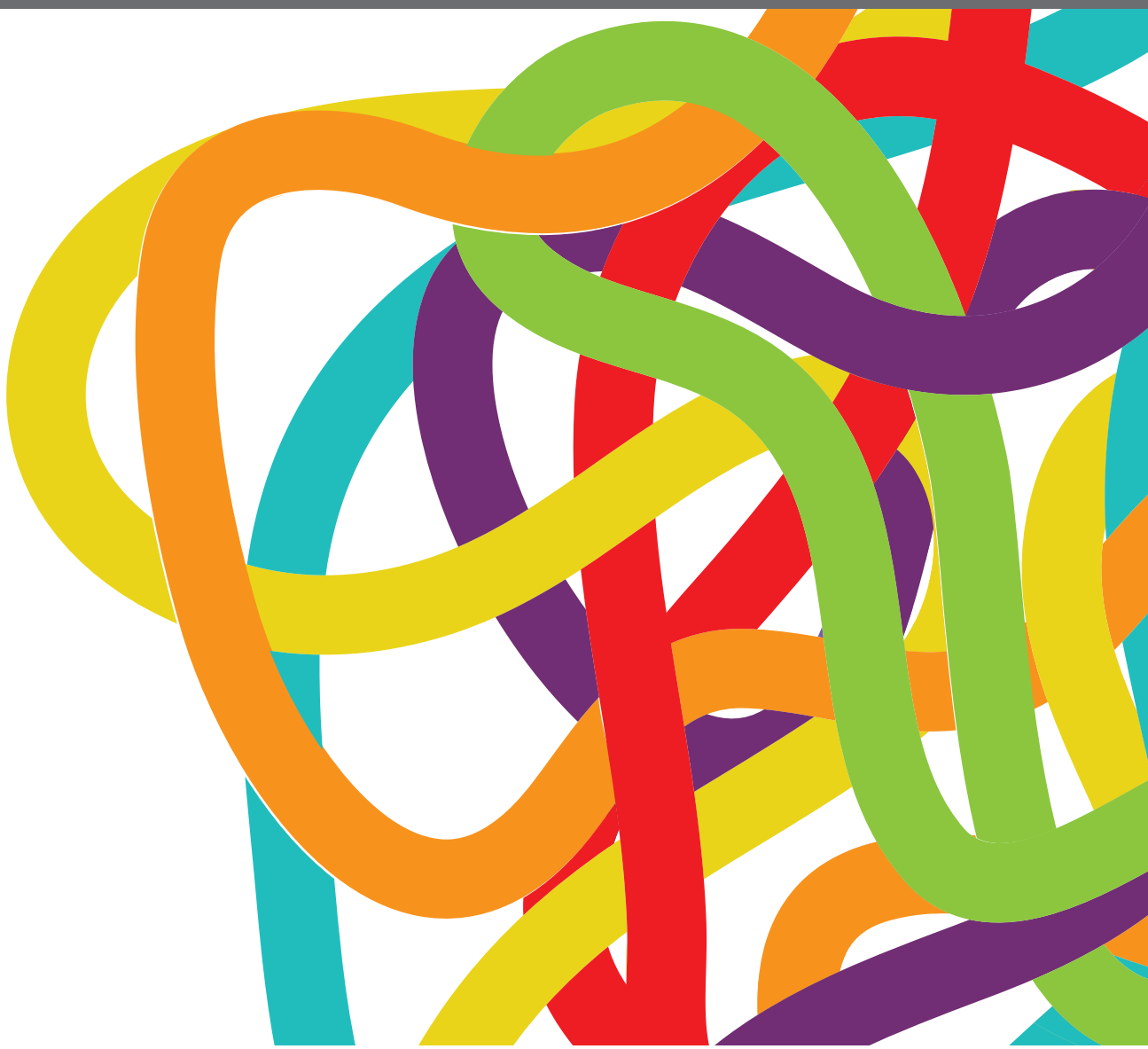
EMERGING BIOMARKERS FOR NSCLC: RECENT ADVANCES IN DIAGNOSIS AND THERAPY

EDITED BY: Umberto Malapelle, Etienne Giroux Leprieur, Christian Rolfo, Paul Takam Kamga and Marius Tresor Chiasseu

PUBLISHED IN: Frontiers in Oncology



frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88966-915-8
DOI 10.3389/978-2-88966-915-8

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

EMERGING BIOMARKERS FOR NSCLC: RECENT ADVANCES IN DIAGNOSIS AND THERAPY

Topic Editors:
Umberto Malapelle, University of Naples Federico II, Italy
Etienne Giroux Leprieur, Service de Pneumologie et d'Oncologie Thoracique, France
Christian Rolfo, University of Maryland Medical System, United States
Paul Takam Kamga, Université de Versailles Saint-Quentin-en-Yvelines, France
Marius Tresor Chiasseu, Yale University, United States

Citation: Malapelle, U., Leprieur, E. G., Rolfo, C., Kamga, P. T., Chiasseu, M. T., eds. (2021). Emerging Biomarkers for NSCLC: Recent Advances in Diagnosis and Therapy. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88966-915-8

Table of Contents

frontiers in Oncology

EDITORIAL
published: 14 May 2021
doi: 10.3389/fonc.2021.694578

Editorial: Emerging Biomarkers for NSCLC: Recent Advances in Diagnosis and Therapy

Umberto Malapelle[1]*, *Etienne Giroux Leprieur*[2,3], *Paul Takam Kamga*[3], *Marius Tresor Chiasseu*[4,5] *and Christian Rolfo*[6]

[1] Department of Public Health, University of Naples Federico II, Naples, Italy, [2] Department of Respiratory Diseases and Thoracic Oncology, APHP-Hopital Ambroise Pare, Boulogne-Billancourt, France, [3] EA 4340 BECCOH, UVSQ, Université Paris-Saclay, Boulogne-Billancourt, France, [4] Cellular Neuroscience, Neurodegeneration and Repair Program, Yale University School of Medicine, New Haven, CT, United States, [5] Departments of Neurology, Yale University School of Medicine, New Haven, CT, United States, [6] Thoracic Medical Oncology, Marlene and Stewart Greenebaum Cancer Center, University of Maryland, Baltimore, Baltimore, MD, United States

Keywords: biomarkers, NSCLC, diagnosis, therapy, NGS

Editorial on the Research Topic

Emerging Biomarkers for NSCLC: Recent Advances in Diagnosis and Therapy

The advent of precision medicine and predictive molecular pathology has significantly modified the clinical management of patients with non-small cell lung cancer (NSCLC). A plethora of different biomarkers has been approved for predictive molecular purposes (Zhu et al.; Li et al.; Lin et al.). In this scenario, molecular techniques able to optimize the limited amount of nucleic acids extracted from small tissue and/or liquid biopsy samples are essential for the different clinically relevant biomarkers evaluation. Next generation sequencing (NGS) is a fascinating molecular approach able to analyze different gene alterations from different patients, simultaneously, starting from low input material. However, it should be remembered that a careful process of validation and harmonization of wet and dry procedures are strongly warranted (Malapelle et al.). Beyond the administration of tyrosine kinase inhibitors, a high percentage of NSCLC patients without any driver alteration can benefit from the administration of immune-checkpoint inhibitors (ICIs). Despite the established role of the evaluation of programmed death-ligand 1 (PD-L1) expression through immunohistochemistry or immunocytochemistry on tissue specimens, several other biomarkers are currently under investigation. Among these, tumor mutational burden (TMB) evaluated on tissue samples is the most commonly studied. However, TMB evaluation suffers from some technical issues. Thus, the adoption of surrogate biomarkers, such as *MSH2* expression (Jia et al.), may be a valid option. In addition, blood TMB evaluation may be a valid opportunity to assess TMB status and monitor ICIs response (Friedlaender et al.). Liquid biopsy adoption is increasing due to a not negligible percentage (about 30%) of NSCLC patients who do not have tissue availability for molecular analysis. Beyond the predictive purposes, liquid biopsy may play a pivotal role in the early diagnosis and prognosis evaluation of lung cancer (Dong et al.; Xi et al.). In the setting of prognostic biomarkers, many data have emerged on NSCLC. In particular, micro RNA (miRNA) 1323 with high expression in lung adenocarcinomas, promoting cancer cell migration, is associated with a poor prognosis (Zhao H et al.). Other prognostic biomarkers are currently under investigation, in particular those related to metabolic reprogramming, extracellular matrix, and tumor microenvironment remodeling (Bi et al.; Yang et al.; Czarnecka et al.; Ma et al.; Ahmed).

OPEN ACCESS

Edited and reviewed by:
Vamsi Velcheti,
New York University, United States

***Correspondence:**
Umberto Malapelle
umberto.malapelle@unina.it

Specialty section:
This article was submitted to Thoracic Oncology, a section of the journal Frontiers in Oncology

Received: *13 April 2021*
Accepted: *19 April 2021*
Published: *14 May 2021*

Citation:
Malapelle U, Leprieur EG, Kamga PT, Chiasseu MT and Rolfo C (2021) Editorial: Emerging Biomarkers for NSCLC: Recent Advances in Diagnosis and Therapy. Front. Oncol. 11:694578. doi: 10.3389/fonc.2021.694578

Another interesting field of investigation concerns the development of prognostic models (Wu L-L et al.) and immunoscoring strategies to stratify early stage patients (Zhao Z et al.). Finally, careful attention should be paid to the novel approaches related to machine learning algorithms used to predict lymph node involvement in early T stage patients (Wu Y et al.), the possibility to isolate and characterize stem-like cells (Masciale et al.), and the possibility to adopt a radiomics-based nomogram to predict *EGFR* mutation subtypes (Zhao W et al.).

Taken together, the papers published in Research Topic "Emerging Biomarkers for NSCLC: Recent Advances in Diagnosis and Therapy" represent a critical discussion focused on the role of different novel biomarkers for both predictive and prognostic purposes.

AUTHOR CONTRIBUTIONS

Writing the original draft: UM and CR. All authors contributed to the article and approved the submitted version.

Conflict of Interest: UM has received personal fees (as consultant and/or speaker bureau) from Boehringer Ingelheim, Roche, MSD, Amgen, Thermo Fisher Scientifics, Eli Lilly, Diaceutics, GSK, Merck and AstraZeneca, unrelated to the current work. CR reported receiving research grants at Antwerp University Hospital, Belgium, from Novartis and Sanofi, receiving speaker fees from Guardant Health, Merck Sharp & Dohme, and Novartis, receiving scientific advisor fees from Mylan, serving on a steering scientific committee for Oncompass, and participating in research collaborations for OncoDNA and Biomark Inc, unrelated to the current work. EG has received personal fees (as consultant and/or speaker bureau) from AstraZeneca, Boehringer Ingelheim, Bristol-Myers-Squibb, Eli Lilly, MSD, Novartis, Roche, Takeda, and research grants (institution) from AstraZeneca, Bristol-Myers-Squibb, and Roche.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

ORIGINAL RESEARCH
published: 23 September 2019
doi: 10.3389/fonc.2019.00880

The Impact of Clinical Factors, ALK Fusion Variants, and BIM Polymorphism on Crizotinib-Treated Advanced EML4–ALK Rearranged Non-small Cell Lung Cancer

Yen-Ting Lin[1,2], *Yi-Nan Liu*[2] *and Jin-Yuan Shih*[1,2]*

[1] Graduate Institute of Clinical Medicine, College of Medicine, National Taiwan University, Taipei, Taiwan, [2] Department of Internal Medicine, National Taiwan University Hospital and College of Medicine, National Taiwan University, Taipei, Taiwan

Patients' clinical factors and genetics factors such as anaplastic lymphoma kinase (ALK) fusion variants and BIM (Bcl-2-like 11) polymorphism were reported to be associated with clinical outcome in crizotinib-treated advanced non-small cell lung cancer (NSCLC). However, the results were still controversial. We analyzed outcome of 54 patients with known ALK fusion variants who received crizotinib for advanced NSCLC. Thirty of them had successful BIM polymorphism analysis and 6 (20%) had a BIM deletion. Multivariate Cox regression analysis found that previous anticancer therapy [adjusted hazard ratio (aHR) 1.35, 95% confidence interval (CI), 1.04–1.76 for each additional line of therapy, $p = 0.025$] and Eastern Cooperative Oncology Group (ECOG) performance status ≥2 (aHR 8.35, 95% CI, 1.52–45.94, $p = 0.015$) were independent factors for progression-free survival (PFS). Only ECOG performance status ≥2 (aHR 7.20, 95% CI, 1.27–40.79, $p = 0.026$) was an independent factor for overall survival (OS). Neither ALK fusion variants nor the presence of a BIM deletion was associated with crizotinib PFS or OS. After adjusting with clinical factors, different ALK variants and BIM polymorphism might not be independent factors for crizotinib PFS or OS in advanced NSCLC with ALK rearrangement.

Keywords: non-small cell lung cancer, ALK, ALK variant, BIM, crizotinib

OPEN ACCESS

Edited by:
Umberto Malapelle,
University of Naples Federico II, Italy

Reviewed by:
Timothy F. Burns,
University of Pittsburgh, United States
Rabab Mohamed Gaafar,
Cairo University, Egypt

***Correspondence:**
Jin-Yuan Shih
jyshih@ntu.edu.tw

Specialty section:
This article was submitted to Thoracic Oncology, a section of the journal Frontiers in Oncology

Received: *03 June 2019*
Accepted: *27 August 2019*
Published: *23 September 2019*

Citation:
Lin Y-T, Liu Y-N and Shih J-Y (2019) The Impact of Clinical Factors, ALK Fusion Variants, and BIM Polymorphism on Crizotinib-Treated Advanced EML4–ALK Rearranged Non-small Cell Lung Cancer. Front. Oncol. 9:880. doi: 10.3389/fonc.2019.00880

INTRODUCTION

In 2007, the echinoderm microtubule-associated protein-like 4 (EML4)–anaplastic lymphoma kinase (ALK) gene rearrangement was first discovered as a driver oncogene for non-small cell lung cancer (NSCLC) (1). Inversion in chromosome 2p fused the N-terminal domain of EML4 to the intracellular kinase domain of ALK, causing constitutive activation of tyrosine kinase, leading to uncontrolled cell growth and proliferation. During the following 10 years, targeting ALK with tyrosine kinase inhibitors (TKIs) has achieved great success. The first-generation ALK TKI crizotinib had better progression-free survival (PFS) (10.9 vs. 7.0 months) and a better overall response rate (ORR) (74 vs. 45%) than chemotherapy in treating naïve ALK rearranged {ALK positive [ALK(+)]} NSCLC in the PROFILE 1014 study (2). Crizotinib has been approved by the US Food and Drug Administration (US FDA) as first-line treatment for ALK(+) advanced NSCLC (3).

The second-generation ALK TKIs alectinib (CH5424802/RO5424802) and ceritinib (LDK378) also showed promising activity in controlling ALK(+) NSCLC in phase 3 trials (4, 5). Moreover, in ALK TKI-naïve patients treated with brigatinib, a next-generation ALK TKI, the PFS was longer than patients treated with crizotinib (6). The development of ALK TKI for use against ALK(+) NSCLC is one of the best stories in the history of developing anticancer therapy.

However, most patients still experienced disease progression after ALK TKI treatment. The latest released data for East Asian patients in PROFILE 1029 revealed that median PFS was 11.1 (95% confidence interval, CI: 8.3–12.6) months for first-line crizotinib-treated advanced ALK(+) NSCLC patients (7). Several factors were reported to be associated with crizotinib PFS, but the two main groups were clinical factors and genetic factors. Traditional clinical factors such as the patient's performance status (8, 9) and brain metastasis (10–12) prior to crizotinib treatment were reported to influence crizotinib PFS. Among the genetic factors, one of the most common was ALK fusion variants. In the preclinical data, different ALK fusion variants were associated with crizotinib sensitivity. ALK fusion variant 2 had lower crizotinib IC_{50} than variant 3. Longer ALK fusion variants were the most unstable and were supposed to be more sensitive to crizotinib than shorter ALK fusion variants (13). The presence of a tandem atypical beta-propeller in the EML protein (TAPE) domain was reported to influence the stability of EML4–ALK protein (14); "short variants," such as variants 3a/b and 5a/b, lack a TAPE domain (15) and might be less responsive to crizotinib than the longer TAPE-containing variants, such as variant 1 and variant 2 (16). A circular RNA F-circEA found in only variant 3 was reported to promote cancer cell migration and proliferation (17). However, in spite of the supposed mechanism, the real-world data were conflicting. ALK variant 1 (18), variant 2 (19), and variants other than variant 3 (16, 20) were reported to have a better crizotinib PFS, but there were also several reports indicating that all variants had a similar outcome (21, 22). In fact, the largest cohort to date reported there was no crizotinib PFS difference between variant 1 and variant 3 (23). Whether or not different EML4–ALK fusion variants influence crizotinib PFS remains controversial.

Another interesting genetic factor was Bcl-2-like 11 (BIM). BIM is a pro-apoptotic member of the B-cell CLL/lymphoma 2 (BCL2) family of proteins, discovered in Asia only. Its upregulation is required for TKIs to induce apoptosis in kinase-driven cancers (24). The BIM deletion polymorphism was reported to be associated with primary resistance to or a short PFS with epidermal growth factor receptor (EGFR) TKI in advanced EGFR-mutant NSCLC (24, 25). Another report indicated that BIM deletion was related to a poor crizotinib response in advanced ALK(+) NSCLC (26). However, in our previous study, we could not find a relationship between the BIM deletion polymorphism and primary EGFR TKI resistance among our 327 Taiwanese patients, while 52 (16%) of them were positive for BIM deletion (27). In this study, we aimed to analyze the association of clinical factors and genetic factors, including ALK fusion variants and BIM polymorphism, with crizotinib PFS and overall survival (OS) in advanced EML4–ALK(+) NSCLC patients.

METHODS

Patients

This study retrospectively enrolled patients receiving crizotinib for EML4–ALK rearrangement stage IV or postoperative recurrent (advanced) NSCLC between December, 31, 2010, and December, 31, 2017, at the National Taiwan University Hospital. Only patients with data on EML4–ALK variants using reverse transcription polymerase chain reaction (RT-PCR) were included. Patients who stopped crizotinib within 30 days due to intolerable side effects were excluded. Patients' baseline characteristics, including age, gender, smoking status, previous anticancer therapy, Eastern Cooperative Oncology Group (ECOG) performance status (28), prior brain metastasis, EML4–ALK variants, and status of BIM polymorphism, were checked. The patients were treated and followed up based on the clinician's decision. A blinded chest physician who was not involved in patient management and did not know the laboratory data on EML4–ALK variants and BIM polymorphism retrospectively reviewed the chart and images to determine disease progression according to RECIST criteria version 1.1 (29). PFS was defined as the duration from the first dose of crizotinib to disease progression or death during treatment. OS was defined as the duration from the first dose of crizotinib to the patient's death. Each patient's best overall response, PFS, and OS were recorded. This study was approved by the Institutional Review Board of National Taiwan University Hospital. Written informed consent was obtained from all patients before checking their cancer specimens for molecular studies. All methods were performed in accordance with the relevant guidelines and regulations.

Analysis for EML4–ALK Fusion Gene

Using immunohistochemistry (IHC) stain, we checked the patients' cancer specimens for ALK using Ventana ALK (D5F3) antibody. We further analyzed cancer specimens for EML4–ALK variants using RNA RT-PCR, as previously described (30). In brief, RNA extracted from patients' tissue specimens were collected for RT-PCR amplification by a OneStep RT-PCR Kit (Qiagen) using the following primers: 5′-TGGCTGATGTTTTGAGGCGT-3′ (forward, on exon 2 of EML4), 5′-AGAGCCCACACCTGGGAAAG-3′ (forward, on exon 13 of EML4), 5′-CCACACAGACGGGAATGAAC-3′ (forward, on exon 18 of EML4), and 5′-AGCAAAGCAGTAGTTGGGGT-3′ (reverse, on exon 20 of ALK). The PCR conditions were as follows: 50°C for 30 min, 95°C for 15 min (94°C for 50 s, 60°C for 50 s, 72°C for 60 s) × 40 cycles, and 72°C for 10 min. RT-PCR amplicons were purified and sequenced with Sanger sequencing in both sense and antisense directions. Because the length of the ALK fusion protein may contribute to its stability (13) and probably crizotinib PFS, we also separated different ALK fusion variants into a long group and a short group. Short ALK fusion variants were defined as variants that do not have the TAPE main, i.e., variant 3a, variant 3b, variant 5a, and variant 5b. Long ALK

fusion variants were defined as EML4–ALK fusion variants that contain the TAPE, i.e., all variants other than variant 3a, 3b, 5a, or 5b (15, 22).

Analysis for the BIM Polymorphism

We checked patient cancer specimens with known EML4–ALK fusion variants for further BIM polymorphism analysis, as previously described (24). Cancer DNA was extracted from cancer specimens using the QIAamp DNA Mini Kit (Qiagen). PCR reactions were done to determine the presence of wild-type or deletion alleles using high-fidelity JumpStart™ REDAccuTaq® LA DNA Polymerase (Sigma) with the following conditions: 96°C for 30 s (94°C for 15 s, 60°C for 60 s, 68°C for 10 min) × 30 cycles, and 68°C for 20 min. The forward primer was 5′-AATACCACAGAGGCCCACAG-3′ and the reverse primer was 5′-GCCTGAAGGTGCTGAGAAAG-3′. The PCR products for the deletion (1,323 bp) and the wild-type (4,226 bp) alleles were applied on a 1% agarose gel and were sequenced.

Statistical Analysis

Continuous variables were reported as median with interquartile range (IQR). Categorical data were compared using the chi-square test. PFS and OS were plotted using the Kaplan–Meier method and compared by log-rank test. A Cox proportional hazard model was used for univariate and multivariate analysis for crizotinib PFS and OS. Variables with $p < 0.2$ in the univariate analysis and clinically important variables such as ALK variant type, BIM deletion, and brain metastasis prior to crizotinib were forced into the final model. Statistical significance was set at $p < 0.05$. All statistical analyses were performed using the Statistical Package for the Social Sciences, version 18.0K (SPSS, Inc., Chicago, IL, USA). The data cutoff date was September 23, 2018.

RESULTS

Patient Demographic and Clinical Characteristics

A total of 104 ALK IHC(+) patients received crizotinib for advanced NSCLC during the study period. Fifty-five patients had known EML–ALK fusion variants, as determined by RT-PCR. One patient who received crizotinib for <30 days because of side effects was excluded. A total of 54 patients with known EML4–ALK fusion variants were included in the study. Because of the overlapping enrollment interval, 13 of the 54 patients were included in another published article (31). Thirty of the total 54 patients had adequate tissue for BIM polymorphism analysis.

Twenty-three patients had ALK variant 1; six patients had ALK variant 2; 18 patients had ALK variant 3a/b; and seven patients had other ALK variants (two with variant 5, one with variant V5a, two with variant 6, one with variant 8, and one with variant 1 plus insertion of 117 base pairs). The median follow-up time of the cohort was 13.8 (IQR, 7.4–25.4) months. Most patients had received prior anticancer therapy (median, 2 lines of prior anticancer therapy before crizotinib, range, 0–12) and three patients had received crizotinib as first-line therapy. The crizotinib response rate and the median follow-up time did not differ between the different ALK variant groups (**Table 1**). Patients with variant 2 had better ECOG performance status (0 or 1) (**Table 1**). Patients with long ALK variants were younger than patients with short ALK variants ($p = 0.03$) (**Supplementary Table 1**). In patients with long ALK variants, the baseline characteristics were not different significantly between variant 2 and other long ALK variants (**Supplementary Table 2**).

