normal reference interval of NSE is less than 16.3 ng/mL, the normal reference interval for serum ProGRP ranged from 25.3 to 69.2 pg/ml.

Statistical analysis.

The statistical analyses of this study were performed using SPSS software version 21.0 (SPSS Inc., Chicago, IL, USA). The Chi-square test was used to analyze the count data. Measurement data must first be analyzed for normal distribution. Measurement data that conforms to the normal distribution are expressed as the mean $\pm$ standard deviation (SD), the comparison between the two groups were performed by unpaired *t*-test. Measurement data that does not conform to the normal distribution are expressed as the median means (Quartile) [M (*P*25, *P*75)], the Mann-Whitney U-test was used for comparison between the two groups. The correlation between indicators were analyzed using Spearman correlation analysis. Logistic regression analysis and receiver operating characteristic (ROC) curves were

performed to assess the diagnostic capability of biomarkers in NSCLC patients. All statistical tests were considered statistically significant at p -value < 0.05.

Results

DEGs in NSCLC and normal tissues

In this study, four series of matrix files (GSE18842, GSE19188, GSE30219, GSE40791), for a total of 380 NSCLC samples and 224 normal samples, were used to identify DEGs ($P < 0.05$, $|\log FC| \geq 2$). The Venn diagrams showed that 258 genes were co-expression, in which 91 genes were clearly up-regulated and 167 genes were clearly down-regulated (**Fig. 1**). The more detailed results are shown in **Supplemental file 2**.

GO and KEGG pathway enrichment analysis

To further investigate the biological function of the 258 common DEGs. Next, GO enrichment and KEGG pathway analysis were performed via using DAVID database. As shown in **Fig. 2**, GO analysis showed that DEGs were particularly enriched in 35 terms, of which 18 cell component (CC), 9 biological processes (BP) and 8 molecular function (MF). KEGG pathway analysis showed that the enriched signaling pathway of DEGs was only related to the Cell cycle ($P = 5.65067 \times 10^{-5}$, gene count = 10). The more detailed results are shown in **Supplemental file 3**.

PPI Network and Module Analysis

In order to further identify which of 258 common DEGs are most likely to be central regulatory genes for NSCLC. The Protein-Protein Interaction (PPI) network was constructed by using the STRING database and Cytoscape software. The gene that is down-regulated was presented by arrows, and the gene that is up-regulated was presented by circles. Different colors indicate that the gene is involved in different biological content and pathways (**Fig. 3**). Subsequently, according to the MCC algorithm, the top 10 hub genes were screened through the cytoHubba, including RRM2, CDK1, UBE2C, MAD2L1, BUB1B, CCNA2, KIF20A, BUB1, KIF11, CCNB2. And we found that 10 hub genes were all significantly up-regulated. The more detailed results are shown in **Supplemental file 4**.

Validation and Survival Analysis of the Hub Genes

The expression levels of 10 hub genes in NSCLC and normal tissues were further verified by GEPIA database (483 LUAD vs 347 Normal, 486 LUSC vs 338 Normal). It was clearly shown that the expression on box-plots of 10 hub genes were all increased in both LUSC and LUAD, among which RRM2, UBE2C, MAD2L1, KIF20A, CDK1, CCNB2 were statistically significant ($P < 0.05$), KIF11, CCNA2, BUB1B, and BUB1 were only statistically significant in LUSC ($P < 0.05$) but not in LUAD ($P > 0.05$). In addition, the overall survival analysis of 10 hub genes were also performed through the GEPIA database. Kaplan-Meier Plotter showed that the survival rates of NSCLC patients with high expression of RRM2 or 2KIF11 were

statistically significant ($P < 0.05$), while UBE2C, MAD2L1, KIF20A, CDK1, CCNB2, CCNA2, BUB1B and BUB1 were not (**Fig. 4**). The expression levels and the survival analysis of other hub genes were shown in the **Supplemental file 5**. Based on the above analysis, we harbor the idea that RRM2 may play a crucial role in the tumorigenesis and survival prognosis of NSCLC.