BIM deletion polymorphism was found in 20% (6/30) of the patients. There was no significant difference in demographic data between patients with deletion polymorphism and wild type (**Supplementary Table 3**).

Progression-Free Survival

The median crizotinib PFS was 7.3 [95% confidence interval (CI), 4.2–10.4] months in this cohort. The median PFS did not differ significantly among the four ALK variant groups [variant 1, 6.1 (95% CI, 1.6–10.6) months; variant 2, 11.0 (95% CI, 0–22.1) months; variant 3, 7.3 (95% CI, 3.6–10.9) months; other variants, 5.5 (95%, 3.1–8.0) months, $p = 0.33$ by log-rank test, **Figure 1A**]. The median PFS also did not differ significantly between variant 2 and all other variants [variant 2, 11.0 (95% CI, 0–22.1) months; all other variants, 6.1 (95% CI, 2.7–9.5) months, $p = 0.21$ by log-rank test, **Figure 1B**], between long ALK variants and short variants [long ALK variants, 6.1 (95% CI 2.3–9.8) months; short ALK variants, 8.2 (95% CI, 3.7–12.7) months, $p = 0.97$ by log-rank test, **Figure 1C**], and between BIM deletions and not [BIM deletion, 5.5 (95% CI 0–26.6) months; wild-type BIM, 8.6 (95% CI, 3.5–13.7) months, $p = 0.57$ by log-rank test, **Figure 1D**]. Multivariate analysis found that ECOG performance status ≥2 [adjusted hazard ratio (aHR) 8.35, 95% CI, 1.52–45.94, $p = 0.015$] and previous anticancer therapy (aHR 1.35, 95% CI, 1.04–1.76 for each additional line of therapy, $p = 0.025$) were independent factors for crizotinib PFS (**Table 2**). However, EML4–ALK fusion variants and BIM deletion were not independent factors for crizotinib PFS. ALK variant 1, variant 2, and variant 3a/b had nearly equal aHR (1.00 as the reference, 0.99 and 1.30, respectively). BIM deletion had a nearly neutral aHR 0.88, as well.

Overall Survival

The median OS was 22.0 [95% confidence interval (CI), 15.3–28.7) months in the cohort. The median OS did not differ significantly among the four ALK variant groups [variant 1, 16.1 (95% CI, 10.6–21.5) months; variant 2, not reached; variant 3, 25.1 (95% CI, 5.4–44.7) months; other variants, 10.3 (95%, 7.7–12.9) months, $p = 0.45$ by log-rank test, **Figure 2A**]. The median OS also did not differ significantly between variant 2 and all other variants [variant 2, not reached; all other variants, 19.7 (95% CI, 11.9–27.4) months, $p = 0.21$ by log-rank test, **Figure 2B**], between long ALK variants and short variants [long ALK variants, 18.5 (95% CI 10.3–26.8) months; short ALK variants, 25.1 (95% CI, 9.2–40.9) months, $p = 0.85$ by log-rank test, **Figure 2C**], and between BIM deletion and not [BIM deletion, 25.1 (95% CI 0–71.6) months; wild-type BIM, 22.0 (95% CI, 11.1–32.9) months,

TABLE 1 | Demographic data ($n = 54$).

Variable	Variant 1 ($n = 23$)	Variant 2 ($n = 6$)	Variant 3a/b ($n = 18$)	Other variants ($n = 7$)	p-value
Median age (years) (IQR)	56 (47–62)	50 (45–57)	62 (55–65)	56 (46–61)	0.16
Male	13 (57%)	2 (33%)	12 (67%)	5 (71%)	0.46
Never-smoker	13 (57%)	6 (100%)	13 (72%)	6 (86%)	0.14
Previous anticancer therapy (line)	2 (1–5)	1 (1–5)	2 (1–3)	4 (4, 5)	0.15
ECOG ≥2 before crizotinib	5 (22%)	0 (0%)	2 (11%)	4 (57%)	0.04
Brain metastasis before crizotinib	10 (44%)	1 (17%)	5 (28%)	4 (57%)	0.34
Best crizotinib response*					0.81
PR	9 (43%)	4 (66%)	11 (60%)	4 (57%)	
SD	8 (38%)	1 (17%)	4 (22%)	1 (14%)	
PD	4 (19%)	1 (17%)	3 (18%)	2 (29%)	
BIM deletion ($n = 30$)	3/12 (25%)	0/4 (0%)	2/9 (22%)	1/5 (20%)	0.75
Median follow-up time (months) (IQR)	15.4 (5.0–22.3)	18.2 (10.2–43.9)	15.9 (8.7–35.2)	9.9 (7.7–20.3)	0.47

**Two patients with variant 1 were not evaluable for crizotinib response.*
IQR, interquartile range; ECOG, Eastern Cooperative Oncology Group performance score; BIM, Bcl-2-like 11.

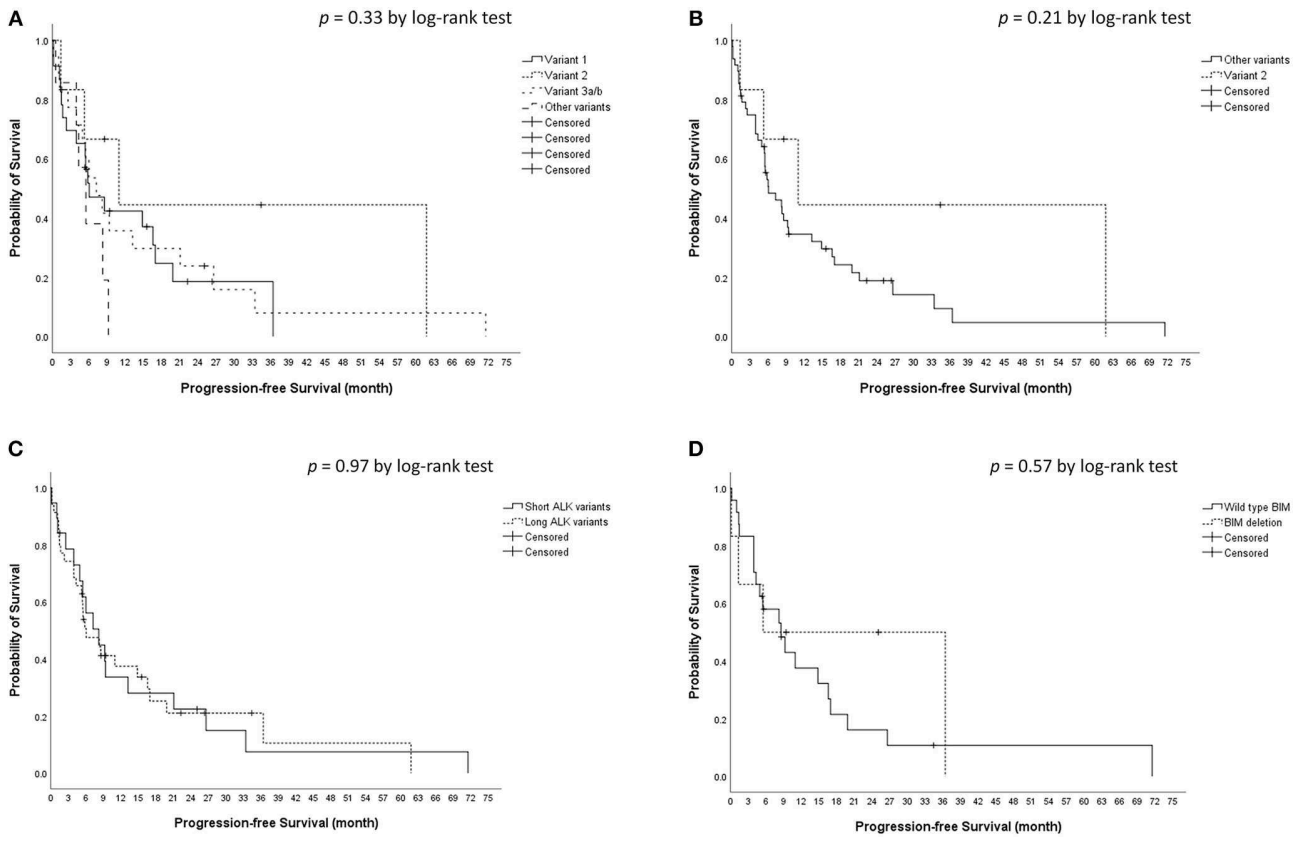


FIGURE 1 | Kaplan–Meier analyses for progression-free survival (PFS). **(A)** PFS among different ALK fusion variants. **(B)** PFS between ALK fusion variant 2 and other fusion variants. **(C)** PFS between short (variant 3a/b and 5a/b) and long (all other variants) ALK fusion variants. **(D)** PFS between BIM deletion and wild type BIM.

$p = 0.57$ by log-rank test, **Figure 2D**]. Multivariate analysis found that ECOG performance status ≥2 (aHR 7.20, 95% CI, 1.27–40.79, $p = 0.026$) was an independent factor for OS (**Table 3**), while ALK fusion variants and BIM deletion were not.

DISCUSSION

We found that clinical factors such as prior anticancer therapy and ECOG performance status were independent factors for crizotinib PFS in advanced NSCLC bearing EML4–ALK fusion,

TABLE 2 | Progression-free survival: univariate and multivariate analysis ($n = 54$).

Variable	Univariate analysis			Multivariate analysis		
	Hazard ratio	95% CI	*p*-value	Adjusted hazard ratio	95% CI	*p*-value
Age (≥65)	0.79	0.35–1.80	0.58			
Male sex	1.04	0.56–1.93	0.89			
Never-smoker	0.94	0.49–1.81	0.86			
ALK variants			0.36			0.87
Variant 1	1.00[b]			1.00[b]		
Variant 2	0.52	0.17–1.59	0.25	0.99	0.18–5.35	0.99
Variant 3a/b	0.93	0.46–1.86	0.93	1.30	0.38–4.43	0.68
Other variants[a]	1.73	0.67–4.47	0.26	0.64	0.14–2.94	0.57
ECOG ≥2	3.76	1.72–8.21	0.001	8.35	1.52–45.94	0.015
Previous anticancer therapy (per line)	1.14	1.02–1.27	0.02	1.35	1.04–1.76	0.025
Initial brain metastasis	1.53	0.82–2.85	0.18	0.72	0.18–2.87	0.64
BIM deletion	0.73	0.25–2.18	0.58	0.88	0.27–2.86	0.83

[a]*ALK variants other than variants 1, 2, or 3a/b.*

[b]*As a reference compared to other ALK variants.*

CI, confidence interval; ALK, anaplastic lymphoma kinase; ECOG, Eastern Cooperative Oncology Group performance score; BIM, Bcl-2-like 11.

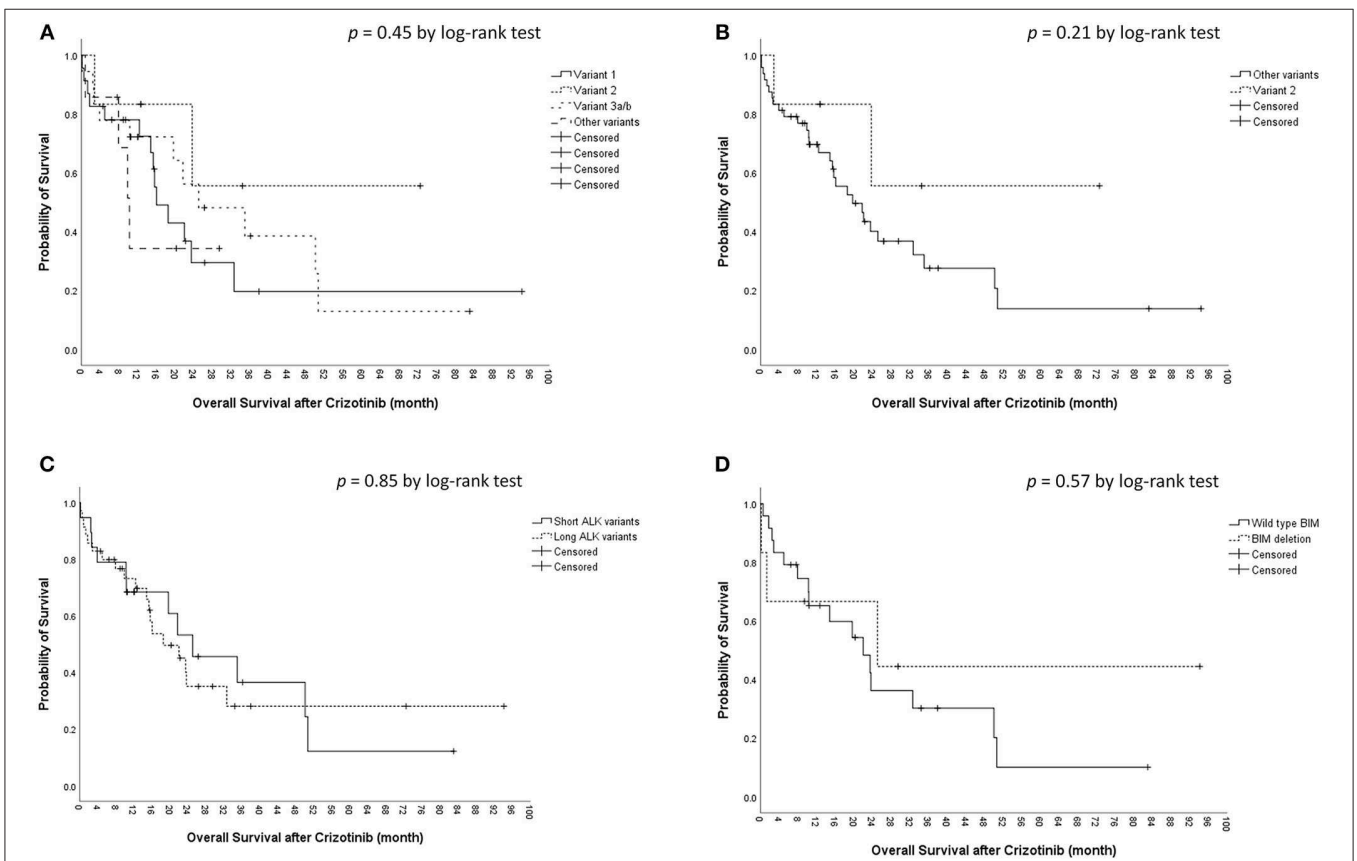


FIGURE 2 | Kaplan–Meier analyses for overall survival (OS). **(A)** OS among different ALK fusion variants. **(B)** OS between ALK fusion variant 2 and other fusion variants. **(C)** OS between short (variant 3a/b and 5a/b) and long (all other variants) ALK fusion variants. **(D)** OS between BIM deletion and wild-type BIM.

while ALK fusion variants and BIM polymorphism were not. In this cohort with mainly previously treated patients, for each additional line of anti-cancer therapy, the adjusted HR was 1.42 (95% CI, 1.06–1.86). With regard to the first-line PROFILE 1014 (2) and second-line PROFILE 1007 (32) phase 3 trials for ALK(+) NSCLC, the median crizotinib PFS in

TABLE 3 | Overall survival: univariate and multivariate analysis ($n = 54$).

Variable	Univariate analysis			Multivariate analysis		
	Hazard ratio	95% CI	*p*-value	Adjusted hazard ratio	95% CI	*p*-value
Age (≥65)	1.07	0.44–2.63	0.89			
Male sex	1.40	0.68–2.90	0.36			
Never-smoker	0.72	0.34–1.51	0.38			
ALK variants			0.48			0.48
Variant 1	1.00[b]			1.00[b]		
Variant 2	0.37	0.08–1.64	0.19	1.15	0.20–6.75	0.88
Variant 3a/b	0.74	0.33–1.64	0.45	1.13	0.28–4.49	0.86
Other variants[a]	1.27	0.41–3.88	0.68	0.22	0.03-1.53	0.13
ECOG ≥2	3.09	1.38–6.93	0.006	7.20	1.27–40.79	0.026
Previous anticancer therapy (per line)	1.15	1.02–1.29	0.028	1.27	0.97–1.66	0.09
Initial brain metastasis	1.65	0.79–3.43	0.18	0.89	0.19–4.13	0.88
BIM deletion	0.70	0.20–2.43	0.57	0.97	0.24–3.95	0.97

[a]*ALK variants other than variants 1, 2, or 3a/b.*

[b]*As a reference compared to other ALK variants.*

CI, confidence interval; ALK, anaplastic lymphoma kinase; ECOG, Eastern Cooperative Oncology Group performance score; BIM, Bcl-2-like 11.

the first-line trial seemed to be better (10.9 vs. 7.0 months). Zhou et al. reported on 73 ALK(+) NSCLC patients that received first-line crizotinib, pemetrexed/platinum, or non-pemetrexed chemotherapy/platinum. Poor ECOG performance status and crizotinib after non-pemetrexed chemotherapy were two independent factors for poor crozitinib PFS in multivariate analysis (33). Lin et al. reported on 94 advanced ALK(+) NSCLC patients and found that crizotinib had a better PFS in first-line use than in second-line use (median PFS 10.5 vs. 8.3 months, $p = 0.020$) (21). Unlike EGFR-TKIs, whose performance was the same in first-line or second-line treatment for advanced EGFR-mutant NSCLC (34), crizotinib had a tendency to do better in first-line use. Poor ECOG performance prior to crizotinib therapy was another independent factor for crizotinib PFS and OS. Poor performance status is a traditional negative prognostic marker among oncology patients (35). It was also found in crizotinib-treated advanced ALK(+) NSCLC patients in previous reports (8, 9, 33, 36).

Different EML4–ALK fusion variants were reported to influence crizotinib efficacy, but results from different reports are conflicting. The results of the current study and of previous reports regarding EML4–ALK fusion variants and crizotinib PFS are summarized in **Table 4** (16, 18–23, 37). All studies were conducted in a single center except for the report by Mitiushkina et al., which included three different hospitals in St. Petersburg, Russia (22). Yoshida et al. *first* reported that ALK fusion variant 1 had better crizotinib PFS in 35 Japanese patients (18). This was the first clinical report on the influence of different ALK fusion variants on crizotinib PFS. The patient numbers were relatively small and it only included first-line crizotinib-treated patients. Moreover, Lin et al. reported that 55 patients with variant 1 and variant 3 received first-line crizotinib, and the PFS was similar (23). Li et al. (19), Woo et al. (16), and Christopoulos et al. (20) reported responsiveness of patients from China, South Korea, and Germany to crizotinib in 2018, and the results were similar. Although Li et al. concluded that variant 2 had better crizotinib PFS, there was still a tendency for non-variant 3a/b patients to have a longer crizotinib PFS (median, 18.4 vs. 13.1 months, $p = 0.24$), which was consistent with the findings reported by Woo et al. and Christopoulos et al. The three studies had a similar characteristic: the majority of patients had variant 3a/b. However, while variant 3 had disadvantages in both PFS and OS as reported by Christopoulos et al. (20), the OS was almost the same ($p = 0.96$), as reported by Woo et al. (16). On the other hand, the largest cohort to date by Lin et al. showed that there was no difference between variant 1 and variant 3, in patients treated with both first-line ALK TKI as crizotinib and first-line crizotinib (23). Lei et al. (21), Cha et al. (37), Mitiushkina et al. (22), and our study found no difference between different ALK fusion variants. In the five studies from the United States, China, South Korea, Russia, and Taiwan, the majority of patients had variant 1. Both positive and negative reports included Caucasian and Asian patients, so race may not have contributed to the differences in results. Is it possible that the composite of variants in study cohorts had some influence on the results? The patient percentages of variant 3 in the five studies, which did not find differences between variants, were 48%, 30%, 19%, 25%, and 33%, respectively. In fact, patients with variant 3 were the second largest group among the cohorts, which does not lend support to the hypothesis of smaller patient numbers leading to an overestimation of PFS for variant 3 in the studies. One of the possible explanations may be the use of multivariate analysis. Only studies by Yoshida et al. and Li et al., and our study used a multivariate analysis to determine the independent factors for crizotinib PFS. Although the clinicopathologic characteristics seemed to be similar between the two analyzed groups (such as variant 3a/b or non-variant 3a/b), multivariate analysis that includes clinically relevant variables may still be a better method to find independent factors. Different patient groupings may influence crizotinib PFS if they are not adjusted appropriately.

TABLE 4 | Current reports of ALK fusion variants and crizotinib progression-free survival.

References	Study site	Detection of ALK fusion	ALK TKI	Timing of ALK TKI	Patient number	Prominent ALK fusion variant	Non-EML4-ALK fusion	PFS difference	Median ALK TKI PFS (months)	Multi-variate analysis
Yoshida et al. (18)	Japan	RT-PCR	Crizotinib	First-line	35	**V1, 54%** (54%/14%/12%)*	No	V1 longer	11.0 vs. 4.2	Yes
Li et al. (19)	China	NGS	Crizotinib	Mixed	49	**V3a/b, 33%** (23%/15%/33%)*	Yes (18%)	V2 longer	34.5 vs. 12.3	Yes
Woo et al. (16)	South Korea	RT-PCR	Crizotinib	Mixed	44	**V3a/b, 44%** (33%/11%/44%)*	Yes (6%)	Non-V3a/b longer	Not-reached vs. 11.0	No
Christopoulos et al. (20)	Germany	RT-PCR, NGS	Crizotinib, Alectinib, Ceritinib	Mixed	67	**V3a/b, 51%** (39%/10%/51%)*	No	Non-V3a/b Longer	39.3 vs. 7.3	No
Lin et al.† (23)	United States	RT-PCR, DNA direct sequencing or NGS	Crizotinib	First-line ALK TKI[a] and first-line[b]	99†[a] 55[b]	**V1† 52%** (V3 48%)[a] **V3† 51%** (V1 49%)[b]	No	No difference[a,b]	9.2 vs. 7.5[a] 8.9 vs. 6.9 [b]	No
Lei et al. (21)	China	RACE-coupled PCR	Crizotinib	Mixed	61	**V1, 36%** (36%/12%/30%)*	Yes (3%)	No difference	V1 vs. V3 vs. others: 11 vs. 10.9 vs. 7.4	No
Cha et al. (37)	South Korea	RT-PCR	Crizotinib	Mixed	32	**V1, 39%** (39%/6%/19%)*	Yes (37%)	No difference	Not disclosed in numbers[#]	No
Mitiushkin et al. (22)	Russia	RT-PCR	Crizotinib, Alectinib, Ceritinib	Mixed	64	**V1, 52%** (52%/5%/25%)*	Yes (2%)	No difference	Not disclosed in numbers[#]	No
Current study	Taiwan	RT-PCR	Crizotinib	Mixed	54	**V1, 43%** (43%/11%/33%)*	No	No difference	V1 vs. V2 vs. V3 vs. others: 6.1 vs. 11.0 vs. 7.3 vs. 5.5	Yes

**Proportion of variant 1/variant 2/variant 3a/b in study cohorts.*

† It is the largest cohort to date. It only compared variant 1 with variant 3. Data from patients with other ALK variants were not disclosed. Patients who received crizotinib as first-line ALK TKI[a] and crizotinib as first-line therapy[b] were listed.

[#]Only Kaplan–Meier curves were available.

The bold values were used to emphasize the prominent ALK fusion variant and its percentage only.

This may partly explain the discordance of OS data between the reports from Woo et al. and Christopoulos et al. In Christopoulos's cohort, variant 3a/b patients had more initial metastatic sites, either thoracic or extra-thoracic, and fewer patients with variant 3a/b had cancer recurrence from an early-stage cancer rather than initial stage IV (20). More metastatic sites and less cancer recurrence from early-stage NSCLC had survival disadvantages (38, 39), and might have contributed to shorter PFS and OS in variant 3a/b patients in the cohort. On the other hand, in our cohort, although patients with variant 2 tended to have a longer PFS, they might also have clinical advantages (tended to be younger, never-smokers, with better baseline performance status, and with less initial brain metastasis) (**Table 1**). The PFS between variant 1, variant 2, and variant 3a/b were almost equal after multivariate analysis (aHR, 1.00 as reference, 0.99, 1.30, respectively, **Table 2**). We hypothesize that although different ALK fusion variants might contribute to different crizotinib PFS, the impact may not be significant after adjusting for clinical factors.

In this study, there were 30 patients with enough tissue for BIM analysis and six were positive for BIM deletion (20%). The prevalence rate was consistent with previous reports (11–19%) (24–27). The BIM deletion polymorphism was not associated with a difference in crizotinib PFS (**Figure 1D**) or OS (**Figure 2D**). Using the multivariate Cox proportional hazard model, BIM deletion was also not related to differences in PFS or OS (**Tables 2**, **3**). BIM deletion was associated with shorter PFS in 47 ALK(+) NSCLC patients receiving crizotinib (26). BIM polymorphism was also reported to be associated with primary resistance or short PFS with EGFR TKIs (24, 25). However, Lee et al. checked 193 patients who received EGFR TKI for EGFR-mutant NSCLC and there was also no difference in EGFR TKI PFS between patients with and those without a BIM deletion (40). The result was similar to our previous analysis (27). Although BIM is a pro-apoptotic protein and may be related to TKI-induced cancer cell death, lung cancer cells may not be totally dependent on this pathway, and the concentration of BIM protein may also matter. Furthermore, the BIM deletion polymorphism is found only in Asians, and not in Caucasians (24). If the BIM deletion polymorphism was associated with shorter PFS, the effectiveness of crizotinib among Asians would be worse than in Western countries, but this is not true. Whether or not a simple BIM gene deletion influences TKI efficacy in NSCLC patients remains questionable.

There were several limitations to this study. First, it was a retrospective cohort study in a single center, as in previous reports. Because of the rarity of ALK(+) NSCLC, the patient number was still limited. The BIM deletion polymorphism in ALK(+) NSCLC patients, which is found in only 10–20% of ALK(+) patients, is even rarer. This may also be the reason that different reports have had different findings to date. As a result of limited patient numbers, the resistance mechanisms could not be addressed. Further larger multicenter or international prospective cohorts are warranted. Second, this was a cohort with mainly previously treated patients. Our results may not be generalizable to patients receiving first-line crizotinib therapy. Because reimbursement of crizotinib as first-line therapy was not approved by Taiwan's National Health Insurance until November 1, 2017, only three of 54 patients in our cohort used crizotinib as first-line therapy. However, although the U.S. FDA approved first-line crizotinib therapy, almost other studies also included mixed-line therapy with crizotinib, and purely first-line crizotinib data were rare (**Table 4**). Third, we used RT-PCR to determine ALK fusion variants. As in prior reports, not all ALK fusion variants could be detected. With the development of next-generation sequencing, more ALK fusion variants can be found, and the entire picture of ALK fusion lung cancer will become clearer.

In conclusion, clinical factors such as more prior anticancer therapies and ECOG performance status $\geq$2 were associated with a poorer crizotinib outcome. Different ALK variants and the BIM polymorphism were not independent factors for crizotinib PFS or OS in this study.

DATA AVAILABILITY

The datasets that were analyzed during the current study are available from the corresponding author on reasonable request.

ETHICS STATEMENT

This study was reviewed and approved by the Institutional Review Board of National Taiwan University Hospital. Written informed consent was obtained from all patients before checking their cancer specimens for molecular studies. All methods were performed in accordance with the relevant guidelines and regulations.

AUTHOR'S NOTE

An earlier version of this study was presented as a poster presentation in the Asian Pacific Society of Respirology 2018 Congress.

AUTHOR CONTRIBUTIONS

Y-TL participated in the study design, review and collection of patients' data, statistical analysis, and drafting of the manuscript. Y-NL participated in collection of patients' data, analyses of ALK and BIM, and revision of the manuscript. J-YS designed the study, interpreted the data, and reviewed and revised the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2019.00880/full#supplementary-material

REFERENCES

1. Soda M, Choi YL, Enomoto M, Takada S, Yamashita Y, Ishikawa S, et al. Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. *Nature.* (2007) 448:561–6. doi: 10.1038/nature05945
2. Solomon BJ, Mok T, Kim DW, Wu YL, Nakagawa K, Mekhail T, et al. First-line crizotinib versus chemotherapy in ALK-positive lung cancer. *N Engl J Med.* (2014) 371:2167–77. doi: 10.1056/NEJMoa1408440
3. Malik SM, Maher VE, Bijwaard KE, Becker RL, Zhang L, Tang SW, et al. U.S. Food and Drug Administration approval: Crizotinib for treatment of advanced or metastatic non-small cell lung cancer that is anaplastic lymphoma kinase positive. *Clin Cancer Res.* (2014) 20:2029–34. doi: 10.1158/1078-0432.CCR-13-3077
4. Peters S, Camidge DR, Shaw AT, Gadgeel S, Ahn JS, Kim DW, et al. Alectinib versus crizotinib in untreated ALK-positive non-small-cell lung cancer. *N Engl J Med.* (2017) 377:829–38. doi: 10.1056/NEJMoa1704795
5. Soria JC, Tan DSW, Chiari R, Wu YL, Paz-Ares L, Wolf J, et al. First-line ceritinib versus platinum-based chemotherapy in advanced ALK-rearranged non-small-cell lung cancer (ASCEND-4): a randomised, open-label, phase 3 study. *Lancet.* (2017) 389:917–29. doi: 10.1016/S0140-6736(17)30123-X
6. Camidge DR, Kim HR, Ahn MJ, Yang JC, Han JY, Lee JS, et al. Brigatinib versus crizotinib in ALK-positive non-small-cell lung cancer. *N Engl J Med.* (2018) 379:2027–39. doi: 10.1056/NEJMoa1810171
7. Wu YL, Lu S, Lu Y, Zhou J, Shi YK, Sriuranpong V, et al. Results of PROFILE 1029, a Phase III comparison of first-line crizotinib versus chemotherapy in East Asian patients with ALK-positive advanced non-small cell lung cancer. *J Thorac Oncol.* (2018) 13:1539–48. doi: 10.1016/j.jtho.2018.06.012
8. Ock CY, Yoo SH, Keam B, Kim M, Kim TM, Jeon YK, et al. Clinical factors affecting progression-free survival with crizotinib in ALK-positive non-small cell lung cancer. *Korean J Intern Med.* (2019) 34:1116–1124. doi: 10.3904/kjim.2018.011
9. Cao Y, Xiao G, Qiu X, Ye S, Lin T. Efficacy and safety of crizotinib among Chinese EML4-ALK-positive, advanced-stage non-small cell lung cancer patients. *PLoS ONE.* (2014) 9:e114008. doi: 10.1371/journal.pone.0114008
10. Xing P, Wang S, Hao X, Zhang T, Li J. Clinical data from the real world: efficacy of crizotinib in Chinese patients with advanced ALK-rearranged non-small cell lung cancer and brain metastases. *Oncotarget.* (2016) 7:84666–74. doi: 10.18632/oncotarget.13179
11. Yoshida T, Oya Y, Tanaka K, Shimizu J, Horio Y, Kuroda H, et al. Clinical impact of crizotinib on central nervous system progression in ALK-positive non-small lung cancer. *Lung Cancer.* (2016) 97:43–7. doi: 10.1016/j.lungcan.2016.04.006
12. Lei YY, Yang JJ, Zhong WZ, Chen HJ, Yan HH, Han JF, et al. Clinical efficacy of crizotinib in Chinese patients with ALK-positive non-small-cell lung cancer with brain metastases. *J Thorac Dis.* (2015) 7:1181–8. doi: 10.3978/j.issn.2072-1439.2015.06.04
13. Heuckmann JM, Balke-Want H, Malchers F, Peifer M, Sos ML, Koker M, et al. Differential protein stability and ALK inhibitor sensitivity of EML4-ALK fusion variants. *Clin Cancer Res.* (2012) 18:4682–90. doi: 10.1158/1078-0432.CCR-11-3260
14. Richards MW, Law EW, Rennalls LP, Busacca S, O'Regan L, Fry AM, et al. Crystal structure of EML1 reveals the basis for Hsp90 dependence of oncogenic EML4-ALK by disruption of an atypical beta-propeller domain. *Proc Natl Acad Sci USA.* (2014) 111:5195–200. doi: 10.1073/pnas.1322892111
15. Bayliss R, Choi J, Fennell DA, Fry AM, Richards MW. Molecular mechanisms that underpin EML4-ALK driven cancers and their response to targeted drugs. *Cell Mol Life Sci.* (2016) 73:1209–24. doi: 10.1007/s00018-015-2117-6
16. Woo CG, Seo S, Kim SW, Jang SJ, Park KS, Song JY, et al. Differential protein stability and clinical responses of EML4-ALK fusion variants to various ALK inhibitors in advanced ALK-rearranged non-small cell lung cancer. *Ann Oncol.* (2017) 28:791–7. doi: 10.1093/annonc/mdw693
17. Tan S, Gou Q, Pu W, Guo C, Yang Y, Wu K, et al. Circular RNA F-circEA produced from EML4-ALK fusion gene as a novel liquid biopsy biomarker for non-small cell lung cancer. *Cell Res.* (2018) 28:693–5. doi: 10.1038/s41422-018-0033-7
18. Yoshida T, Oya Y, Tanaka K, Shimizu J, Horio Y, Kuroda H, et al. Differential crizotinib response duration among ALK fusion variants in ALK-positive non-small-cell lung cancer. *J Clin Oncol.* (2016) 34:3383–9. doi: 10.1200/JCO.2015.65.8732
19. Li Y, Zhang T, Zhang J, Li W, Yuan P, Xing P, et al. Response to crizotinib in advanced ALK-rearranged non-small cell lung cancers with different ALK-fusion variants. *Lung Cancer.* (2018) 118:128–33. doi: 10.1016/j.lungcan.2018.01.026
20. Christopoulos P, Endris V, Bozorgmehr F, Elsayed M, Kirchner M, Ristau J, et al. EML4-ALK fusion variant V3 is a high-risk feature conferring accelerated metastatic spread, early treatment failure and worse overall survival in ALK(+) non-small cell lung cancer. *Int J Cancer.* (2018) 142:2589–98. doi: 10.1002/ijc.31275
21. Lei YY, Yang JJ, Zhang XC, Zhong WZ, Zhou Q, Tu HY, et al. Anaplastic lymphoma kinase variants and the percentage of ALK-positive tumor cells and the efficacy of crizotinib in advanced NSCLC. *Clin Lung Cancer.* (2016) 17:223–31. doi: 10.1016/j.cllc.2015.09.002
22. Mitiushkina NV, Tiurin VI, Iyevleva AG, Kholmatov MM, Filippova EA, Moiseyenko FV, et al. Variability in lung cancer response to ALK inhibitors cannot be explained by the diversity of ALK fusion variants. *Biochimie.* (2018) 154:19–24. doi: 10.1016/j.biochi.2018.07.018
23. Lin JJ, Zhu VW, Yoda S, Yeap BY, Schrock AB, Dagogo-Jack I, et al. Impact of EML4-ALK variant on resistance mechanisms and clinical outcomes in ALK-positive lung cancer. *J Clin Oncol.* (2018) 36:1199–206. doi: 10.1200/JCO.2017.76.2294
24. Ng KP, Hillmer AM, Chuah CT, Juan WC, Ko TK, Teo AS, et al. A common BIM deletion polymorphism mediates intrinsic resistance and inferior responses to tyrosine kinase inhibitors in cancer. *Nat Med.* (2012) 18:521–8. doi: 10.1038/nm.2713
25. Isobe K, Hata Y, Tochigi N, Kaburaki K, Kobayashi H, Makino T, et al. Clinical significance of BIM deletion polymorphism in non-small-cell lung cancer with epidermal growth factor receptor mutation. *J Thorac Oncol.* (2014) 9:483–7. doi: 10.1097/JTO.0000000000000125
26. Zhang L, Jiang T, Li X, Wang Y, Zhao C, Zhao S, et al. Clinical features of Bim deletion polymorphism and its relation with crizotinib primary resistance in Chinese patients with ALK/ROS1 fusion-positive non-small cell lung cancer. *Cancer.* (2017) 123:2927–35. doi: 10.1002/cncr.30677
27. Wu SG, Liu YN, Yu CJ, Yang PC, Shih JY. Association of BIM deletion polymorphism with intrinsic resistance to EGFR tyrosine kinase inhibitors in patients with lung adenocarcinoma. *JAMA Oncol.* (2016) 2:826–8. doi: 10.1001/jamaoncol.2016.0016
28. Oken MM, Creech RH, Tormey DC, Horton J, Davis TE, McFadden ET, et al. Toxicity and response criteria of the Eastern Cooperative Oncology Group. *Am J Clin Oncol.* (1982) 5:649–55. doi: 10.1097/00000421-198212000-00014
29. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer.* (2009) 45:228–47. doi: 10.1016/j.ejca.2008.10.026
30. Lin YT, Yu CJ, Yang JC, Shih JY. Anaplastic lymphoma kinase (ALK) kinase domain mutation following ALK inhibitor(s) failure in advanced ALK positive non-small-cell lung cancer: analysis and literature review. *Clin. Lung Cancer.* (2016) 17:e77–94. doi: 10.1016/j.cllc.2016.03.005
31. Lin YT, Chen CY, Shih JY. Real-world crizotinib use for anaplastic lymphoma kinase (ALK)-positive advanced non-small cell lung cancer under first-year national health insurance coverage in Taiwan. *Thorac Med.* (2018) 33:1–13.
32. Shaw AT, Kim DW, Nakagawa K, Seto T, Crino L, Ahn MJ, et al. Crizotinib versus chemotherapy in advanced ALK-positive lung cancer. *N Engl J Med.* (2013) 368:2385–94. doi: 10.1056/NEJMoa1214886
33. Zhou J, Zheng J, Zhang X, Zhao J, Zhu Y, Shen Q, et al. Crizotinib in patients with anaplastic lymphoma kinase-positive advanced non-small cell lung cancer versus chemotherapy as a first-line treatment. *BMC Cancer.* (2018) 18:10. doi: 10.1186/s12885-017-3720-8
34. Wu JY, Yu CJ, Yang CH, Wu SG, Chiu YH, Gow CH, et al. First- or second-line therapy with gefitinib produces equal survival in non-small cell lung cancer. *Am J Respir Crit Care Med.* (2008) 178:847–53. doi: 10.1164/rccm.200803-389OC

35. Stanley KE. Prognostic factors for survival in patients with inoperable lung cancer. *J Natl Cancer Inst.* (1980) 65:25–32.
36. Noronha V, Ramaswamy A, Patil VM, Joshi A, Chougule A, Kane S, et al. ALK positive lung cancer: clinical profile, practice and outcomes in a developing country. *PLoS ONE.* (2016) 11:e0160752. doi: 10.1371/journal.pone.0160752
37. Cha YJ, Kim HR, Shim HS. Clinical outcomes in ALK-rearranged lung adenocarcinomas according to ALK fusion variants. *J Transl Med.* (2016) 14:296. doi: 10.1186/s12967-016-1061-z
38. Bates JE, Milano MT. Prognostic significance of sites of extrathoracic metastasis in patients with non-small cell lung cancer. *J Thorac Dis.* (2017) 9:1903–10. doi: 10.21037/jtd.2017.06.117
39. Ko R, Kenmotsu H, Hisamatsu Y, Akamatsu H, Omori S, Nakashima K, et al. The effect of gefitinib in patients with postoperative recurrent non-small cell lung cancer harboring mutations of the epidermal growth factor receptor. *Int J Clin Oncol.* (2015) 20:668–73. doi: 10.1007/s10147-014-0761-8
40. Lee JK, Shin JY, Kim S, Lee S, Park C, Kim JY, et al. Primary resistance to epidermal growth factor receptor (EGFR) tyrosine kinase inhibitors (TKIs) in patients with non-small-cell lung cancer harboring TKI-sensitive EGFR mutations: an exploratory study. *Ann Oncol.* (2013) 24:2080–7. doi: 10.1093/annonc/mdt127

Conflict of Interest Statement: Y-TL has received speaking honoraria from AstraZeneca, Boehringer Ingelheim, Bristol-Myers Squibb, Pfizer, Roche, and TTY Biopharm; and travel expense from Pfizer. J-YS has received personal fees for advisory boards from AstraZeneca, Roche, Boehringer Ingelheim, Eli Lilly, Merck Sharp & Dohme, Ono Pharmaceutical, Chugai Pharmaceutical, AbbVie, and Bristol-Myers Squibb; speaking honoraria from AstraZeneca, Roche, Boehringer Ingelheim, Eli Lilly, Pfizer, Novartis, Merck Sharp & Dohme, Ono Pharmaceutical, Chugai Pharmaceutical, AbbVie, and Bristol-Myers Squibb; and travel expense from Roche, Boehringer Ingelheim, Pfizer, Merck Sharp & Dohme, Chugai Pharmaceutical, and Bristol-Myers Squibb.

The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

ORIGINAL RESEARCH
published: 01 October 2019
doi: 10.3389/fonc.2019.00975

Combining Plasma miRNAs and Computed Tomography Features to Differentiate the Nature of Pulmonary Nodules

Kexing Xi[1,2†], Weidong Wang[1,2†], Yingsheng Wen[1,2†], Yongqiang Chen[1,2], Xuewen Zhang[2,3], Yaobo Wu[4], Rusi Zhang[1,2], Gongming Wang[1,2], Zirui Huang[1,2] and Lanjun Zhang[1,2]*

[1] Department of Thoracic Surgery, Sun Yat-sen University Cancer Center, Guangzhou, China, [2] State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Guangzhou, China, [3] Department of Medical Oncology, Sun Yat-sen University Cancer Center, Guangzhou, China, [4] State Key Laboratory of Organ Failure Research, Guangdong Provincial Key Laboratory of Viral Hepatitis Research, Department of Infectious Diseases, Nanfang Hospital, Southern Medical University, Guangzhou, China

OPEN ACCESS

Edited by:
Umberto Malapelle,
University of Naples Federico II, Italy

Reviewed by:
Rabab Mohamed Gaafar,
Cairo University, Egypt
Francesco Pepe,
Department of Public Health,
University of Naples Federico II, Italy

***Correspondence:**
Lanjun Zhang
zhanglj@sysucc.org.cn

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Thoracic Oncology, a section of the journal Frontiers in Oncology

Received: 23 June 2019
Accepted: 13 September 2019
Published: 01 October 2019

Citation:
Xi K, Wang W, Wen Y, Chen Y, Zhang X, Wu Y, Zhang R, Wang G, Huang Z and Zhang L (2019) Combining Plasma miRNAs and Computed Tomography Features to Differentiate the Nature of Pulmonary Nodules. Front. Oncol. 9:975. doi: 10.3389/fonc.2019.00975

Objective: The purpose of this study was to evaluate the diagnostic efficiency of combining plasma microRNAs (miRNAs) and computed tomography (CT) features in the diagnosis of pulmonary nodules.

Methods: Ninety-two pulmonary nodule patients who had undergone surgery were enrolled in our study from July 2016 to March 2018 at the Sun Yat-sen University Cancer Center. A prediction model was established by combining 3 miRNAs (miRNA-146a, -200b, and -7) and CT features to identify the pulmonary nodules of these patients. We evaluated the diagnostic performance of this prediction model for pulmonary nodules using the Receiver Operating Characteristic (ROC) curve.

Results: The expression levels of miRNA-146a, -200b, and -7 in early-stage non-small cell lung cancer (NSCLC) patients are significantly higher than those in benign nodule patients. We used these three miRNAs and CT features (pleural indentation and speculation) to establish a prediction model for early-stage NSCLC, with a sensitivity and specificity of 92.9%, 83.3% in the training set, respectively. For the validation process, with the sensitivity of 71.8% and the specificity of 69.2%. For ROC curve analyses, area under the curve (AUC) for tumor identification in the training stage and validation stage were 0.929 and 0.781, respectively.

Conclusion: Plasma miRNA-146a, miRNA-200b, and miRNA-7 may be potential biomarkers for the early diagnosis of lung cancer. Our prediction model can help to identify the nature of pulmonary nodules with a relatively high diagnostic efficiency.

Keywords: microRNA, pulmonary nodules, prediction model, diagnosis, NSCLC

INTRODUCTION

Lung cancer is the most common cancer in the world with high morbidity and mortality based on the global cancer statistics from 2012 (1). In all, 80∼85% of lung cancer cases are non-small cell lung cancer (NSCLC) (2). To date, the 5-year overall survival of lung cancer is only approximately 15%, despite the advances of new diagnostic techniques and treatments over the last few decades (3). One of main reasons for the poor prognosis of lung cancer is that more than 60% of patients have advanced NSCLC at the time of diagnosis (4). The early diagnosis of lung cancer is vital to improve the overall survival rate and prognosis of lung cancer.

At present, the diagnosis of lung cancer clinically depends on a pathology examination. However, this method requires obtaining tissue from patients, which involves invasive procedures, such as percutaneous needle lung biopsy, bronchoscopy, and video-assisted thoracoscopic surgery (VATS). And a non-invasive test is more valuable, which can reduce the economic burden and surgical wound. Thanks to the increasing widespread use of computed tomographic (CT) scans, some early-stage lung cancer cases can be discovered. Moreover, based on the National Lung Screening Trial (NSLT), it was found that high-risk individuals who received low-dose computed tomographic (LDCT) screening had a 20% reduction in the mortality of lung cancer (5). However, some studies reported that screening CT had a high false positive rate which may lead to unnecessary treatment (5, 6). In the meantime, some tumor markers including Neuron-specific enolase (NSE), Carcinoembryonic antigen (CEA), Squamous cell carcinoma antigen (SCCA), and Cytokeratin-19 fragments (Cyfra21-1), have been widely used in the clinic. However, the specificity and sensitivity of these tumor markers for differentiating malignant or benign pulmonary nodules are limited (7). Furthermore, most of them were increased in advanced-stage lung cancer patients. Thus, it is urgent to find new methods and molecular biomarkers for detecting and diagnosing lung cancer.

MicroRNAs are endogenous noncoding RNAs of approximately−22nt in length (8). They participate in post-transcriptional gene regulation, and may act as potential molecular biomarkers (9). Numerous studies have reported the diagnostic value of miRNAs in pulmonary nodules and that they can serve as novel biomarkers for the early detection of lung cancer. According to the published literature and our previous study, we selected 10 miRNAs (miR-17, -146a, -200b, -182, -221, -205, -7, -21, -145, and miR-210) for our study (10–14). We aimed to measure the differences in expression of miRNAs between benign and malignant pulmonary nodules and then select the target miRNAs in this study. Finally, we combine miRNAs and CT features to diagnosis the nature of pulmonary nodules.