ELISA Validation

The protein encoded by the RRM2 gene is a secreted protein which can be detected in serum. The serum level of RRM2 in NSCLC and healthy control were further verified by ELISA test. At the same time, serum levels of CEA, CYFRA21-1, NSE and ProGRP were analyzed by ECLIA. As expected, RRM2 is significantly up-regulated in NSCLC patient compared with the healthy control (**Table 1**). Subsequently, Spearman correlation analysis was further used to perform correlation analysis between indicators. We found that there was a weak positive correlation between RRM2 and CEA, CYFRA21-1, NSE ($P < 0.05$), but no correlation between RRM2 and PROGRP ($P > 0.05$) (**Supplemental file 6**).

In addition, we further analyzed the association between serum RRM2 level and clinical characteristics of the patient (**Table 2**). It was clearly shown that the serum RRM2 level of NSCLC patients were related to distant metastasis, TNM stage and histological type ($P < 0.05$), but not to tumor size, lymph node metastasis ($P > 0.05$).

Finally, receiver operating characteristic (ROC) curve analysis was used to calculate the sensitivity and specificity of these five biomarkers in NSCLC patients. The ROC analysis in **Table 3** showed that the ability to distinguish NSCLC patients from healthy controls using RRM2 was 0.798 (95% CI 0.731-0.866) and the cut-off value was 131.31 ng/mL. The sensitivity and specificity were 73.6% and 74.0%, respectively. Although the AUC value of CEA is higher than that of RRM2, when the cut-off value of CEA is 5.00ng/mL, the accuracy of CEA for NSCLC diagnosis is only 65.0%, while the accuracy of RRM2 and NSE are both 73.8%. Then, logistic regression analysis was used to combine these three biomarkers for further analysis (**Supplemental file 7**). The AUC value of the combined score (AUC = 0.958, Sensitivity = 90.9%, Specificity = 94.0%) indicated that the combined detection of RRM2, CEA and NSE has the highest efficacy in diagnosing NSCLC (**Table 3, Fig. 5**).

Discussion

In the current study, we first performed bioinformatics analysis of series matrix files of GSE18842, GSE19188, GSE30219 and GSE40791 using GEO, GEPIA, STRING databases and R, Cytoscape software. As a result, 258 genes were coexpression, of which 91 were up-regulated and 167 were down-regulated. 10 hub genes were further screened, including RRM2, CDK1, UBE2C, MAD2L1, BUB1B, CCNA2, KIF20A, BUB1, KIF11, CCNB2. Based on expression and survival prognosis analysis, we harbor the view that RRM2 may play a crucial role in the tumorigenesis and survival prognosis of NSCLC. Next, we further validated serum RRM2 level in NSCLC and healthy control by ELISA assay and the serum levels of CEA, CYFRA21-1, NSE and ProGRP were analyzed by ECLIA. We found that RRM2 is not only significantly up-regulated in NSCLC patient compared with the healthy control, but also the serum RRM2 level of NSCLC patient is related to

distant metastasis, TNM stage and histological type ($P < 0.05$), but not to tumor size, lymph node metastasis ($P > 0.05$). Finally, ROC curve analysis showed that the combined detection of RRM2, CEA and NSE has the highest efficacy in diagnosing NSCLC. Based on the above analysis, RRM2 may become a new potential serum diagnostic marker for non-small cell lung cancer.

RRM2 (ribonucleotide reductase regulatory subunit M2) encodes one of two different subunits of ribonucleotide reductase, this reductase catalyzes the conversion of ribonucleotides to deoxyribonucleotides [18]. Multiple studies have demonstrated that RRM2 is associated with non-small cell lung cancer. Mah V et al. research manifested that RRM2 is a useful predictor of survival in certain subgroups of NSCLC patients [19]. Grossi et al. showed that the expression of RRM2 is associated with poor prognosis in patients with resected stage I-III NSCLC [20]. Our study found that RRM2 is elevated in patients with non-small cell lung cancer. As well as, survival analysis showed that the survival rates of patients with high expression of RRM2 are significantly lower than that of patients with low expression. However, it is worth noting that the expression of RRM2 in the serum of patients with adenocarcinoma is higher than that of patients with squamous cell carcinoma, which is inconsistent with the gene expression data in the GEPIA database. Therefore, the mechanism of inconsistency between protein and gene expression is still to be further studied.