MATERIALS AND METHODS

Patients

In this retrospective study, all 92 patients were enrolled from July 2016 to March 2018 at the Sun Yat-sen University Cancer Center. Patients were collected based on the following eligibility criteria: (1) Patients were diagnosed with a pulmonary nodule by CT before the surgery; (2) Patients diagnosed as stage 0∼IIa non-small cell lung cancer (NSCLC) or benign lung disease pathologically after surgery; (3) Patients diagnosed with NSCLC can obtain accurate pathologic staging; and (4) Patients had complete clinicopathologic information. The exclusion criteria in this study were as follows: (1) Patients received radiotherapy or chemotherapy before the surgery; (2) Patients had a second primary tumor; (3) Patients were not pathologically confirmed after surgery; and (4) Patients with stage IIb∼ IV lung cancer. The patients selected for the training set and validation sets were according to the time of diagnosis. Patients were staged before the surgery according to the examination including brain MRI, thoracoabdominal CT, even positron emission tomography (PET). Patients were staged after the surgery according to the result of pathology. The pathologic staging of tumor was according to the International Association for the Study of Lung Cancer (8th version). This study was approved by the Ethics Committee of Sun Yat-sen University Cancer Center (NO. B2017-050) and written informed consent was obtained from all patients.

RNA Extraction and RT-PCR

Five milliliter of venous blood was collected from every patient before the surgery, centrifuged within 2 h at 3,200 rpm for 10 min at 4°C and then stored at −80°C until use.

Total RNA was extracted from 200 μl of plasma samples using miRNeasy Serum/Plasma Kit (Qiagen, USA) according to the manufacturer's instructions, and eluted to a final volume of 14 μl. The total reaction volume for Poly(A) tailing was 25 μl including eluted RNA, 10 μl; cel-miR-39, 1×10^9 copy; 5 × PAP buffer solution, 4 ul; PolyA polymerase (Life, 74225Y/Z), 2∼5U; and appropriate RNase-free water to the volume of 25 μl. The reaction conditions for Poly(A) tailing were 37°C for 10∼20 min, 65°C for 10 min. Adding a total of PolyA-tailed RNA, 10 μl; RT buffer solution, 2 μl; dNTPs, 2 μl; reverse primers, 20 μM; Omniscript (Qiagen Cat No.205111), 4U and appropriate RNase-free water to the Reverse transcription (RT) reactions, and the final volume is 20 μl. The conditions for reverse transcription (RT) reactions were incubated at 37°C for 1 h, 85°C for 5 min, and terminated at 4°C. The total volume for real-time PCR reaction was 20 μl and contained 1 μl cDNA, 10 μl 2 × SYBR Green Mix, 10 μM forward primers, 10 μM reverse primers and appropriate H_2O. The conditions for real-time PCR reaction were as follow: 95°C for 3 min; 40 cycles of 95°C for 15 s and 60°C for 35 s; 95°C for 15 s and 60°C for 1 min.

We used the Omniscript RT Kit (Qiagen, Germany) to accomplish reverse transcription (RT) reactions. The SYBR Green Mix (Qiagen, Cat No.208054) was used for real-time quantitative polymerase chain reaction (RT-PCR) analysis to detect the expression levels of miRNAs. The relative expression of miRNAs was analyzed with the $2^{-\Delta\Delta Ct}$ method (15). The cel-miR-39 was chosen as the inner control.

TABLE 1 | Clinical features of benign pulmonary nodule patients and NSCLC patients (n, %).

	Training set		Validation set	
Characteristic	**Benign (n = 12)**	**NSCLC (n = 28)**	**Benign (n = 13)**	**NSCLC (n = 39)**
Gender				
Men	4 (33.33)	15 (53.57)	9 (69.23)	17 (43.59)
Women	8 (66.67)	13 (46.43)	4 (30.77)	22 (56.41)
Age				
≤60	10 (83.33)	14 (50.00)	9 (69.23)	19 (48.72)
>60	2 (16.67)	14 (50.00)	4 (30.77)	20 (51.28)
Smoking index				
<400	10 (83.33)	20 (71.43)	8 (61.54)	28 (71.79)
≥400	2 (16.67)	8 (28.57)	5 (38.46)	11 (28.21)
pTNM stage				
0∼Ia		18 (64.29)		31 (79.49)
Ib		7 (25.00)		7 (17.95)
IIa		3 (10.71)		1 (2.56)
Differentiation grade				
Poor		7 (25.00)		6 (15.38)
Moderate		14 (50.00)		25 (64.10)
Well		7 (25.00)		8 (20.51)
Nodule diameter (cm)	1.90 ± 1.25[a]	2.51 ± 1.31[a]	1.40 ± 0.53[a]	1.83 ± 0.83[a]
speculation	0 (0.00)	14 (50.00)	4 (30.77)	21 (53.85)
Pleural indentation	1 (8.33)	13 (46.43)	1 (7.69)	16 (41.03)
Air bronchogram	0 (0.00)	3 (10.71)	2 (16.67)	5 (12.82)
Vessels sign	1 (8.33)	3 (10.71)	0 (0.00)	2 (5.13)

NSLCL, non-small cell lung cancer; TNM, tumor-node-metastasis. [a]Data are presented as mean ± SD.

TABLE 2 | Expression levels of 10 miRNAs between 28 NSCLC patients and 12 benign pulmonary nodule patients [mean ± SD and median (Q25–Q75)].

miRNAs	**NSCLC group**	**Benign group**	***P*-value**
miR-17	8.716 (8.054–9.281)	8.117 (7.341–8.390)	0.059[a]
miR-146a	8.371 (7.562–9.085)	7.628 (6.720–7.961)	0.021[a]
miR-200b	6.678 (5.482–7.585)	5.319 (3.916–5.817)	0.048[a]
miR-182	6.898±1.377	6.212±1.187	0.141[b]
miR-221	8.409±1.060	7.800±0.904	0.091[b]
miR-205	6.743 (6.235-8.008)	6.278 (4.866-6.891)	0.157[a]
miR-7	7.312 (6.745-8.346)	6.379 (5.635-6.731)	0.040[a]
miR-21	9.045 (8.165-9.705)	8.347 (7.732-8.909)	0.092[a]
miR-145	8.004±1.000	7.471±0.962	0.127[b]
miR-210	7.756 (6.536-8.569)	6.269 (5.558−7.267)	0.082[a]

NSCLC, non-small cell lung cancer. [a]P-value of Mann-Whitney U, [b]P-value of t-test.

TABLE 3 | The results of miRNAs and the prediction model in the diagnosis of pulmonary nodules in the training set.

Category	**AUC**	**Sensitivity**	**Specificity**	**Cutoff value**	**95%CI**	***P***
miR-146a	0.732	67.9%	91.7%	8.0528	0.561∼0.903	0.021
miR-200b	0.699	67.9%	83.3%	6.0335	0.503∼0.896	0.048
miR-7	0.707	75.0%	83.3%	6.7858	0.511∼0.903	0.040
Prediction model	0.929	92.9%	83.3%	0.4404	0.847∼1.000	<0.001

AUC, area under the curve; CI, confidence interval.

Statistical Analysis

Statistical analyses were performed using SPSS software version 21 (SPSS Inc., Chicago, IL). We used the Shapiro-Wilk test for the test for normal distribution. The Wilcoxon rank-sum test was used when the data were not normally distributed. When the data were normally distributed, we used the *t*-test for univariate analysis. The Mann-Whitney *U*-test and *t*-test were used to compare the difference of the expression of miRNAs between benign and malignant pulmonary nodule patients. The *t*-test, Mann-Whitney *U*-test, and Kruskal-Wallis test were used to evaluate the relationship between the expression levels of the plasma miRNAs and clinicopathologic characteristics. The risk score model was built using logistic regression. The Receiver Operating Characteristic (ROC) curve was used to evaluate the diagnostic performance of plasma miRNAs and the risk score model for distinguishing benign and malignant pulmonary nodules. Because of the magnitude of the miRNA levels measured, the results were log transformed for analysis when needed. *P*-values <0.05 was considered to indicate a statistically significant difference.

RESULTS

Patient Description

In all, 25 benign pulmonary nodule patients and 67 malignant pulmonary nodule patients were enrolled in this study (**Table 1**). The age of all participants ranged from 32 to 81 years. The median ages of benign pulmonary nodule patients and NSCLC patients were 57 and 61 years, respectively. In the training set, there were 12 participants in the benign group and 28 patients in the NSCLC group. In the validation set, 13 benign pulmonary nodule patients and 39 NSCLC patients were included. Among the NSCLC patients, including 1 patient of stage 0, 48 patients of stage Ia, 14 patients of stage Ib, and 4 patients of stage IIa. The clinical and imaging features of the patients are shown in **Table 1**.

miRNA Selection

We detected the levels of 10 miRNAs (miR-17, -146a, -200b, -182, -221, -205, -7, -21, -145, and miR-210) in 28 NSCLC patients and 12 benign subjects in the training stage. The results revealed that the levels of miR-146a, miR-200b, and miR-7 in NSCLC group were statistically higher than those in benign subjects. The *P*-values were 0.021, 0.048, 0.040, respectively. However, for the expression levels of the other seven miRNAs, there were no significant differences between these two groups. The details are shown in **Table 2**.

Establish a Prediction Model

In the first stage, we selected 3 miRNAs (miR-146a, miR-200b, and miR-7) from 10 miRNAs (miR-17, -146a, -200b, -182, -221, -205, -7, -21, -145, and miR-210) in 28 NSCLC patients and 12 benign subjects. Then, we used logistic regression analyses on variables, including miR-146a, miR-200b, miR-7, gender, age, smoking index, nodule diameter, speculation, pleural indentation, air bronchogram, and vessel sign, to establish a prediction model. The CT features were evaluated by two radiologists independently. A third radiologist made the final decision when the two radiologists disagreed. The following variables were enrolled into the prediction model, including miR-146a, miR-200b, miR-7, speculation, and pleural indentation. The prediction model was described by the following equations: $Y = 1.011 \times$ miR-146a$+0.907 \times$ miR-200b$-0.795 \times$ miR-7$+23.109 \times$ speculation$+4.291 \times$ pleural indentation-8.648.

Note: Smoking index is a parameter which is used to quantify cumulative smoking exposure. Usually, multiply the number of cigarettes smoked every day by the years smoked to get the Smoking index.

TABLE 4 | The results of miRNAs and the prediction model in the diagnosis of pulmonary nodules in the validation set.

Category	AUC	95%CI	*P*
miR-146a	0.696	0.530∼0.863	0.035
miR-200b	0.724	0.558∼0.889	0.016
miR-7	0.717	0.551∼0.883	0.020
Prediction model	0.781	0.636∼0.926	0.003

AUC, area under the curve; CI, confidence interval.

Diagnostic Value of the Prediction Model

We used the ROC curve to assess the efficacy of miRNAs and the prediction model for the diagnosis of pulmonary nodules. In the training set, the AUC, sensitivity, and specificity were 0.732, 67.9, and 91.7% for miR-146a; 0.699, 67.9, and 83.3% for miR-200b; 0.707, 75.0, 83.3% for miR-7, respectively. The AUC of the prediction model was 0.929, with a sensitivity of 92.9%, and a specificity of 83.3% using the optimal cutoff value of 0.4404 (**Table 3**). At this point, the sensitivity + specificity were considered to be maximal. In the validation stage, 13 benign patients and 39 NSCLC patients were included. The same cutoff value was used to determine the risk score for the cases. The ROC curve analysis of the prediction model allowed us to distinguish early-stage NSCLC from benign pulmonary nodule patients with 71.8% sensitivity and 69.2% specificity (AUC = 0.781), other results were shown in **Table 4**. The ROC curves of three miRNAs (miR-146a, miR-200b, and miR-7) and the prediction model were shown in **Figures 1**, **2**, respectively.

Association of the Expression of miR-146a, miR-200b, miR-7 With Clinical Characteristics

We explored the relationship between the expression of these three miRNAs and the patient clinical status in all NSCLC samples. As shown in **Table 5**, there are no significant connections between the expression of miR-146a, miR-200b, miR-7, and gender, age, smoking index, pTNM stage, or differentiation grade.

DISCUSSION

Lung cancer is the leading cause of cancer-related death in the world (1). The poor prognosis of lung cancer is mostly

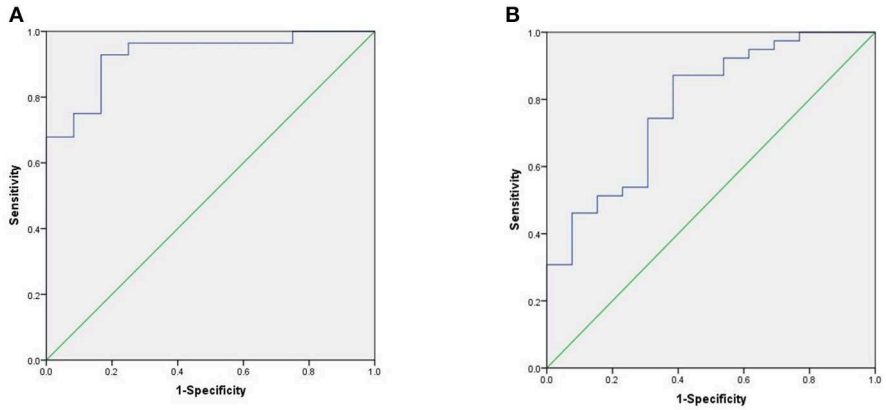


FIGURE 1 | ROC curve of the prediction model. **(A)** ROC curve of training stage (AUC = 0.929). **(B)** ROC curve of validation stage (AUC = 0.781).

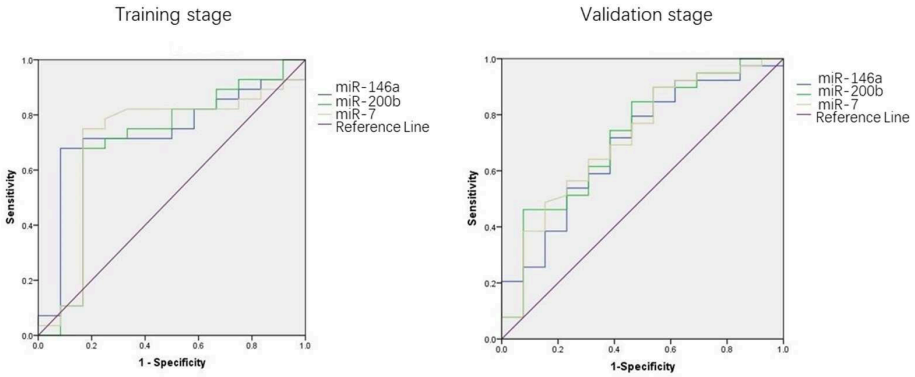


FIGURE 2 | ROC curves of three miRNAs.

due to the advanced stage of the disease. Improving the early diagnostic rate of lung cancer is very important. With the help of CT, more patients with lung nodules can be found. However, CT has a relatively high false positive rate, which may diagnose benign nodules as lung cancer and lead to unnecessary anxiety and surgery (5, 6, 16). It is imperative to find novel biomarkers for the auxiliary diagnosis of lung cancer together with CT. In this study, three miRNAs were enrolled into a model to predict the nature of pulmonary nodules. Current literatures have shown that miRNAs may serve as novel potential biomarkers for cancers (17, 18). Novel potential biomarkers like miRNAs was non-invasive method which was more valuable. Invasive operations, such as bronchoscopy and VATS not only increase the treatment costs but also the time of treatment, and some patients need to be hospitalized. What is more, invasive operations can bring patients surgical trauma. Compared with invasive test, a blood test/imaging is repeatable, easy sample management, cost-effectiveness, and non-invasive.

Plasma miRNAs have been reported to be involved in the occurrence and progression of lung cancer (19). Many studies have demonstrated the high diagnostic efficiency of miRNAs for early cancer detection (20–22). Based on previous studies, the diagnostic value of single miRNA for diagnosing pulmonary nodules was still poor (23). Thus, we combined of three miRNAs (miR-146a, miR-200b, and miR-7) for analysis in our study. Different from the early studies focused on a single or a few miRNAs (18, 24), the result of our study revealed that combining miRNAs and other clinical imaging features can obtain better diagnosis efficacy than those approaches using a single miRNA. A single miRNA only revealed some aspects of tumorigenesis, the combination of miRNAs constituted a better indicator for tumor occurrence and progression. And the diagnostic efficiency of a panel of miRNAs is usually better than that of a single miRNA (25). In our study, three out of 10 candidate miRNAs (miR-17, -146a, -200b, -182, -221, -205, -7, -21, -145, and miR-210) were found to increase significantly in NSCLC patients compared with those of benign individuals. Which demonstrated miR-146a, miR-200b, miR-7 may be the potential biomarkers for early-stage NSCLC. Nevertheless, the expressions of the other candidate miRNAs may be different significantly between NSCLC and benign patients when the

TABLE 5 | Relationship between clinicopathologic characteristics and miRNAs expressions in NSCLC.

Characteristics	Number	miR-146a	*P*	miR-200b	*P*	miR-7	*P*
Gender			0.620		0.425		0.990
Men	32	8.068 (7.122–8.545)[a]		6.249 (4.145–6.912)[a]		7.312 (5.603–7.525)[a]	
Women	35	8.137 (7.636–8.547)[a]		5.842 (4.592–6.465)[a]		7.230 (6.615–7.565)[a]	
Age			0.778		0.792		0.744
≤60	33	8.068 (7.194–8.525)[a]		6.049 (4.527–7.052)[a]		7.301 (5.926–7.782)[a]	
>60	34	8.117 (7.367–8.588)[a]		6.090 (4.503–6.521)[a]		7.268 (6.048–7.502)[a]	
Smoking index			0.906		0.372		0.513
<400	48	8.072 (7.246–8.659)[a]		5.622±1.435[b]		7.223 (5.891–7.525)[a]	
≥400	19	8.127 (7.490–8.504)[a]		5.985±1.623[b]		7.316 (6.726–7.531)[a]	
pTNM stage			0.408		0.420		0.821
0∼Ia	49	8.068 (7.105–8.552)[a]		6.037 (4.409–6.502)[a]		7.310 (5.630–7.518)[a]	
Ib/IIa	18	8.218 (7.367–8.662)[a]		6.276 (4.828–7.132)[a]		7.177 (6.102–7.808)[a]	
Differentiation grade			0.589		0.207		0.711
Poor	13	8.324 (8.029–8.506)[a]		6.193 (5.734–6.772)[a]		7.301 (7.078–7.535)[a]	
Moderate	39	8.017 (7.265–8.558)[a]		5.758 (4.462–6.465)[a]		7.230 (5.856–7.531)[a]	
Well	15	8.107 (6.763–8.598)[a]		6.360 (4.356–7.288)[a]		7.318 (5.365–8.037)[a]	

NSCLC, non-small cell lung cancer. [a]data are presented as median (Q25–Q75), [b]data are presented as mean ± SD.

sample size increased. We did not find the relationships between the expression of miR-146a, miR-200b, miR-7, and clinical characteristics in early-stage NSCLC. However, future studies may be needed to conduct to test other stage NSCLC to determine the relationships between the expression of above miRNAs and clinical characteristics.

Regarding the diagnosis of the lung nodules, most clinicians prefer to evaluate the pretest probability using clinical experience and judgment. Under this situation, the pretest probability is easily affected by subjectivity, and the result is not reliable. In this study, we established a prediction model with a mathematical formula, to distinguish benign and malignant pulmonary nodules, which is more reliable, measurable. Because a combination of molecular biological information and imaging features reflects various aspects of tumorigenesis. And we can get a risk score through the mathematical formula which contributes to decrease subjective. For imaging findings, the Swensen model and the VA model were used to diagnose the nature of lung nodules by some clinicians (26, 27). In the Swensen model, 6 variables of clinical and imaging features were predictors of malignancy, including age, smoking history, history of cancer, nodule diameter, upper lobe location, and speculation (26). In our prediction model, 2 variables of clinical and imaging features were included. We found that nodules with speculation and pleural indentation are more likely to be malignant, which was consistent with previous studies (26). Unlike many early studies focused on using clinical and imaging features to distinguish benign and malignant pulmonary nodules (28, 29), we added the new potential biomarker- miRNAs into our prediction model and achieved a relatively good discriminant result. Which can also provide the theoretical foundation for further exploring the role of combining miRNAs and, clinical and imaging features in the diagnosis of lung nodules.

Reviewing the overall existing literature of biomarkers for distinguishing NSCLC from benign pulmonary nodules, we may apply plasma miRNAs, ctDNA, circulating cancer cell (CTC), together with existing tumor biomarkers (such as CEA, NSE, Cyfra21-1) to the early diagnosis of lung cancer in clinic in the future.

More and more patients with pulmonary nodules are discovered due to the popularity of LDCT. The management of pulmonary nodules remains challenging (30), and evaluating the pretest probability of malignancy of pulmonary nodules is important. The result of the present study discloses that this model can diagnosis early-stage NSCLC with relatively high sensitivity and specificity, which will help to improve the strategies for pulmonary nodule management. In face of a patients with pulmonary nodules, we can choose different treatment programs, such as surgery, needle biopsy, or watchful waiting, according to the risk score of the model in the future. Furthermore, the model can be used for early detection of lung cancer, which can help to improve the prognosis of lung cancer.

This study has several limitations. Above all, this is a retrospective study with a small sample size, which may lead to potential bias. Furthermore, the proportion of benign nodules patients was relatively low, and this may can reduce diagnostic efficiency. Finally, the predictive model appears to be validated using a separate small set of patients from the same institution. Conducting a large-scale, multicentric and prospective clinical study will be more valuable in the future.

In conclusion, we found that Plasma miRNA-146a, miRNA-200b, and miRNA-7 may be the potential biomarkers for the early diagnosis of lung cancer. Our prediction model can help to identify the nature of pulmonary nodules with a relatively high diagnostic efficiency. Which will serve for the detection for early-stage lung cancer and improve the management of pulmonary nodules.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the manuscript/supplementary files.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of Sun Yat-sen University Cancer Center (No. B2017-050). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

LZ and KX: conception and design. KX, WW, YW, GW, and ZH: provision of study materials or patients. KX, YC, XZ, YWu, and RZ: collection and assembly of data. KX, WW, and YW: data analysis and interpretation. All authors: manuscript writing and final approval of manuscript.

FUNDING

This study was funded by the National Key Research and Development Program, China (No. 2016YFC0905400).

REFERENCES

1. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA Cancer J Clin.* (2015) 65:87–108. doi: 10.3322/caac.21262
2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2015. *CA Cancer J Clin.* (2015) 65:5–29. doi: 10.3322/caac.21254
3. Wu YL, Zhou Q. Clinical trials and biomarker research on lung cancer in China. *Exp Opin Therapeut Targets.* (2012) 16 (Suppl. 1):S45–50. doi: 10.1517/14728222.2011.630663
4. Patz EF Jr, Goodman PC, Bepler G. Screening for lung cancer. *N Eng J Med.* (2000) 343:1627–33. doi: 10.1056/NEJM200011303432208
5. National Lung Screening Trial Research Team, Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med.* (2011) 365:395–409. doi: 10.1056/NEJMoa1102873
6. Barba M, Felsani A, Rinaldi M, Giunta S, Malorni W, Paggi MG. Reducing the risk of overdiagnosis in lung cancer: a support from molecular biology. *J Cell Physiol.* (2011) 226:2213–4. doi: 10.1002/jcp.22558
7. Nisman B, Lafair J, Heching N, Lyass O, Baras M, Peretz T, et al. Evaluation of tissue polypeptide specific antigen, CYFRA 21-1, and carcinoembryonic antigen in nonsmall cell lung carcinoma: does the combined use of cytokeratin markers give any additional information? *Cancer.* (1998) 82:1850–9. doi: 10.1002/(SICI)1097-0142(19980515)82:10<1850::AID-CNCR6>3.0.CO;2-R
8. Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell.* (2004) 116:281–97. doi: 10.1016/S0092-8674(04)00045-5
9. Hayes J, Peruzzi PP, Lawler S. MicroRNAs in cancer: biomarkers, functions and therapy. *Trends Mol Med.* (2014) 20:460–9. doi: 10.1016/j.molmed.2014.06.005
10. Zheng D, Haddadin S, Wang Y, Gu LQ, Perry MC, Freter CE, et al. Plasma microRNAs as novel biomarkers for early detection of lung cancer. *Int J Clin Exp Pathol.* (2011) 4:575–86.
11. Shen J, Todd NW, Zhang H, Yu L, Lingxiao X, Mei Y, et al. Plasma microRNAs as potential biomarkers for non-small-cell lung cancer. *Lab Invest.* (2011) 91:579–87. doi: 10.1038/labinvest.2010.194
12. Chen X, Ba Y, Ma L, Cai X, Yin Y, Wang K, et al. Characterization of microRNAs in serum: a novel class of biomarkers for diagnosis of cancer and other diseases. *Cell Res.* (2008) 18:997–1006. doi: 10.1038/cr.2008.282
13. Wang RJ, Zheng YH, Wang P, Zhang JZ. Serum miR-125a-5p, miR-145 and miR-146a as diagnostic biomarkers in non-small cell lung cancer. *Int J Clin Exp Pathol.* (2015) 8:765–71.
14. Xi KX, Zhang XW, Yu XY, Wang WD, Xi KX, Chen YQ, et al. The role of plasma miRNAs in the diagnosis of pulmonary nodules. *J Thoracic Dis.* (2018) 10:4032–41. doi: 10.21037/jtd.2018.06.106
15. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods.* (2001) 25:402–8. doi: 10.1006/meth.2001.1262
16. Wieskopf B, Demangeat C, Purohit A, Stenger R, Gries P, Kreisman H, et al. Cyfra 21-1 as a biologic marker of non-small cell lung cancer. Evaluation of sensitivity, specificity, and prognostic role. *Chest.* (1995) 108:163–9. doi: 10.1378/chest.108.1.163
17. Mitchell PS, Parkin RK, Kroh EM, Fritz BR, Wyman SK, Pogosova-Agadjanyan EL, et al. Circulating microRNAs as stable blood-based markers for cancer detection. *Proc Natl Acad Sci USA.* (2008) 105:10513–8. doi: 10.1073/pnas.0804549105
18. Ng EK, Chong WW, Jin H, Lam EK, Shin VY, Yu J, et al. Differential expression of microRNAs in plasma of patients with colorectal cancer: a potential marker for colorectal cancer screening. *Gut.* (2009) 58:1375–81. doi: 10.1136/gut.2008.167817
19. Yanaihara N, Caplen N, Bowman E, Seike M, Kumamoto K, Yi M, et al. Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. *Cancer Cell.* (2006) 9:189–98. doi: 10.1016/j.ccr.2006.01.025
20. Di Leva G, Garofalo M, Croce CM. MicroRNAs in cancer. *Ann Rev Pathol.* (2014) 9:287–314. doi: 10.1146/annurev-pathol-012513-104715
21. Wozniak MB, Scelo G, Muller DC, Mukeria A, Zaridze D, Brennan P. Circulating MicroRNAs as non-invasive biomarkers for early detection of non-small-cell lung cancer. *PLoS ONE.* (2015) 10:e0125026. doi: 10.1371/journal.pone.0125026
22. Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, Peck D, et al. MicroRNA expression profiles classify human cancers. *Nature.* (2005) 435:834–8. doi: 10.1038/nature03702
23. Yu L, Todd NW, Xing L, Xie Y, Zhang H, Liu Z, et al. Early detection of lung adenocarcinoma in sputum by a panel of microRNA markers. *Int J Cancer.* (2010) 127:2870–8. doi: 10.1002/ijc.25289
24. Lawrie CH, Gal S, Dunlop HM, Pushkaran B, Liggins AP, Pulford K, et al. Detection of elevated levels of tumour-associated microRNAs in serum of patients with diffuse large B-cell lymphoma. *Br J Haematol.* (2008) 141:672–5. doi: 10.1111/j.1365-2141.2008.07077.x
25. Chen X, Hu Z, Wang W, Ba Y, Ma L, Zhang C, et al. Identification of ten serum microRNAs from a genome-wide serum microRNA expression profile as novel noninvasive biomarkers for nonsmall cell lung cancer diagnosis. *Int J Cancer.* (2012) 130:1620–8. doi: 10.1002/ijc.26177
26. Swensen SJ, Silverstein MD, Ilstrup DM, Schleck CD, Edell ES. The probability of malignancy in solitary pulmonary nodules. Application to small radiologically indeterminate nodules. *Arch Int Med.* (1997) 157:849–55. doi: 10.1001/archinte.157.8.849
27. Gould MK, Ananth L, Barnett PG, Veterans Affairs SCSG. A clinical model to estimate the pretest probability of lung cancer in patients with solitary pulmonary nodules. *Chest.* (2007) 131:383–8. doi: 10.1378/chest.06-1261
28. Herder GJ, van Tinteren H, Golding RP, Kostense PJ, Comans EF, Smit EF, et al. Clinical prediction model to characterize pulmonary nodules: validation

and added value of 18F-fluorodeoxyglucose positron emission tomography. *Chest*. (2005) 128:2490–6. doi: 10.1378/chest.128.4.2490

29. Gould MK, Donington J, Lynch WR, Mazzone PJ, Midthun DE, Naidich DP, et al. Evaluation of individuals with pulmonary nodules: when is it lung cancer? Diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. *Chest*. (2013) 143(Suppl. 5):e93S−120S. doi: 10.1378/chest.12-2351
30. Ost D, Fein AM, Feinsilver SH. Clinical practice. The solitary pulmonary nodule. *N Engl J Med*. (2003) 348:2535–42. doi: 10.1056/NEJMcp012290

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

ORIGINAL RESEARCH
published: 15 October 2019
doi: 10.3389/fonc.2019.01011

Integrated Network Analysis Reveals FOXM1 and MYBL2 as Key Regulators of Cell Proliferation in Non-small Cell Lung Cancer

Firoz Ahmed [1,2*]

[1] Department of Biochemistry, University of Jeddah, Jeddah, Saudi Arabia, [2] University of Jeddah Center for Scientific and Medical Research, University of Jeddah, Jeddah, Saudi Arabia

Background: Loss of control on cell division is an important factor for the development of non-small cell lung cancer (NSCLC), however, its molecular mechanism and gene regulatory network are not clearly understood. This study utilized the systems bioinformatics approach to reveal the "driver-network" involve in tumorigenic processes in NSCLC.

Methods: A meta-analysis of gene expression data of NSCLC was integrated with protein-protein interaction (PPI) data to construct an *NSCLC network*. MCODE and iRegulone were used to identify the local clusters and its upstream transcription regulators involve in NSCLC. Pair-wise gene expression correlation was performed using GEPIA. The survival analysis was performed by the Kaplan-Meier plot.

Results: This study identified a local "driver-network" with highest MCODE score having 26 up-regulated genes involved in the process of cell proliferation in NSCLC. Interestingly, the "driver-network" is under the regulation of TFs FOXM1 and MYBL2 as well as miRNAs. Furthermore, the overexpression of member genes in "driver-network" and the TFs are associated with poor overall survival (OS) in NSCLC patients.

Conclusion: This study identified a local "driver-network" and its upstream regulators responsible for the cell proliferation in NSCLC, which could be promising biomarkers and therapeutic targets for NSCLC treatment.

Keywords: non-small cell lung cancer, gene expression, meta-analysis, systems bioinformatics, gene network

OPEN ACCESS

Edited by:
Umberto Malapelle, University of Naples Federico II, Italy

Reviewed by:
Francesco Pepe, University of Naples Federico II, Italy
Francesco Passiglia, Palermo University Hospital, Italy

***Correspondence:**
Firoz Ahmed
fahmed1@uj.edu.sa

Specialty section:
This article was submitted to Thoracic Oncology, a section of the journal Frontiers in Oncology

Received: *26 June 2019*
Accepted: *20 September 2019*
Published: *15 October 2019*

Citation:
Ahmed F (2019) Integrated Network Analysis Reveals FOXM1 and MYBL2 as Key Regulators of Cell Proliferation in Non-small Cell Lung Cancer. Front. Oncol. 9:1011. doi: 10.3389/fonc.2019.01011

INTRODUCTION

Lung cancer is one of the most commonly diagnosed cancer with high mortality around the world (1). The global prevalence of lung cancer and mortality rate is rising at an alarming pace with an estimated number of newly diagnosed lung cancer was 2.1 million while the number of deaths was 1.8 million in 2018 (https://gco.iarc.fr). Based upon histology, lung cancer is divided into two classes: (i) Non-small cell lung cancer (NSCLC) which represents approximately 85–90% of all lung cancer, and (ii) Small-cell lung cancer (SCLC) which represents approximately 10–15% of the lung cancer (1). NSCLC has three major sub-classes including (a) lung squamous cell carcinoma (LUSC), (b) lung adenocarcinoma (LUAD), and (c) large cell carcinoma. However, due to lack of clinical symptom and effective diagnostic screening, the NSCLC is generally diagnosed at an advanced stage. The 5-year overall

survival rate of metastatic NSCLC is 6% and has not been significantly improved in spite of having advancement in surgery, chemotherapy, and radiation (https://www.cancer.org).

Molecular profiling of NSCLC identified mutations in the tumor suppressor genes (TP53, RB1), oncogenes (EGFR, KRAS, AKT, MAPK) and translocations in oncogenes (ALK, RET, ROS1, NTRK1, NRG1), which alter the important signal-transduction pathways (2, 3). EGFR mutants have been reported more frequently in NSCLC in nonsmokers Asians patients and showed highly sensitive to therapy with EGFR tyrosine kinase inhibitors such as gefitinib and erlotinib (4, 5). Similarly, ALK rearranged gene fusion was also highly reported in NSCLC and has been proven more effective treatment with ALK-targeted inhibitors (crizotinib and alectinib) (1, 6). The genomic mutations not only alter the protein structure but also affect the expression level of genes involved in the cell division resulting in uncontrolled cell proliferation, cell survival, and NSCLC. Previous studies mainly focused on understanding the alteration in gene expression in NSCLC tumors (7), and identified overexpressed genes including CDC20 (8), CCNB1 (9), ASPM (10), and KIF4A (11), which contributes to the proliferation of tumor cells and also associated with poor prognosis. Furthermore, the role of transcription factors (TFs) including MYBL2 (12), FOXM1 (13–15), and E2F4 (16) in cell proliferation and cell survival in NSCLC has been reported. The miRNAs, a class of small non-coding RNAs which regulate gene expression at the post-transcriptional level through binding to 3′UTR of mRNA (17, 18), are also emerging as promising biomarkers for detecting NSCLC (19). However, the molecular mechanism and regulatory network of the mRNAs, TFs, miRNAs, and proteins underlying dysregulated cell division and cell proliferation in NSCLC are still largely remain unclear. Addressing these challenges are most pivotal for developing anticancer drugs and diagnostic and prognostic biomarkers for better management and personalized treatment of NSCLC.

The emergence of high-throughput genomics, transcriptomics, proteomics, and interactome data and their integrative analysis opens a new avenue for a deep understanding of etiology of cancer (20, 21). This work is focused on applying a systems bioinformatics approach to uncover interaction and regulatory mechanism of mRNAs, TFs, miRNAs, and proteins underlying cell proliferation and progression of NSCLC. Gene expression profiles have been integrated to identify the high confidence up- and down-regulated genes in the NSCLC compared to adjacent non-tumor tissues. Moreover, using the transcriptome-interactome data, *NSCLC network* was constructed and analyzed to understand the molecular mechanism underlying the development and proliferation of NSCLC. Our analysis revealed one important "driver-network" consists of 26 genes and its upstream regulators TFs FOXM1 and MYBL2 whose overexpression are associated with dysregulation of cell cycle and enhance cell proliferation in NSCLC. Furthermore, NSCLC associated miRNAs regulating the genes of "driver-network" were also identified. Combination of genes in the "driver-network" and upstream regulators could be potential biomarkers for diagnosis and prognosis; and therapeutic targets for better treatment of NSCLC.

MATERIALS AND METHODS

Gene Expression Data Collection

In February 2019, microarray gene expression data were searched in Gene Expression Omnibus database (GEO: www.ncbi.nlm.nih.gov/geo/) using following criteria: (a) Lung cancer; (b) Human; and (c) Expression profiling by array; which gave 304 unique GEO series (GSEs). Then, a careful manually selected the GSEs data using following criteria: (d) Each GSE must have the profile of NSCLC along with adjacent non-tumor tissues as a control; (e) Each group (NSCLC/control) must have more than 20 samples; (f) All GSEs are from same microarray platform. Based upon the above criteria, three GSEs data [GSE27262 (22, 23), GSE18842 (7), and GSE19804 (24)] were selected and downloaded for further study (**Table S1**).

Identification of Differentially Expressed Genes (DEGs)

This study analyzed the gene expression data having 131 samples from NSCLC and 130 samples from adjacent non-tumor tissues as control (Normal). Preprocessing of each microarray raw data including background correction, normalization and log2 transformation were performed separately with RMA of Oligo package version 1.46 in Bioconductor/R version 3.5.2 (25). Each normalized expression data was integrated into a single file and batch effects were removed with ComBat of sva package version 3.30 in R (26). After that, differential expression analysis of genes between NSCLC compared to control was calculated using the linear modeling features of the limma package version 3.38 in Bioconductor/R (27). Affymetrix probe set ids were mapped to gene symbol using DAVID 6.8 (https://david.ncifcrf.gov/) (28). The gene is considered as differentially expressed (DEGs) if log2 Fold Change $|\log2FC|$ is >2 and adjusted P-value is < 0.001. If multiple probe id mapped with the same gene, probe id with highest log2FC were selected. The expression data of the significant DEGs were selected and transformed into Z-score (row-wise of value), then a hierarchical clustering across rows were performed to create a heatmap using Morpheus tool (https://software.broadinstitute.org/morpheus/).

Functional Annotation and Pathway Enrichment Analysis

In order to investigate the biological processes altered in NSCLC, we performed the functional annotation including Gene Ontology (GO) enrichment analysis for Biological Process, Molecular Function, Cellular Component, and Kyoto Encyclopedia of Genes and Genomes (KEGG: www.kegg.jp) to the list of DEGs. All these functional annotations were performed with clusterProfiler v3.10.1 in Bioconductor/R using pvalueCutoff = 0.01, pAdjustMethod = "BH," qvalueCutoff = 0.05, minGSSize = 5 (29).

Construction and Analysis of the NSCLC Network

To construct the *NSCLC network*, DEGs were mapped to the STRING version 11 application (30). The setting parameters of STRING were: (a) meaning of network edges (confidence);

(b) active interaction sources (selected all); (c) minimum required interaction score (high confidence >0.900); (d) max number of interactors to show (1st shell, none/query protein only); (2nd shell, none). The PPI data was downloaded, and the network was visualized in Cytoscape software version (3.5.1), where each node represents a protein while the edge represents an interaction between proteins (31). We also integrated the information of differential gene expression level into the network, where the red node indicates up-regulated, while the green node indicates down-regulated expression in NSCLC compared to control. The topological structure of the *NSCLC network* was analyzed using Cytoscape plug-in "NetworkAnalyzer."

Intending to identify highly connected local sub-networks in the *NSCLC network*, we applied the Cytoscape plug-in MCODE clustering algorithm (32). Furthermore, the biological relevance of these modules was analyzed with GO and KEGG pathway.

Identifying the Upstream Regulator of Genes in Cluster

To identify the upstream transcription regulators of genes in MCODE clusters, Cytoscape plug-in iRegulone (V 1.3) was used at default parameters (33). Then, in house program was used to generate a matrix table where each row indicates TF while a column indicates a target gene. A Venn diagram was drawn using (http://www.interactivenn.net/).

Expression Correlation Between Genes and TFs

To establish the relationship between genes in a cluster and its upstream regulators, pair-wise gene expression correlation was performed with GEPIA (http://gepia.cancer-pku.cn/index.html) (34). The web server integrated RNA-seq expression data from 9,736 tumors and 8,587 normal samples of the Cancer Genome Atlas (TCGA) and the Genotype-Tissue Expression (GTEx) projects. The analysis was done on default parameters: Pearson correlation coefficient; and selected LUAD and LUSC as TCGA tumor and TCGA normal dataset.

Effect of Signature Genes on Survival in NSCLC by Kaplan-Meier Plot

The potential effect of expression of relevant genes on the overall survival (OS) was analyzed on the lung cancer patient. An online KM plotter software (http://kmplot.com/analysis/) was used to generate the Kaplan-Meier Plot on 1926 NSCLC cancer patients (LUAD and LUSC) (35). The tool run on the default parameters on hazard ratio (HR) with 95% confidence intervals and log-rank *P*-value which is considered as significant *P*-value < 0.05. The biased arrays ($n = 2{,}435$) were excluded for quality control.

Extension of Cluster 1 With miRNA

In order to understand the regulatory role of miRNAs in NSCLC, the differentially expressed miRNAs (DEMs) in NSCLC compared to normal was downloaded from miRCancer database (http://mircancer.ecu.edu/) (36). After removing redundancy and cleaning of the data, 56 miRNAs were appeared as up-regulated, while 168 miRNAs were appeared as down-regulated in NSCLC compared to control. The targets of these DEMs were identified using miRNet tool (https://www.mirnet.ca/). After that, only those miRNAs were selected for further study which are targeting any of the 31 genes (*Cluster 1* and its associated TFs). Finally, an extended sub-network of *Cluster 1* was generated by integrating *Cluster 1* with its upstream regulators of TFs and miRNAs in Cytoscape. To make sparse visualization of network, interaction within *Cluster 1* as well as between *Cluster 1* and TFs were removed.

Mutational Signatures in NSCLC

The cBio Cancer Genomic Portal (http://cbioportal.org) is a freely available tool to explore cancer genomic data in diverse cancers. We selected the NSCLC from TCGA database and submitted the list of 31 genes from *Cluster 1* and its associated TFs in cBioPortal.

RESULTS

Verification of Each Group of Samples Using Principal Component Analysis

The Principal Component Analysis (PCA) was performed on normalized data of gene expression, which revealed a clear difference between NSCLC and normal samples in each GSEs study (**Figures S1A–C**). The cumulative contribution of PC1, PC2, and PC3 is 38.59, 33.64, and 36.47% for GSE27262, GSE18842, and GSE19804 datasets, respectively. In order to increase the statistical power to discover the DEGs, the expression data of three GSEs was integrated

TABLE 1 | List of top 20 differentially expressed genes is NSCLC.

Up-regulated			Down-regulated		
Gene	**Log2FC**	**Adj. *p*-value**	**Gene**	**Log2FC**	**Adj. *p*-value**
SPP1	4.70	4.49E-66	AGER	−5.06	2.18E-89
COL11A1	4.27	3.81E-51	CLDN18	−5.05	4.65E-54
COL10A1	4.10	1.62E-67	SFTPC	−4.53	9.44E-41
MMP12	3.95	3.41E-47	GPM6A	−4.50	6.61E-89
MMP1	3.86	1.32E-35	ADH1B	−4.34	5.76E-51
GREM1	3.71	3.70E-43	FABP4	−4.30	8.44E-65
HS6ST2	3.56	4.04E-54	TMEM100	−4.28	8.16E-57
GJB2	3.39	3.21E-46	CLIC5	−4.15	2.31E-69
CTHRC1	3.38	6.08E-62	CA4	−4.13	1.45E-85
TOP2A	3.36	4.37E-66	FAM107A	−4.12	2.82E-77
ANLN	3.25	2.25E-59	WIF1	−4.03	1.35E-40
COL1A1	3.13	9.35E-49	FCN3	−4.02	2.07E-62
PSAT1	3.05	4.96E-59	GKN2	−3.90	1.01E-56
TMPRSS4	3.00	8.65E-52	STXBP6	−3.88	9.95E-66
SPINK1	2.93	1.21E-20	CD36	−3.88	2.41E-64
CDCA7	2.90	2.04E-56	Mt1m	−3.87	2.33E-50
CST1	2.90	6.08E-42	AQP4	−3.76	2.70E-37
CXCL14	2.87	5.54E-31	SFTPA1	−3.70	2.40E-28
CEACAM5	2.83	2.54E-23	cpb2	−3.69	1.29E-48
RRM2	2.76	8.65E-52	TNNC1	−3.69	4.28E-84

and technical variability and noise were removed using batch effect correction. The PCA analysis clearly showed the distinction between cancer and normal samples indicating successful removal of the batch effects on the GSEs microarray data (**Figures S1D,E**).

Identification of DEGs

To identify the genes contributing to the NSCLC, differential expression analysis was conducted on gene expression data. A total of 346 DEGs including 97 up-regulated and 249 down-regulated genes were identified with $|\log2FC| > 2$ and adjusted *P*-value is < 0.001. Among DEGs, top 10 genes showing up-regulated expression are SPP1, COL11A1, COL10A1, MMP12, MMP1, GREM1, HS6ST2, GJB2, CTHRC1, and TOP2A; while top 10 genes showing down-regulated expression are AGER, CLDN18, SFTPC, GPM6A, ADH1B, FABP4, TMEM100, CLIC5, CA4, and FAM107A (**Table 1**). Detail information and the complete list of DEGs is provided in **Table S2A** for up-regulated and **Table S2B** for down-regulated genes. A hierarchical cluster heatmap of DEGs across biological samples reveals distinct patterns of gene expressions in NSCLC compared to normal (**Figure S2**).

Functional Annotation and Pathway Enrichment Analysis

To understand the biological function and key pathways altered in NSCLC, function annotation and pathway enrichment analysis was performed for the list of up- and down-regulated genes. Biological process (BP) and Molecular Function (MF) of Gene Ontology analysis revealed that the up-regulated genes are primarily associated with nuclear division, organelle fission, chromosome segregation, regulation of mitotic nuclear division, metaphase/anaphase transition of cell cycle, and mitotic spindle assembly checkpoint, and microtubule binding (**Figure 1A**). The cellular components (CC) of up-regulated genes were significantly associated with spindle, chromosome region, kinetochore, microtubule, and midbody, fibrillar collagen trimer, and spindle microtubule (**Figure 1A**). The KEGG pathway analysis showed the up-regulated genes were significantly enriched in only Cell cycle-G2/M transition (**Figure 1A**).