In addition, KEGG pathway analysis of hub genes revealed that RRM2 is primarily involved in the p53 signaling pathway (**Supplemental file 8**). We all know that p53 tumor suppressor, like Bcl-2, is one of the major apoptotic signaling pathways. Rahman MA et al. observed that RRM2 regulates the stabilization of Bcl-2 protein, with RRM2 inhibition leading to increased Bcl-2 degradation, and found their co-localization in NSCLC cells [21]. Combined with the above analysis, we hypothesize that RRM2 may be involved in the regulation of apoptosis through interaction with the BCL-2 pathway, and RRM2 expression may be positively correlated with the development of non-small cell lung cancer.

In terms of treatment, some studies have shown that low expression levels of RRM2 predicted better prognosis in patients with advanced non-small cell lung cancer who received platinum-based chemotherapy [22, 23]. However, Xian-Jun F et al. found RRM2 expression levels have no effect on platinum-based chemotherapy response and clinical outcomes in patients with NSCLC [24]. Therefore, the effect of RRM2 expression level on patients with NSCLC after platinum-based chemotherapy remains to be further studied.

Conclusion

In summary, we first demonstrated that the serum expression level of RRM2 is significantly up-regulated in NSCLC patient. RRM2 might serve as a potential molecular biomarker for diagnosis indicator of NSCLC. However, our study contains some limitations. First, the patients in the validation cohort were all Chinese. Second, samples for ELISA validation are rather small and RRM2 is not detected in other diseases. Third, the exact mechanism of RRM2 in the occurrence and progression of NSCLC need to be further investigated. Therefore, RRM2 as a molecular biomarker for diagnosis indicator of NSCLC is still

to do more experiments, including increasing sample size and scope as well as the functional mechanism of RRM2 in NSCLC.

List Of Abbreviations

NSCLC

non-small cell lung cancer

GO

gene ontology

CC

cellular components

BP

biological processes

MF

molecular functions

KEGG

the Kyoto Encyclopedia of Genes and Genomes

PPI

protein-protein interaction

GEPIA

the Gene Expression Profiling Interactive Analysis

ELISA

enzyme linked immunosorbent assay

ECLIA

electrochemiluminescence immunoassay

LUAD

lung adenocarcinoma

LUSC

lung squamous cell carcinoma

ROC

receiver operating characteristic

RRM2

ribonucleotide reductase regulatory subunit M2

Declarations

Ethics approval and consent to participate: This study has been approved by the Chongqing Hospital of Traditional Chinese Medicine ethics committee(Supplementary File 9) and written informed consent was obtained from all patients(Supplementary File 10). We hereby declare that all the methods referred to in

the article are carried out in accordance with the relevant guidelines and provisions of the Declaration of Helsinki.

Consent for publication: Not Applicable (NA).

Availability of data and materials: The datasets generated and analysed during the current study are available in the Gene Expression Omnibus (GEO) repository and Supplementary File 11(Clinical data.xlsx).

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

Funding: This work was financially supported by Xinglin Scholar Research Promotion Project of Chengdu University of TCM (Grant no.: YYZX2020069).

Author's contributions: Dandan Zhou and Xiuming Zhai conceived and designed the experiments. Xiuming Zhai and Ruixue Zhang performed the experiments. Dandan Zhou and Zihe Dong prepared the figures and tables. All authors reviewed the manuscript.

Acknowledgements: Not Applicable (NA).