Biological process (BP) and Molecular Function (MF) of Gene Ontology analysis revealed that the down-regulated genes are primarily associated with circulatory system process, leukocytes migration, cell-substrate adhesion, regulation of angiogenesis, receptor-mediated endocytosis, cell chemotaxis, regulation of cell junction assembly, amide binding, peptide binding, and cytokine binding (**Figure 1B**). The cellular components (CC) of down-regulated genes were significantly associated with the extracellular matrix, membrane raft, cell-cell junction, and collagen-containing extracellular matrix (**Figure 1B**). The KEGG pathway analysis showed the down-regulated genes were significantly enriched in only BMP signaling (**Figure 1B**). The complete results of GO and KEGG analyses could be found in **Table S3A** for up-regulated and **Table S3B** for down-regulated genes.

Construction and Analysis of NSCLC Network

Mapping of DEGs on STRING gave PPI network with 151 nodes and 640 edges, which were visualized in Cytoscape software, where each node represents a protein while an edge represents an interaction between proteins. The gene expression level of each protein was integrated into the PPI network, where the red node indicates up-regulated, while the blue node indicates down-regulated gene expression level in NSCLC compared to normal and termed as *NSCLC network* (**Figure 2**). Size of the node is based upon the degree of connectivity of the node. Edges in the network represent direct interactions between nodes. As shown in **Figure 2**, there are 61 and 86 nodes in the network showing up-regulation and down-regulation, respectively; while 4 nodes are not having gene expression level (identified by PPI interaction and not in the list of our DEGs). A highly interconnected sub-network of overexpressed genes could be seen in the NSCLC network. The structural topological of *NSCLC network* including Betweenness Centrality, Closeness Centrality, Clustering Coefficient, and Degree were analyzed and presented in **Table S4**. Furthermore, highly inter-connected 15 sub-network clusters were extracted from *NSCLC network* using Cytoscape plug-in MCODE (**Figure S3**; **Table 2**). Among them, top five clusters with the highest MCODE score were considered for further study. Topologically relevant information of a gene is given as follows:

Hub genes: The highly connected gene in the network is called hub gene. The node CDC20 has the highest degree of connectivity [35] in the *NSCLC network*. Other top-five hub nodes with their degree of connectivity are BUB1 [33], CDK1 [33], UBE2C [32], CCNB1 [31], and CCNB2 [31] (**Table S4**). It is interesting to note that all 26 nodes of *NSCLC network* act as intramodular hubs of *Cluster 1*. Therefore, we considered all genes in *Cluster 1* as hub genes as their degree of connectivity are more than 21 (**Figure S3**; **Table S4**).

Betweenness centrality of Node: The node RHOJ has the highest betweenness centrality of 1, which connects DLC1, ARHGEF26, and ARHGAP6 (**Figure 2**; **Table S4**). Node Interleukin-6 (IL6) has second-highest betweenness centrality of 0.611 in the *NSCLC network,* which connects 6 proteins across *three* sub-networks: *Cluster 1* (UBE2C); *Cluster 2* (CXCL2); and *Cluster 5* (SPP1, CP, GOLM1, CHRDL1) (**Figure 2**; **Table S4**).

Top DEGs in Clusters: It was found that the highest up-regulated gene are TOP2A (*Cluster 1*), CXCL13 (*Cluster 2*), COL11A1 (*Cluster 3*), and SPP1 (*Cluster 5*); while highest down-regulated genes are PPBP (*Cluster 2*), COL6A6 (*Cluster 3*), EDNRB (*Cluster 4*), and IL6 (*Cluster 5*).

In order to understand the functional relevance, these clusters were further analyzed using GO and pathways enrichments. The *Cluster 1* consist of 26 up-regulated gene in *NSCLC network* (**Table 2**). Functional annotation indicates that: (a) *Cluster 1* is significantly associated with nuclear division, spindle, microtubule binding, and protein serine/threonine kinase activity (**Figure S4**; **Table S5**); (b) *Cluster 2* is significantly associated with leukocyte migration, cell chemotaxis, G

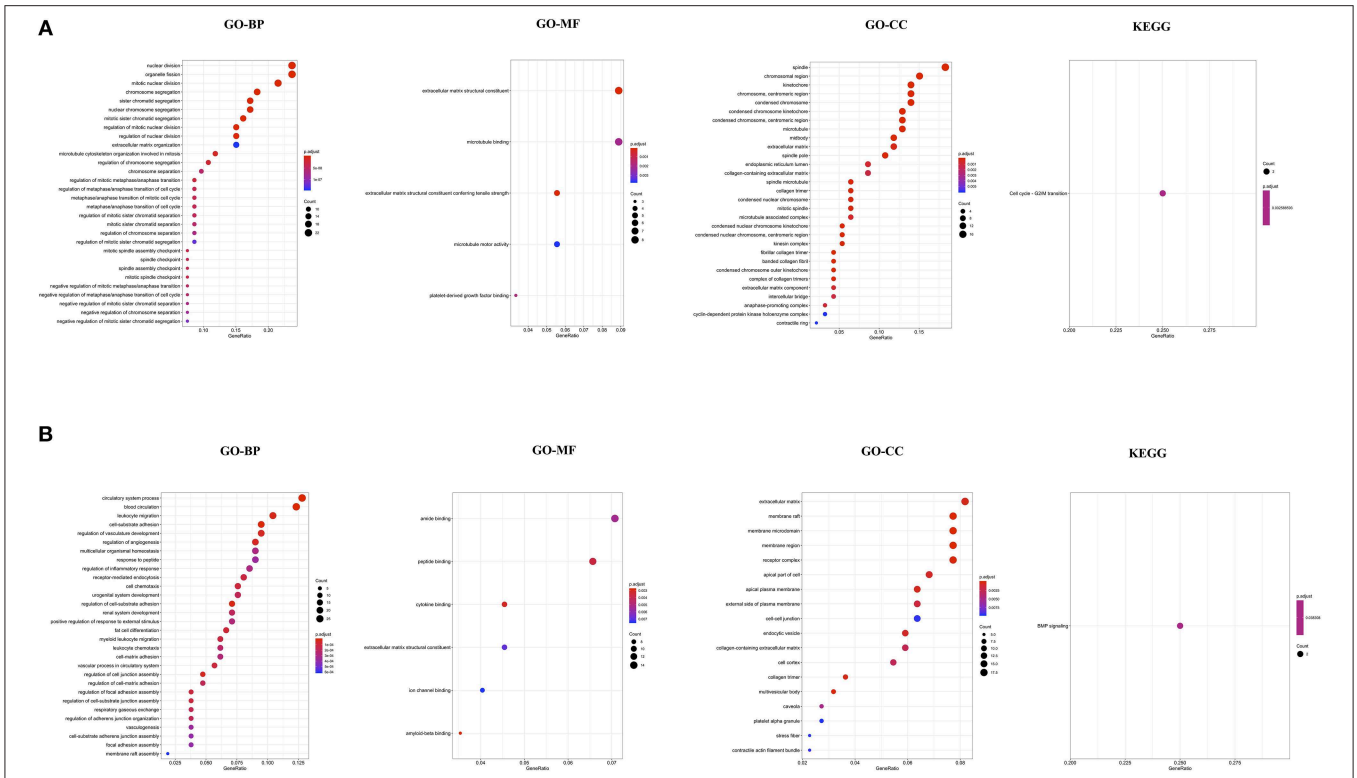


FIGURE 1 | Functional annotation of up-regulated genes in NSCLC **(A)**; and down-regulated genes in NSCLC compared to control **(B)**. GO, Gene Ontology; BP, Biological Processes; MF, Molecular Function; CC, Cell Component; KEGG, Kyoto Encyclopedia of Genes and Genomes.

protein-coupled receptor binding, and chemokine activity; (c) *Cluster 3* is significantly associated with extracellular matrix organization and collagen trimer; (d) *Cluster 4* is significantly associated with G protein-coupled receptor signaling pathway via cyclic nucleotide second messenger; (e) *Cluster 5* is significantly associated with post-translational protein modification, cytokine activity (**Figure S4**).

Upstream Regulator of Cluster

Transcription factors play crucial roles in initiation, progression, and metastasis of cancer. However, the role of TFs and their downstream target genes and their regulatory mechanisms in the development of NSCLC remains largely unknown. Therefore, each MCODE cluster was analyzed to identify the potential upstream TF regulators using the iRegulone tool.

Cluster 1: Our analysis showed that almost all 26 up-regulated genes are under the control of five TFs: FOXM1, MYBL2, TFDP1, E2F4, and SIN3A (**Figure 3A**). However, only FOXM1 and MYBL2 are up-regulated gene showing log2FC >1, while, TFDP1 and E2F4 show slight up-regulated while SIN3A show slight down-regulated in our list of DEGs of NSCLC.

Cluster 2: The cluster 2 contains 10 genes, which is under the regulation of 24 TFs, however, only four TFs showed |log2FC| >1 (TFAP2A and TFAP2C up-regulated; GATA2 and FOS down-regulated) in our list of DEGs (**Figure 3B**).

Cluster 3: The cluster 3 contains 7 genes, which is under the regulation of 7 TFs, however, all of the TF showed |log2FC| < 1 in our list of DEGs (**Figure 3C**).

Cluster 4: The cluster 4 contains 10 genes, which are under the regulation of 20 TFs, however, only GATA6, EBF1, and JUN showed log2FC < −1 in our list of DEGs (**Figure 3D**). ***Cluster 5:*** *Cluster 5* contains 5 genes, which is under the regulation of 22 TFs, however, only six TFs showed |log2FC| >1 (TFAP2A and TFAP2C up-regulated; JUN, FOXA2, JUNB, and FOS down-regulated) in our list of DEGs (**Figure 3E**). *Cluster 1* contains all up-regulated genes, and *Cluster 4* contains all down-regulated genes, however, the rest of the cluster contains both up-and down-regulated genes. Venn diagram showing that few TFs are commonly regulating more than one cluster (**Figure 3F**).

Validation of Upstream Regulator of Cluster

A study found that the expression level of genes and their TFs are highly correlated in spite of cell diversity; while the expression level of randomly selected genes and TFs show very weak correlation (37). Therefore, TFs interacting with its potential target gene in clusters of *NSCLC network* were analyzed for their expression correlation.

Cluster 1: Using Pearson correlation coefficients, all genes in *Cluster 1* are showing significantly highly positive correlation with upstream TFs FOXM1, and MYBL2 in NSCLC (**Figure 4A**; **Figure S5**). As revealed in the figures, their gene expression is induced in NSCLC compared to the normal sample. Top five highly correlated expressed genes with a FOXM1 are (a) CCNB2 ($R = 0.74$); (b) KIF4A ($R = 0.73$); (c) ASPM ($R = 0.72$); (d) KIF11 ($R = 0.70$); and (e) BUB1 ($R = 0.69$). Top five highly correlated

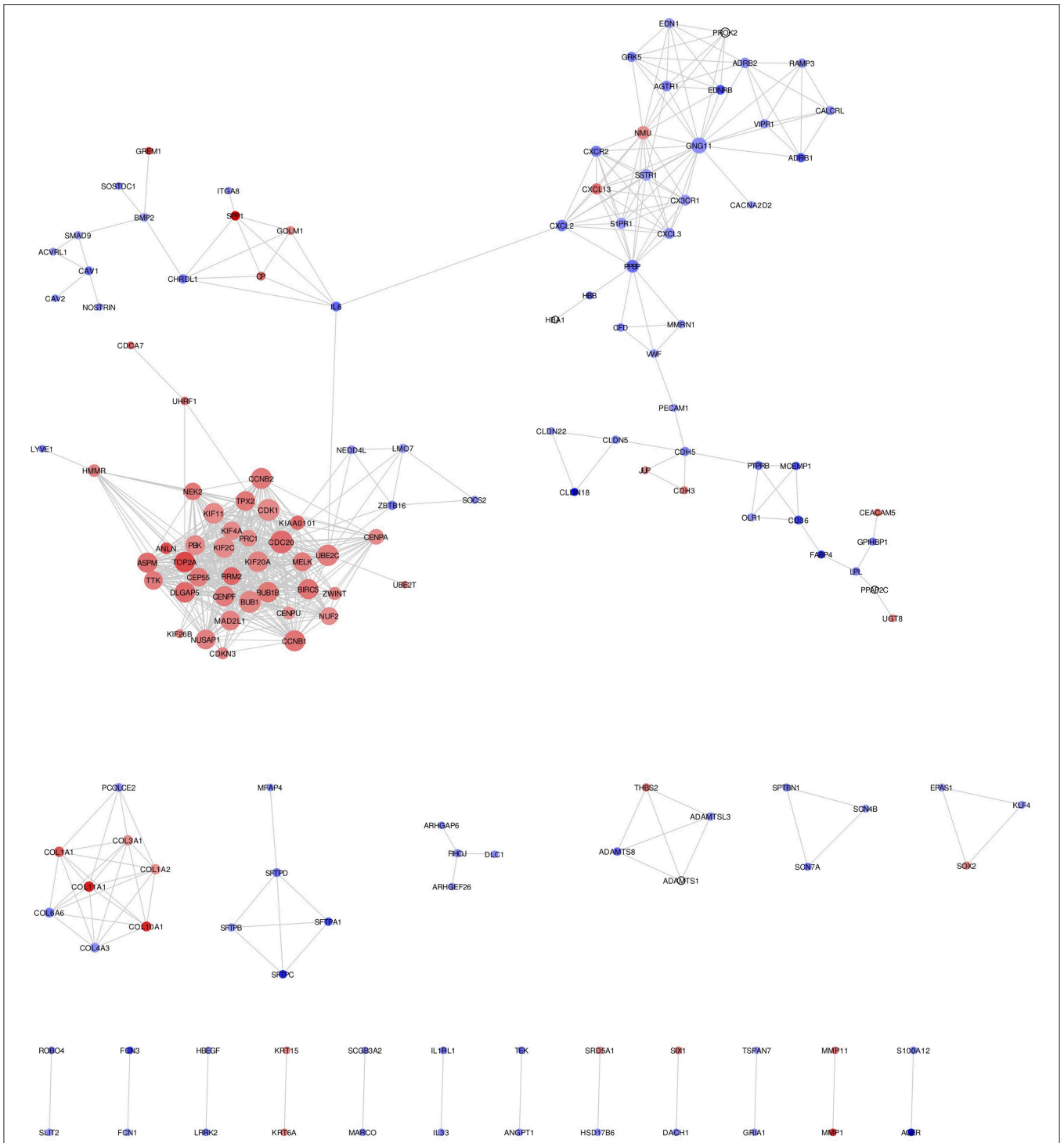


FIGURE 2 | *NSCLC network* showing protein-protein integration network in NSCLC. Red node indicates up-regulated; while blue node indicates down-regulated mRNAs in NSCLC compared to normal. Size of the node is based upon degree of connectivity of the node. Edges in the network represent direct interactions between nodes.

expressed genes with TF MYBL2 are; (a) BUB1 ($R = 0.72$); (b) KIF4A ($R = 0.71$); (c) KIF2C ($R = 0.70$); (d) KIF11 ($R = 0.68$); and (e) NEK2 ($R = 0.67$).

Cluster 2: The highest correlation of $R = 0.53$ was observed between FOS and CXCL2 (**Figure S6**). However, our data showed these both genes are down-regulated in NSCLC.

TABLE 2 | List of 15 highest score clusters identified from *NSCLC Network* by MCODE.

Cluster	Score (Density*#Nodes)	# Nodes	# Edges	Node IDs
1	25.2	26	315	CEP55, KIF20A, DLGAP5, KIF4A, RRM2, KIF11, NUF2, NUSAP1, UBE2C, CCNB1, BUB1, NEK2, TPX2, BUB1B, TOP2A, MAD2L1, PBK, CCNB2, MELK, CDC20, BIRC5, CDK1, CENPF, ASPM, KIF2C, TTK
2	10	10	45	GNG11, CXCL13, NMU, CXCL2, S1PR1, CX3CR1, CXCL3, PPBP, SSTR1, CXCR2
3	7	7	21	COL3A1, COL11A1, COL6A6, COL1A1, COL4A3, COL1A2, COL10A1
4	5.111	10	23	PROK2, VIPR1, ADRB2, ADRB1, EDNRB, EDN1, RAMP3, CALCRL, GRK5, AGTR1
5	5	5	10	GOLM1, SPP1, IL6, CHRDL1, CP
6	4	4	6	PTPRB, MCEMP1, CD36, OLR1
7	4	4	6	THBS2, ADAMTSL3, ADAMTS8, ADAMTS1
8	4	4	6	SFTPA1, SFTPC, SFTPB, SFTPD
9	3	3	3	ZWINT, CENPA, CENPU
10	3	3	3	ZBTB16, NEDD4L, LMO7
11	3	3	3	VWF, MMRN1, CFD
12	3	3	3	SCN7A, SPTBN1, SCN4B
13	3	3	3	SOX2, KLF4, EPAS1
14	3	3	3	ACVRL1, CAV1, SMAD9
15	2.8	6	7	CLDN18, CLDN5, CDH5, CLDN22, CDH3, JUP

Cluster 3: It showed the highest correlation of $R = 0.36$ between TCF12 and COL1A2 (**Figure S7**). However, our data showed COL1A2 is up-regulated, while TCF12 is slightly down-regulated in NSCLC.

Cluster 4: The highest correlation of $R = 0.34$ was observed between GATA6 and RAMP3 (**Figure S8**). Our data showed these two genes are down-regulated in NSCLC.

Cluster 5: Cluster 5 showed the highest correlation of $R = 0.33$ between FOXA2 and GOLM1 (**Figure S9**). However, our data showed GOLM1 is up-regulated, while FOXA2 is down-regulated in NSCLC.

Gene Expression-Based Survival Analysis in NSCLC by Kaplan-Meier Plot

The topologically significant genes in the global *NSCLC network*, genes in MCODE clusters, and upstream regulator TFs (showing $|\log2FC|>1$) were analyzed for association with OS in NSCLC using Kaplan-Meier plots. Kaplan-Meier plots of each cluster and their associated TFs are presented as follow: *Cluster 1* in **Figure 4B** and **Figure S10**; *Cluster 2* in **Figure S11**; *Cluster 3* in **Figure S12**; *Cluster 4* in **Figure S13**; and *Cluster 5* in **Figure S14**. Kaplan-Meier plots showed that high expression of all the up-regulated genes of *Cluster 1* make worse the OS [HR >1], while high expression of down-regulated gene SIN3A makes better the OS [HR <1] in NSCLC (**Figure 4B**; **Figure S10**). Kaplan-Meier plots of the gene of other clusters showed very much similar patterns that high expression of up-regulated genes make worse the OS, while high expression of down-regulated genes make better the OS in NSCLC (**Figures S11–S14**).

Extension of Cluster 1 With miRNA

Our analysis identified 30 up-regulated and 70 down-regulated miRNAs targeting 25 genes of *Cluster 1* and associated five TFs. These data were used to generate miRNA network of *Cluster 1* consisting of 130 nodes and 218 interactions (**Figure 5**). Our analysis found none of the miRNA is targeting NUSAP1 gene of *Cluster 1*. The top five genes targeted by highest number of down-regulated miRNAs in NSCLC are: (a) RRM2 targeted by 17 miRNAs; (b) BIRC5 targeted by 14 miRNAs; (c) CEP55 targeted by 12 miRNAs; (d) KIF2C targeted by 11 miRNAs; (e) CDK1 targeted by 9 miRNAs (**Table S6**). Interestingly, it was found that the expression of TFs, regulators of *Cluster 1* genes, are also under the control of miRNAs as following: (a) FOXM1 is targeted by 10 miRNA (9 down- and 1 up-regulated); (b) MYBL2 is targeted by 6 miRNAs (5 down- and 1 up-regulated); (c) TFDP1 is targeted by 7 miRNAs (6 down- and 1 up-regulated); (d) E2F4 is targeted by 5 miRNAs (4 down- and 1 up-regulated); (e) SIN3A is targeted by 6 miRNAs (4 down- and 2 up-regulated). The complete list of gene and its associated miRNAs are provided in **Table S6**.

The miRNA targeting highest number of genes are as following: (a) hsa-miR-193b-3p is targeting 11 genes (ASPM, BUB1, BUB1B, CDC20, CDK1, KIF11, MELK, RRM2, TOP2A, TPX2, UBE2C); (b) hsa-miR-215-5p is targeting 10 genes (ASPM, BUB1B, CDC20, CENPF, CEP55, DLGAP5, KIF20A, MAD2L1, NUF2, TTK); (c) hsa-miR-186-5p is targeting 7 genes (BUB1, DLGAP5, FOXM1, KIF11, NEK2, RRM2, TOP2A); (d) hsa-miR-16-5p is targeting 7 genes (BIRC5, CDC20, CDK1, CENPF, CEP55, KIF2C, UBE2C); and (e) hsa-miR-30a-5p is targeting 5 genes (CDC20, KIF11, MYBL2, RRM2, TFDP1). The complete list of miRNAs and their targets are provided in the **Table S7**. Interestingly, hsa-miR-193b-3p, hsa-miR-215-5p, hsa-miR-186-5p, hsa-miR-16-5 and hsa-miR-30a-5p are down-regulated and their targets are up-regulated in NSCLC compared to control indicating their role in the development of NSCLC.

Mutational Signatures in NSCLC

Analysis of mutational signatures in 31 genes (*Cluster 1* and associated TFs) in NSCLC studies showed that queried genes are altered in 1966 (37%) out of 5279 samples across TCGA datasets

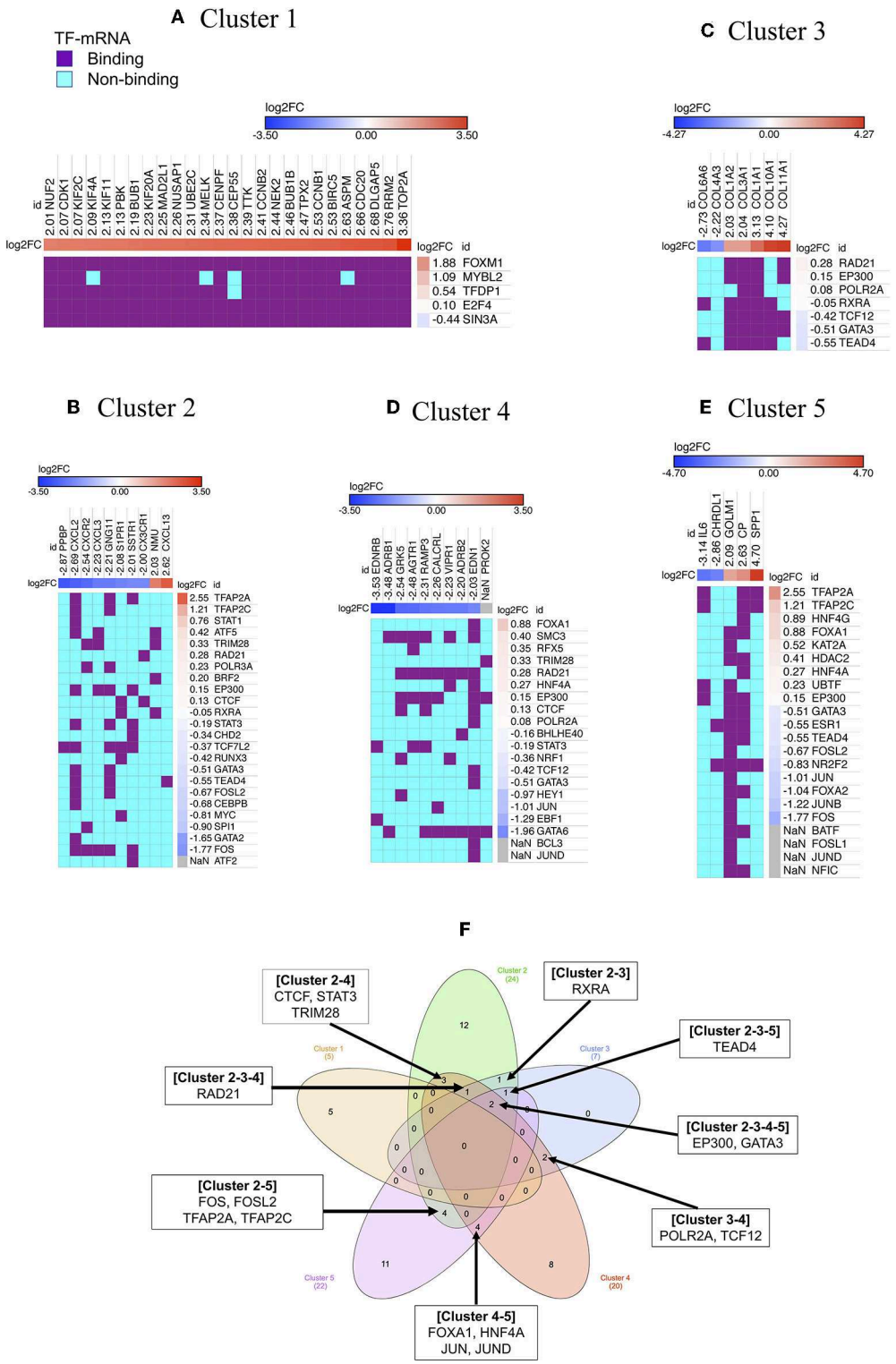


FIGURE 3 | Regulators of gene cluster 1–5. Each column indicates gene in a cluster, while each row indicates TF identified by iRegulone **(A–E)**. Up-regulated DEGs in the cluster is red with positive log2FC; while down-regulated DEGs is blue with negative log2FC. TF binding with the mRNA is in purple, while non-binding in cyan. "NaN" If the log2FC is not available in our list of DEGs. **(F)** Venn diagram showing common TFs regulating different clusters.

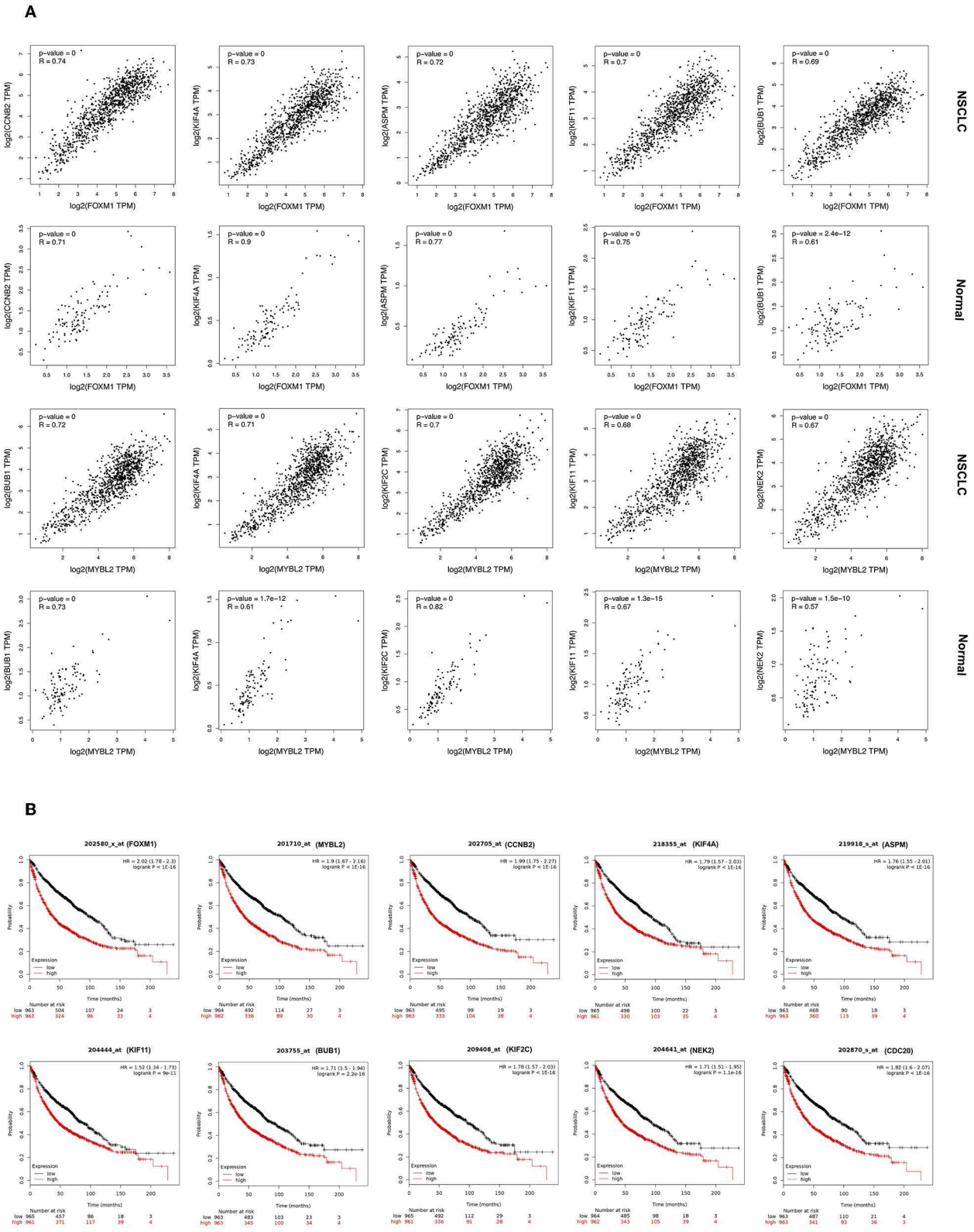


FIGURE 4 | (A) Correlation analysis of expression of genes in *Cluster 1* and its TFs. Expression of gene is on Y-axis while TF is on X-axis. **(B)** Overall survival analysis in NSCLC patients using Kaplan-Meier plots for genes of *Cluster 1* and associated TFs.

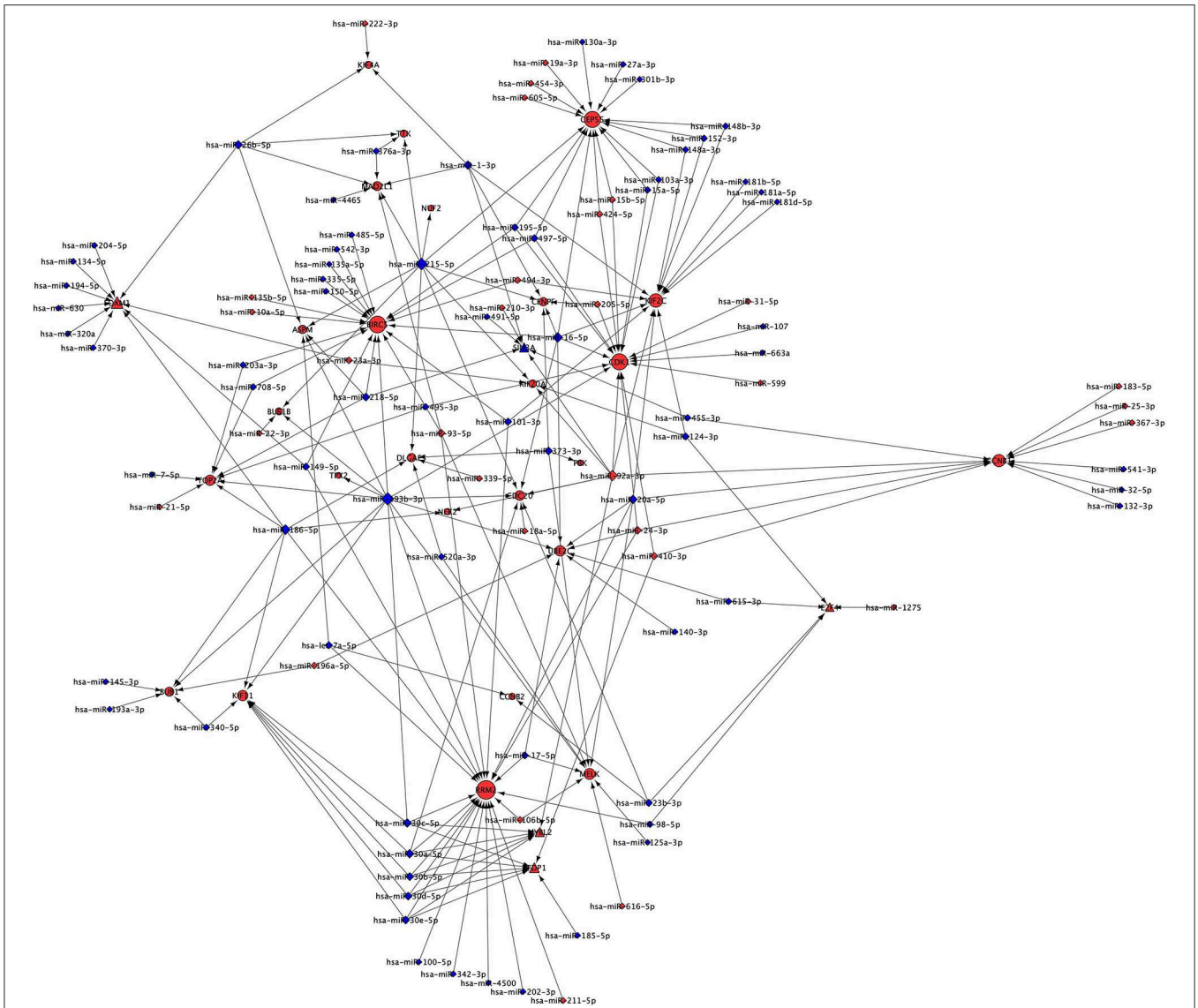


FIGURE 5 | miRNA network of *Cluster 1* showing miRNAs targeting mRNAs and TFs of *Cluster 1*. Red node indicates up-regulated; while the blue node indicates down-regulated expression in NSCLC compared to normal. Size of the node is based upon degree of connectivity of the node. Nodes shape with triangle, round, and diamond represent TFs, mRNAs, and miRNAs, respectively.

(**Figure 6**). The top three highest altered genes are ASPM (10%), NUF2 (6%) and CENPF (6%) (**Figures 6A–D**).

Our accumulating results indicate *Cluster 1* is working as a "**driver-network**" for the initiation of uncontrolled cell proliferation and development of NSCLC.

DISCUSSION

The availability of huge and diverse genome-scale molecular data provide great opportunity to integrate and analyze them to discover new mechanisms and experimentally testable models for initiation and proliferation of cancer (20, 38). Furthermore, the pan-cancer studies utilized the genomics and transcriptomics data and identified differences and commonalities in dysregulation of biological process across multiple cancer types (39, 40). NSCLC is a commonly diagnosed cancer with a high mortality rate. Previous studies identified numerous "driver-genes" as well as abnormally expressed genes and their functional enrichment associated with NSCLC (7, 22–24, 41). However, such studies lack the information of the regulatory network of abnormally expressed genes, which makes difficult to understand the molecular mechanism of development of NSCLC as well as to identify the potential therapeutic target genes. An earlier study integrated the gene expression data, DNA copy number alteration (CAN) and PPI data, and identified "driver-networks" containing potential target genes in breast cancer (20).

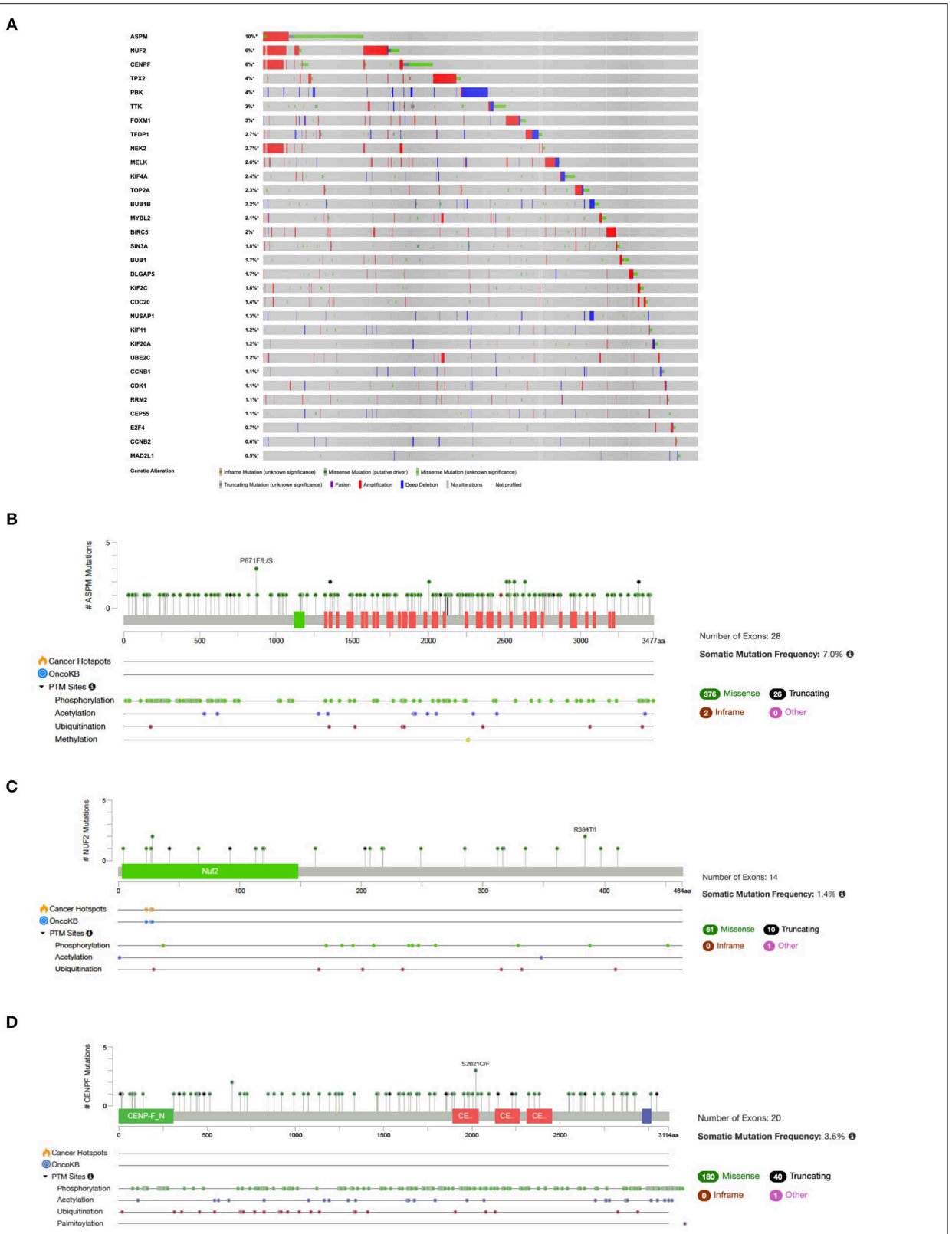


FIGURE 6 | (A) OncoPrint of genes in *Cluster 1* and associated TFs alteration in NSCLC. Lollipop plot with distribution of mutations in NSCLC across protein domains of **(B)** ASPM; **(C)** NUF2; and **(D)** CENPF.

In the current study, a meta-analysis of large gene expression samples (131 NSCLC and 130 control) identified 346 DEGs (97 up-regulated and 249 down-regulated) in NSCLC compared to adjacent non-tumor lung tissues. After integrating the PPI data into DEGs, the *NSCLC network* was created and analyzed to understand the dysregulated sub-networks and pathways in NSCLC. Furthermore, the sub-networks were studied to identify a "driver-network" and its upstream regulators by integrating the data of TFs and miRNAs. The topologically important gene in *NSCLC network,* the "driver-network," and its upstream regulators could be candidate genes for biomarkers and therapeutic target for NSCLC. However, selecting candidate genes for biomarker and therapeutic target requires their overexpression or underexpression have a deep connection at the molecular level for the initiation and progression of tumorigenesis. Therefore, the literature and databases mining was performed on the selected candidate genes and regulatory sub-network to understand their molecular mechanism in NSCLC.

DEGs, Functional Annotation and Pathway Enrichment

Our study found that SPP1 is the most up-regulated gene and its high expression is associated with reducing OS (**Table 1**; **Figure S14**) and thus, support the previous finding showed enhanced expression of SPP1 in several types of tumors including NSCLC (42). Network analysis showed SPP1 is the member of *Cluster 5* and under the regulation of TFs TFAP2A and TFAP2C (**Figure 3E**). However, in spite of overexpression of both TFs in NSCLC, our study found no positive correlation between the expression of these TFs and SPP1 and therefore need further investigation (**Figure S9**). SPP1 binds to CD44 and integrin receptor in the lung cancer cell and activates the FAK/PI3K/AKT pathway which induces the secretion of vascular endothelial growth factor (VEGF) resulting in increased cell survival, cell proliferation and tumor metastasis (42). Silencing the expression of SPP1 using siRNA decreased the NSCLC tumor volume and weight in mice demonstrated it as a promising therapeutic target (43). Furthermore, our analysis showed that AGER is the highly down-regulated gene in NSCLC compared to normal tissue. AGER is a multi-ligand receptor that binds various ligands derived from a damaged cell and its up-regulation at both mRNA and protein level is associated with the majority of cancers including gastric, breast, hepatocellular, colorectal carcinoma (44, 45). However, unlike other cancers, AGER is down-regulated in NSCLC and also supported by the previous finding suggested its role as a tumor suppressor in lung cancer (46). S100A12 is a small protein express by neutrophil granulocytes and binds with AGER receptor, which induces the production of proinflammatory cytokines (47). AGER and S100A12 are interacting in *NSCLC network* and both are down-regulated suggested their role in reducing the inflammation to escape from the immune response in NSCLC (**Figure 2**) (47). A previous study also supports our finding that SPP1 and AGER are highly up-regulated and down-regulated, respectively, in NSCLC (48). A pan-cancer analysis of pediatric leukemias and solid tumors identified 142 and 82 driver genes (40). Comparing with pan-cancer driver genes, our study found that SIX1 was up-regulated, while TAL1 and ID4 were down-regulated in NSCLC (40). The pan-cancer study reported mutation in SIX1, TAL1, and ID4 were present in the Wilms Tumors, T-lineage acute lymphoblastic leukemias, and B-lineage acute lymphoblastic leukemias, respectively (40). Taken together the functional annotation and pathway enrichment analysis of DEGs indicated that up-regulated genes could lead to enhance tumor cell proliferation, while down-regulated genes decreased in immune cell migration in NSCLC, which are vital for uncontrolled cell division and survival in cancer, and to escape from the proper immune response (**Figure 1**).

Topology of NSCLC Network

Topological properties of *NSCLC network* identified several other key proteins. Hub gene plays a key role in the proper maintaining the architecture of the biological network (49, 50). The study found that the intramodular hubs are significantly related to cell proliferation and survival time in cancer (49, 51). All of the protein of *Cluster 1* genes are highly interacting and therefore act as hub genes in the *NSCLC network* (**Table S4**). Top six hub genes with more than 30 degrees of connectivity are CDC20 [i.e. 35], BUB1 [33], CDK1 [33], UBE2C [32], CCNB1 [31], and CCNB2 [31] (**Table S4**). CDC20 regulates cell division through activating the anaphase-promoting complex/cyclosome (APC/C), which begins chromatid separation to enter into anaphase. Overexpression of CDC20 is reported in various cancer including breast cancer, cervical cancer, urinary bladder cancer, and associated with poor prognosis of ovarian tumors (52). It was reported that overexpression of CDC20 is associated with poor prognosis in NSCLC, which support our findings (**Figure 4B**) (8). Therefore, various studies considered CDC20 as a therapeutic target for cancer treatment (52).

The ubiquitin-conjugating enzyme E2 (UBE2C) is a member of the APC/C complex and promotes the degradation of various target proteins required for cell cycle progression. The aberrantly high expression of UBE2C was reported in various cancers. It was experimentally showed that the TF FOXM1 binds to the promoter region of UBE2C and activates its high expression in esophageal squamous cell carcinoma, which supports our finding that the FOXM1 as an upstream regulator of UBE2C of *Cluster 1* (**Figure 3A**) (53). CCNB1 interact with CDK1 to form a complex that phosphorylate their substrates and promotes G2/M transition in the cell cycle. *Cluster 1* contains various protein including CCNB1, which are degraded by APC/C E3 ubiquitin ligase complex. Overexpression of CCNB1 resulting in cell proliferation and was reported in various cancers including NSCLC (9). Inhibiting the expression of CCNB1 using siRNAs promotes apoptosis in colorectal cancer cells (54). BUB1 is component of spindle checkpoint for proper chromosome segregation and its up-regulation was reported in human prostate cancer (55).

Node with the highest betweenness centrality controls the flow of information between two nodes and therefore, could be crucial protein in signaling network and potential drug target to stop the flow of communication in a disease state. The node

RHOJ has the highest betweenness centrality and interacts with DLC1, ARHGEF26, and ARHGAP6 and they all are down-regulated in NSCLC (**Figure 2**; **Table S4**). A study found that the encoded protein of RHOJ is activated by VEGF and may regulate angiogenesis. DLC1 encodes a Rho GTPase-activating protein that functions as a tumor suppressor and down-regulated in more than 95% of NSCLC and other cancers (56). A previous study supports the role of DLC1 as an inducer of apoptosis in NSCLC (57) and as a metastasis suppressor in breast cancer cells (58). ARHGEF26 function as guanine nucleotide exchange factors (GEFs), which catalyze the activation of RHOG by displacing GDP (from inactive enzyme) with GTP (active enzyme). ARHGAP6 involves in regulating the actin polymerization at the cell plasma membrane. Our network analysis indicates that Rho family proteins forming a complex of RHOJ- DLC1-ARHGEF26-ARHGAP6, which need detail study in NSCLC. Node IL6 has second highest betweenness centrality of 0.611 in the *NSCLC network,* which connects 6 proteins across *three* sub-networks. IL-6 is cytokines secreted during inflammation and chronic disease like cancer. It binds with interleukin-6 receptor alpha (IL-6Rα) present on the surface of T-cell, NK cell, B-cell and activates them. IL6 and its interacting partners CXCL2 and CHRDL1 are down-regulated; while other interacting partners UBE2C, SPP1, CP, GOLM1 are up-regulated in the *NSCLC network*. CXCL2 gene encodes secreted proteins and plays an important role in inflammation and immunoregulation. A study found CXCL2 role in the resistance of anti-cancer drug, anlotinib, in NSCLC (59). The overexpression of GOLM1 is reported in prostate cancer (60) and lung adenocarcinoma (61). Because of high betweenness centrality of down-regulated IL6, it may be playing a pivotal role in the down-regulation of the inflammatory response in NSCLC.