References

1. Bray, F., et al., *Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries*. CA Cancer J Clin, 2018. **68**(6): p. 394–424.
2. Blandin Knight, S., et al., *Progress and prospects of early detection in lung cancer*. Open Biol, 2017. **7**(9).
3. Siegel, R.L., K.D. Miller, and A. Jemal, *Cancer statistics*, 2016. CA Cancer J Clin, 2016. **66**(1): p. 7–30.
4. Pan, Q., et al., *Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing*. Nat Genet, 2008. **40**(12): p. 1413–5.
5. Geng, H., et al., *Survival prediction for patients with lung adenocarcinoma: A prognostic risk model based on gene mutations*. Cancer Biomark, 2020. **27**(4): p. 525–532.
6. Sanchez-Palencia, A., et al., *Gene expression profiling reveals novel biomarkers in nonsmall cell lung cancer*. Int J Cancer, 2011. **129**(2): p. 355–64.
7. Hou, J., et al., *Gene expression-based classification of non-small cell lung carcinomas and survival prediction*. PLoS One, 2010. **5**(4): p. e10312.
8. Rousseaux, S., et al., *Ectopic activation of germline and placental genes identifies aggressive metastasis-prone lung cancers*. Sci Transl Med, 2013. **5**(186): p. 186ra66.
9. Zhang, Y., et al., *USP44 regulates centrosome positioning to prevent aneuploidy and suppress tumorigenesis*. J Clin Invest, 2012. **122**(12): p. 4362–74.
10. Barrett, T., et al., *NCBI GEO: archive for functional genomics data sets—update*. Nucleic Acids Res, 2013. **41**(Database issue): p. D991-5.

11. Detterbeck, F.C., et al., *The Eighth Edition Lung Cancer Stage Classification*. Chest, 2017. **151**(1): p. 193–203.
12. The Gene Ontology, C., *The Gene Ontology Resource: 20 years and still GOing strong*. Nucleic Acids Res, 2019. **47**(D1): p. D330-D338.
13. Kanehisa, M., et al., *New approach for understanding genome variations in KEGG*. Nucleic Acids Res, 2019. **47**(D1): p. D590-D595.
14. Jiao, X., et al., *DAVID-WS: a stateful web service to facilitate gene/protein list analysis*. Bioinformatics, 2012. **28**(13): p. 1805–6.
15. Szklarczyk, D., et al., *STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets*. Nucleic Acids Res, 2019. **47**(D1): p. D607-D613.
16. Chin, C.H., et al., *cytoHubba: identifying hub objects and sub-networks from complex interactome*. BMC Syst Biol, 2014. **8 Suppl 4**: p. S11.
17. Tang, Z., et al., *GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses*. Nucleic Acids Res, 2017. **45**(W1): p. W98-W102.
18. Gagat, M., et al., *Potential role of cyclin F mRNA expression in the survival of skin melanoma patients: Comprehensive analysis of the pathways altered due to cyclin F upregulation*. Oncol Rep, 2018. **40**(1): p. 123–144.
19. Mah, V., et al., *Ribonucleotide reductase subunit M2 predicts survival in subgroups of patients with non-small cell lung carcinoma: effects of gender and smoking status*. PLoS One, 2015. **10**(5): p. e0127600.
20. Grossi, F., et al., *Expression of Ribonucleotide Reductase Subunit-2 and Thymidylate Synthase Correlates with Poor Prognosis in Patients with Resected Stages I-III Non-Small Cell Lung Cancer*. Dis Markers, 2015. **2015**: p. 302649.
21. Rahman, M.A., et al., *RRM2 regulates Bcl-2 in head and neck and lung cancers: a potential target for cancer therapy*. Clin Cancer Res, 2013. **19**(13): p. 3416–28.
22. Zhao, H., et al., *Prognostic significance of BRCA1, ERCC1, RRM1, and RRM2 in patients with advanced non-small cell lung cancer receiving chemotherapy*. Tumour Biol, 2014. **35**(12): p. 12679–88.
23. Wang, L., et al., *Expression of RRM1 and RRM2 as a novel prognostic marker in advanced non-small cell lung cancer receiving chemotherapy*. Tumour Biol, 2014. **35**(3): p. 1899–906.
24. Xian-Jun, F., et al., *ERCC1 and BRCA1 mRNA expression predicts the clinical outcome of non-small cell lung cancer receiving platinum-based chemotherapy*. Pak J Med Sci, 2014. **30**(3): p. 488–92.

Tables

Table 1
Tumor marker level in peripheral blood of NSCLC patients and healthy control

Indicators	NSCLC (n = 110)	Healthy control(n = 50)	Z value	P value
RRM2 (ng/ml)	151.72(129.37,166.98)	111.78(90.17,132.65)	-6.043	0.000
CEA (ng/ml)	5.00(2.85,12.39)	1.36(0.86,1.95)	-8.165	0.000
CYFRA21-1 (ng/ml)	2.97(1.92,5.19)	1.82(1.39,2.25)	-5.279	0.000
NSE (ng/ml)	19.23(16.27,26.14)	15.17(13.19,16.65)	-5.586	0.000
PROGRP (pg/ml)	48.48(37.27,65.60)	37.50(32.48,47.90)	-3.326	0.001

Table 2
Association between serum RRM2 level and clinical characteristics

Parameter	NSCLC	RRM2 (ng/ml)	Z value	P value
Tumor size				
Diameter ≤ 3 cm	14	156.78(115.70,161.41)	-0.271	0.787
Diameter > 3 cm	92	151.42(128.81,167.11)		
Lymph node metastasis				
Yes	79	152.18(135.82,166.72)	-1.279	0.201
No	24	144.21(99.35,167.00)		
Distant metastasis				
Yes	59	158.68(145.03,169.30)	-3.280	0.001
No	49	139.80(113.61,156.87)		
TNM stage				
I + II	16	114.66(91.09,150.21)	-3.069	0.002
III + IV	94	153.81(135.52,167.39)		
Histological type				
Adenocarcinoma	81	154.93(136.11,167.24)	-2.028	0.043
Squamous cell carcinoma	29	137.86(120.40,161.07)		

Table 3
ROC results of indicators as biomarker for NSCLC

Indicators	AUC	SE	<i>P</i>	95% C.I.	Cut off	Sensitivity	Specificity	Accuracy
RRM2	0.798	0.034	0.000	0.731–0.866	131.31	0.736	0.740	0.738
CEA	0.903	0.024	0.000	0.857–0.950	5.00	0.500	0.980	0.650
CYFRA21-1	0.761	0.037	0.000	0.688–0.834	3.31	0.445	0.940	0.600
NSE	0.776	0.037	0.000	0.703–0.849	16.25	0.755	0.720	0.738
PROGRP	0.664	0.045	0.001	0.577–0.752	68.99	0.209	0.940	0.438
RRM2 + NSE + CEA	0.958	0.015	0.000	0.929–0.987	-	0.909	0.94	-

Figures

Figure 1 Venn diagrams of the differentially expressed genes (DEGs).

Venn diagrams of the DEGs from four independent data sets (GSE18842, GSE19188, GSE30219, GSE40791). (a) a total of 258 DEGs were identified, (b) 91 up-regulated DEGs were identified, (c) 167 down-regulated DEGs were identified, using adjust *P* value < 0.05 and $|\log FC| \geq 2$ as cut-off criteria.

Figure 1

See image above for figure legend.

Figure 2 GO enrichment analysis of the DEGs.

GO enrichment analysis of DEGs in terms of 3 aspects: molecular functions(MF); cellular components(CC); biological processes(BP), using "after FDR" (corrected P-Value < 0.05, gene count ≥ 10) set as statistically significant.

Figure 2

See image above for figure legend.

Figure 3 Protein-protein interaction network of the DEGs.

- (a) Down-regulated genes are indicated by arrows and up-regulated genes are indicated by circles.
 Different colors indicate that the gene is involved in different biological content and pathways.
 (b) 10 hub genes screened by cytoHubba according to the MCC algorithm.

Figure 3

See image above for figure legend.

Figure 4 The expression level and survival analysis of RRM2 between NSCLC and normal controls

(a) The expression level of RRM2 was increased in both LUSC and LUAD ($P < 0.01$),

(b) Kaplan-Meier Plotter showed that the survival rates of NSCLC patients with high expression of RRM2 were statistically significant ($P = 0.03$).

Figure 4

See image above for figure legend.

Figure 5 ROC curves for various markers.

ROC curves show the combined detection of RRM2, CEA and NSE has the highest efficacy in identifying NSCLC patients.

Figure 5

See image above for figure legend.

Supplementary Files

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

- [Clinicaldata.xlsx](#)
- [Supplementalfile1.docx](#)

- [Supplementalfile2.docx](#)
- [Supplementalfile3.xls](#)
- [Supplementalfile4.xlsx](#)
- [Supplementalfile5.docx](#)
- [Supplementalfile6.docx](#)
- [Supplementalfile7.docx](#)
- [Supplementalfile8.docx](#)