Driver-Network and Upstream Regulators

Further analysis of *NSCLC network* identified the biologically informative 15 local cluster networks. Among them, the highest scoring *Cluster 1* identified as local "driver-network" having 26 overexpressed gene and their upstream regulators FOXM1 (log2FC = 1.88), MYBL2 (log2FC = 1.09), TFDP1 (log2FC = 0.54), E2F4 (log2FC = 0.10), and SIN3A (log2FC = −0.44) (**Figure 3A**). The "driver-network" is collectively associated with cell proliferation (*Cluster 1* in **Figure S4**). Interestingly, we observed a strong positive correlation between gene expression of each member of "driver-network" and its upstream regulators FOXM1 and MYBL2 in NSCLC (**Figure 4A**; **Figure S5**). Furthermore, NSCLC patients with their overexpression had significantly worse OS (**Figure 4B**; **Figure S10**).

The previous study found that MuvB core proteins interact with E2F4-DP1 and p130 or p107 to form a DREAM complex in G0/G1 phase of the cell cycle, which put the cell in quiescence state by globally repressing more than 800 cell cycle genes (62). When cell exit from quiescence state, MuvB core proteins dissociated from p130 and interacts with MYBL2 to form MMB (MYBL2-MuvB) complex. Subsequently, MMB recruits FOXM1 protein to form MMB-FOXM1 complex, which binds to the promoters of several cell cycle genes and activate their expression in G2/M phase responsible for mitosis (63). A study found that high expression MYBL2 gene disrupts the DREAM complex and increase the MMB complex formation and subsequently triggers the expression of the several target genes driving the cell proliferation in cancer (64). In this way, MMB complex function as opposite of the DREAM complex. A previous study identified the highly confident candidate target genes and regulatory network of DREAM and MMB-FOXM1 complexes involved in the cell cycle (63). Comparing to the study, it was found that all 26 genes of *Cluster 1* are the target of DREAM complex (63). The same study support that 24 genes of *Cluster 1* (except MELK and PBK) are the target of MMB-FOXM1 complex (63). Furthermore, the TFs FOXM1, MYBL2, SIN3A, and TFDP1 are the target gene of DREAM, but not the MMB-FOXM1 complex. Combining all these findings, our study indicates that most of the genes of *Cluster 1* are the common target for both DREAM and MMB-FOXM1 complexes. However, overexpression of MYBL2 and FOXM1 could disrupt the DREAM complex and enhance the formation of MMB-FOXM1 complex resulting high expression of cell cycle genes in *Cluster 1* which consequences uncontrolled cell proliferation and ultimately NSCLC.

FOXM1 is a member of the Forkhead box family of transcriptional factor that expresses in actively dividing cells. Several studies reported the overexpression of FOXM1 stimulates the proliferation of tumor cells during the progression of NSCLC and other types of cancers and also associated with poor overall survival (13–15, 65). A study found FOXM1 overexpressed NSCLC associated with resistance of cisplatin-based chemotherapy, and its inhibition using thiostrepton or siRNA reversed the drug resistance resulted in inhibition of cell proliferation and induce cell death (14). Silencing of FOXM1 expression by siRNA in A549 lung adenocarcinoma cells resulted in significant reduction in cell cycle-promoting cyclin A2 and cyclin B1 genes, as well as DNA replication and mitosis (13). DREAM complex directly represses the transcription of TOP2A, which encode DNA topoisomerase to relief the torsional stress during DNA transcription and replication (62). Furthermore, the study showed that depleting FOXM1 expression decrease the TOP2A mRNA and protein level in A549 human lung adenocarcinoma cells (66). Experimental studies showed the FOXM1 protein directly bind to the promoter region of TOP2A mRNA (66). Our study showed overexpression of TOP2A (log2FC = 3.36) and its upstream regulator FOXM1 indicates that both genes are the promising target for anti-cancer therapy for NSCLC (67). A pan-cancer study found FOXM1 is overexpressed across all studied 32 TCGA cancer types including NSCLC compared to normal tissues (68). A pan-cancer analysis revealed FOXM1 regulatory network as a top predictor of poor prognosis (69). We found all the genes of *Cluster 1* except NUSAP1, PBK, and CDK1 are present in the pan-cancer network associated with mitotic cell cycle and adverse prognostic genes [see Figure 2d of (69)]. In addition, the pan-cancer network contains the TFs FOXM1 and MYBL2 which support our finding.

The MYBL2 is phosphorylated by cyclin A/cyclin-dependent kinase 2 during the S-phase of the cell cycle and activate the cell division (12). Overexpression of MYBL2 is associated with poor patient survival in various cancers patient including NSCLC (12, 70). A previous study showed experimentally that several genes including KIF20A, KIF4A, NUSAP1, CCNB1, TOP2A,

CDK1, CENPF, and KIF2C of *Cluster 1* are transactivated by MYBL2 (12). Abnormal expression of E2F4 and its mutations are reported in several cancers including NSCLC (16, 71, 72).

The genes of *Cluster 1* are mainly involving in the proliferation of cell division. Our study supports the previous finding that overexpression of *Cluster 1* genes in several cancer including NSCLC and associated with poor overall survival such as: KIF2C (73), KIF4A (11, 74), and KIF11 (75) which are kinesin family members of motor proteins regulating the cell mitosis through faithful chromosome condensation and segregation (76). Furthermore, it was reported that silencing of their expression using siRNAs inhibit the cancer cell growth (11, 73). Several inhibitors of KIF11 such as Monastrol, Ispinesib, and Dimethylenastron have been developed and are in clinically used to inhibit cell proliferation and induce apoptosis to treat numerous cancers (77). The high expression of ASPM responsible for mitotic spindle formation (10, 78). NEK2 (79, 80) which triggers centrosome separation are reported in NSCLC. Interestingly, another study also supports that E2F4 and FOXM1 bind to the promoter of NEK2 gene (80). NUF2 component of NDC80 Kinetochore complex regulating the chromosome segregation was overexpressed in NSCLC (81). Suppressing the expression of NUF2 inhibits tumor growth and also stimulates cell apoptosis (81). CENPF (82, 83) which is associated with the centromere-kinetochore complex and requires for chromosome segregation during mitosis.

Therefore, the "driver-network" and the predicted TFs MYBL2 and FOXM1 give more insight about the initiation and progression of NSCLC and also could be therapeutic target genes. However, a further biochemical study is required to understand the effect on cell proliferation in NSCLC by using functional siRNAs targeting a combination of TFs FOXM1, and MYBL2, and their downstream genes of "driver-network" (84).

Our study found that the "driver-network" is not only under the regulation of TFs, but also under the regulation of miRNAs. This study showed that 9 miRNAs (hsa-miR-134-5p, hsa-miR-149-5p, hsa-miR-186-5p, hsa-miR-194-5p, hsa-miR-204-5p, hsa-miR-26b-5p, hsa-miR-320a, hsa-miR-370-3p, hsa-miR-630) targeting FOXM1 are down-regulated in NSCLC. In addition, these miRNAs are also targeting 11 other genes (ASPM, BIRC5, BUB1, DLGAP5, KIF11, KIF4A, MAD2L1, NEK2, RRM2, TOP2A, and TTK), which means these are common targets for miRNAs and FOXM1 (**Table S7**). Previous studies found the down-regulation of has-miR-134-5p and hsa-miR-149-5p were contributing epithelial-to-mesenchymal transition (EMT), a key process of cancer metastasis, in NSCLC (85, 86). These studies also demonstrated that has-miR-134-5p and hsa-miR-149-5p act as tumor suppressors by directly binds to the 3′UTR of FOXM1 and inhibiting its expression and the EMT in NSCLC (85, 86). Accumulating evidence indicate that down-regulation of these miRNAs eliminate their suppressive effect resulting overexpression of FOXM1 and its 11 downstream target genes. The decrease expression of other miRNAs in NSCLC and their role in cell proliferation and EMT has been demonstrated in several studies including hsa-miR-194-5p (87), hsa-miR-204-5p (88), hsa-miR-26b-5p (89), hsa-miR-320a (90), hsa-miR-370-3p (91) and hsa-miR-630 (92). MYBL2 and TFDP1 are targeted by five common miRNAs (hsa-miR-30a-5p, hsa-miR-30b-5p, hsa-miR-30c-5p, hsa-miR-30d-5p, hsa-miR-30e-5p), though they belong from same miRNA family. A previous study showed the down-regulation of hsa-miR-30a-5p which directly targeting MYBL2 mRNA in NSCLC (93). Interestingly, TCGA NSCLC dataset showed higher mutation rates in the genes of "driver-network" as well as its upstream regulators FOXM1 and MYBL1 in the NSCLC (**Figure 6**).

Previous bioinformatics studies were mainly focused on the analysis of gene expression data to identify the DEGs, their function enrichment, the interacting hub genes in NSCLC. Ni *et al.* identified five up-regulated hub genes (TOP2A, CCNB1, CCNA2, UBE2C, and KIF20A) in NSCLC (41). Huang *et al.* identified five up-regulated hub genes (CDC20, CENPF, KIF2C, BUB1, and ZWINT) in NSCLC (94). Another study found 16 hub genes (TEK, ANGPT1, MMP9, VWF, CDH5, EDN1, ESAM, CCNE1, CDC45, PRC1, CCNB2, AURKA, MELK, CDC20, TOP2A, and PTTG1) in NSCLC (95). However, our study has following advantages compared to previous studies: (a) Current study is based upon a large dataset of NSCLC obtained from different GEO microarray dataset; (b) Identified various common DEGs detected by previous studies; (c) The DEGs were integrated with PPI, TFs and miRNAs to understand the regulatory mechanism of NSCLC initiation and progression; (d) Finally, we have identified a "driver-network" consist of 26 up-regulated hub genes and their upstream regulators (FOXM1, MYBL2, and miRNAs) involved in the proliferation of NSCLC and could serve as diagnostic and therapeutic targets to treat NSCLC.

Our study suggested that a gene could be functionally important even at a small level of overexpression such as FOXM1 and MYBL2 if they act as upstream regulators of genes of interacting proteins involved in an important biological process (20). However, our study has following limitations: (a) All the findings are based upon computational analysis using integrated data of gene expression, PPI, TFs, and miRNAs; (b) Our study did not integrate the data of gene mutations and gene copy number variations, which could abolish the cis- and trans-regulatory elements of a gene resulting aberrant gene expression and cancer (96–98); and (c) Finally, our study lacks the experimental validation and therefore need further experimental testing.

CONCLUSION

In this study, potential biomarkers and therapeutic targets has been identified for NSCLC using systems bioinformatics approach on the public gene expression data. All the genes in "driver-network" (*Cluster 1)* and its upstream regulators, FOXM1 and MYBL2, which collectively overexpressed and involve in the cell proliferation, and cell division are particularly promising for further study. Furthermore, we identified several tumor suppressor miRNAs and their interacting target genes in the "driver-network." Targeting two or more genes of the "driver-network" may be synergistic and more effective therapy against NSCLC. In our study, correlation expression, OS, and gene mutations dataset with strong statistical support were used to validate our finding. However, the biochemical study on the potential biomarkers and therapeutic targets are necessary for further validation on clinical samples.

DATA AVAILABILITY STATEMENT

This study was conducted on publicly available data on Gene Expression Omnibus (GEO https://www.ncbi.nlm.nih.gov/geo/) with accession number: GSE27262, GSE18842, and GSE19804.

AUTHOR CONTRIBUTIONS

FA conceptualized the whole study, designed and performed all the Bioinformatics analysis, interpreted the results, wrote, and revised the manuscript.

ACKNOWLEDGMENTS

The author would like to acknowledge the technical support provided by the Department of Biochemistry, and University of Jeddah Center for Scientific and Medical Research (UJCSMR), University of Jeddah.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2019.01011/full#supplementary-material

REFERENCES

1. Gridelli C, Rossi A, Carbone DP, Guarize J, Karachaliou N, Mok T, et al. Non-small-cell lung cancer. *Nat Rev Dis Primers*. (2015) 1:15009. doi: 10.1038/nrdp.2015.48
2. Cancer Genome Atlas Research N. Comprehensive molecular profiling of lung adenocarcinoma. *Nature*. (2014) 511:543–50. doi: 10.1038/nature13385
3. Cancer Genome Atlas Research N. Comprehensive genomic characterization of squamous cell lung cancers. *Nature*. (2012) 489:519–25. doi: 10.1038/nature11404
4. Paez JG, Janne PA, Lee JC, Tracy S, Greulich H, Gabriel S, et al. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science*. (2004) 304:1497–500. doi: 10.1126/science.1099314
5. Pao W, Miller V, Zakowski M, Doherty J, Politi K, Sarkaria I, et al. EGF receptor gene mutations are common in lung cancers from "never smokers" and are associated with sensitivity of tumors to gefitinib and erlotinib. *Proc Natl Acad Sci USA*. (2004) 101:13306–11. doi: 10.1073/pnas.0405220101
6. Vavala T, Novello S. Alectinib in the treatment of ALK-positive non-small cell lung cancer: an update on its properties, efficacy, safety and place in therapy. *Ther Adv Med Oncol*. (2018) 10:1758835918789364. doi: 10.1177/1758835918789364
7. Sanchez-Palencia A, Gomez-Morales M, Gomez-Capilla JA, Pedraza V, Boyero L, Rosell R, et al. Gene expression profiling reveals novel biomarkers in nonsmall cell lung cancer. *Int J Cancer*. (2011) 129:355–64. doi: 10.1002/ijc.25704
8. Kato T, Daigo Y, Aragaki M, Ishikawa K, Sato M, Kaji M. Overexpression of CDC20 predicts poor prognosis in primary non-small cell lung cancer patients. *J Surg Oncol*. (2012) 106:423–30. doi: 10.1002/jso.23109
9. Yoshida T, Tanaka S, Mogi A, Shitara Y, Kuwano H. The clinical significance of Cyclin B1 and Wee1 expression in non-small-cell lung cancer. *Ann Oncol*. (2004) 15:252–6. doi: 10.1093/annonc/mdh073
10. Jung HM, Choi SJ, Kim JK. Expression profiles of SV40-immortalization-associated genes upregulated in various human cancers. *J Cell Biochem*. (2009) 106:703–13. doi: 10.1002/jcb.22063
11. Taniwaki M, Takano A, Ishikawa N, Yasui W, Inai K, Nishimura H, et al. Activation of KIF4A as a prognostic biomarker and therapeutic target for lung cancer. *Clin Cancer Res*. (2007) 13(22 Pt 1):6624–31. doi: 10.1158/1078-0432.CCR-07-1328
12. Musa J, Aynaud MM, Mirabeau O, Delattre O, Grunewald TG. MYBL2 (B-Myb): a central regulator of cell proliferation, cell survival and differentiation involved in tumorigenesis. *Cell Death Dis*. (2017) 8:e2895. doi: 10.1038/cddis.2017.244
13. Kim IM, Ackerson T, Ramakrishna S, Tretiakova M, Wang IC, Kalin TV, et al. The Forkhead Box m1 transcription factor stimulates the proliferation of tumor cells during development of lung cancer. *Cancer Res*. (2006) 66:2153–61. doi: 10.1158/0008-5472.CAN-05-3003
14. Wang Y, Wen L, Zhao SH, Ai ZH, Guo JZ, Liu WC. FoxM1 expression is significantly associated with cisplatin-based chemotherapy resistance and poor prognosis in advanced non-small cell lung cancer patients. *Lung Cancer*. (2013) 79:173–9. doi: 10.1016/j.lungcan.2012.10.019
15. Yang DK, Son CH, Lee SK, Choi PJ, Lee KE, Roh MS. Forkhead box M1 expression in pulmonary squamous cell carcinoma: correlation with clinicopathologic features and its prognostic significance. *Hum Pathol*. (2009) 40:464–70. doi: 10.1016/j.humpath.2008.10.001
16. Bankovic J, Stojsic J, Jovanovic D, Andjelkovic T, Milinkovic V, Ruzdijic S, et al. Identification of genes associated with non-small-cell lung cancer promotion and progression. *Lung Cancer*. (2010) 67:151–9. doi: 10.1016/j.lungcan.2009.04.010
17. Ahmed F, Senthil-Kumar M, Lee S, Dai X, Mysore KS, Zhao PX. Comprehensive analysis of small RNA-seq data reveals that combination of miRNA with its isomiRs increase the accuracy of target prediction in *Arabidopsis thaliana*. *RNA Biol*. (2014) 11:1414–29. doi: 10.1080/15476286.2014.996474
18. Ahmed F, Ansari HR, Raghava GP. Prediction of guide strand of microRNAs from its sequence and secondary structure. *BMC Bioinformatics*. (2009) 10:105. doi: 10.1186/1471-2105-10-105
19. Han Y, Li H. miRNAs as biomarkers and for the early detection of non-small cell lung cancer (NSCLC). *J Thorac Dis*. (2018) 10:3119–31. doi: 10.21037/jtd.2018.05.32
20. Dutta B, Pusztai L, Qi Y, Andre F, Lazar V, Bianchini G, et al. A network-based, integrative study to identify core biological pathways that drive breast cancer clinical subtypes. *Br J Cancer*. (2012) 106:1107–16. doi: 10.1038/bjc.2011.584
21. Sanchez-Vega F, Mina M, Armenia J, Chatila WK, Luna A, La KC, et al. Oncogenic signaling pathways in the cancer genome atlas. *Cell*. (2018) 173:321–37 e10. doi: 10.1016/j.cell.2018.03.035
22. Wei TY, Hsia JY, Chiu SC, Su LJ, Juan CC, Lee YC, et al. Methylosome protein 50 promotes androgen- and estrogen-independent tumorigenesis. *Cell Signal*. (2014) 26:2940–50. doi: 10.1016/j.cellsig.2014.09.014
23. Wei TY, Juan CC, Hisa JY, Su LJ, Lee YC, Chou HY, et al. Protein arginine methyltransferase 5 is a potential oncoprotein that upregulates G1 cyclins/cyclin-dependent kinases and the phosphoinositide 3-kinase/AKT signaling cascade. *Cancer Sci*. (2012) 103:1640–50. doi: 10.1111/j.1349-7006.2012.02367.x
24. Lu TP, Tsai MH, Lee JM, Hsu CP, Chen PC, Lin CW, et al. Identification of a novel biomarker, SEMA5A, for non-small cell lung carcinoma in nonsmoking women. *Cancer Epidemiol Biomarkers Prev*. (2010) 19:2590–7. doi: 10.1158/1055-9965.EPI-10-0332
25. Carvalho BS, Irizarry RA. A framework for oligonucleotide microarray preprocessing. *Bioinformatics*. (2010) 26:2363–7. doi: 10.1093/bioinformatics/btq431
26. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics*. (2012) 28:882–3. doi: 10.1093/bioinformatics/bts034
27. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. (2015) 43:e47. doi: 10.1093/nar/gkv007
28. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. (2009) 4:44–57. doi: 10.1038/nprot.2008.211

29. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. (2012) 16:284–7. doi: 10.1089/omi.2011.0118
30. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* (2015) 43:D447–52. doi: 10.1093/nar/gku1003
31. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. (2003) 13:2498–504. doi: 10.1101/gr.1239303
32. Bader GD, Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics*. (2003) 4:2. doi: 10.1186/1471-2105-4-2
33. Janky R, Verfaillie A, Imrichova H, Van de Sande B, Standaert L, Christiaens V, et al. iRegulon: from a gene list to a gene regulatory network using large motif and track collections. *PLoS Comput Biol*. (2014) 10:e1003731. doi: 10.1371/journal.pcbi.1003731
34. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res*. (2017) 45:W98–W102. doi: 10.1093/nar/gkx247
35. Gyorffy 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. doi: 10.1371/journal.pone.0082241
36. Xie B, Ding Q, Han H, Wu D. miRCancer: a microRNA-cancer association database constructed by text mining on literature. *Bioinformatics*. (2013) 29:638–44. doi: 10.1093/bioinformatics/btt014
37. Wu G, Ji H. ChIPXpress: using publicly available gene expression data to improve ChIP-seq and ChIP-chip target gene ranking. *BMC Bioinformatics*. (2013) 14:188. doi: 10.1186/1471-2105-14-188
38. Snijders AM, Lee SY, Hang B, Hao W, Bissell MJ, Mao JH. FAM83 family oncogenes are broadly involved in human cancers: an integrative multi-omics approach. *Mol Oncol*. (2017) 11:167–79. doi: 10.1002/1878-0261.12016
39. Leiserson MD, Vandin F, Wu HT, Dobson JR, Eldridge JV, Thomas JL, et al. Pan-cancer network analysis identifies combinations of rare somatic mutations across pathways and protein complexes. *Nat Genet*. (2015) 47:106–14. doi: 10.1038/ng.3168
40. Ma X, Liu Y, Liu Y, Alexandrov LB, Edmonson MN, Gawad C, et al. Pan-cancer genome and transcriptome analyses of 1,699 paediatric leukaemias and solid tumours. *Nature*. (2018) 555:371–6. doi: 10.1038/nature25795
41. Ni M, Liu X, Wu J, Zhang D, Tian J, Wang T, et al. Identification of candidate biomarkers correlated with the pathogenesis and prognosis of non-small cell lung cancer via integrated bioinformatics analysis. *Front Genet*. (2018) 9:469. doi: 10.3389/fgene.2018.00469
42. Zhao H, Chen Q, Alam A, Cui J, Suen KC, Soo AP, et al. The role of osteopontin in the progression of solid organ tumour. *Cell Death Dis*. (2018) 9:356. doi: 10.1038/s41419-018-0391-6
43. Cho WY, Hong SH, Singh B, Islam MA, Lee S, Lee AY, et al. Suppression of tumor growth in lung cancer xenograft model mice by poly(sorbitol-co-PEI)-mediated delivery of osteopontin siRNA. *Eur J Pharm Biopharm*. (2015) 94:450–62. doi: 10.1016/j.ejpb.2015.06.017
44. Wang D, Li T, Ye G, Shen Z, Hu Y, Mou T, et al. Overexpression of the receptor for advanced glycation endproducts (RAGE) is associated with poor prognosis in gastric cancer. *PLoS ONE*. (2015) 10:e0122697. doi: 10.1371/journal.pone.0122697
45. Kostova N, Zlateva S, Ugrinova I, Pasheva E. The expression of HMGB1 protein and its receptor RAGE in human malignant tumors. *Mol Cell Biochem*. (2010) 337:251–8. doi: 10.1007/s11010-009-0305-0
46. Wu S, Mao L, Li Y, Yin Y, Yuan W, Chen Y, et al. RAGE may act as a tumour suppressor to regulate lung cancer development. *Gene*. (2018) 651:86–93. doi: 10.1016/j.gene.2018.02.009
47. Meijer B, Gearry RB, Day AS. The role of S100A12 as a systemic marker of inflammation. *Int J Inflam*. (2012) 2012:907078. doi: 10.1155/2012/907078
48. Zhang W, Fan J, Chen Q, Lei C, Qiao B, Liu Q. SPP1 and AGER as potential prognostic biomarkers for lung adenocarcinoma. *Oncol Lett*. (2018) 15:7028–36. doi: 10.3892/ol.2018.8235
49. Langfelder P, Mischel PS, Horvath S. When is hub gene selection better than standard meta-analysis? *PLoS ONE*. (2013) 8:e61505. doi: 10.1371/journal.pone.0061505
50. Barabasi AL, Oltvai ZN. Network biology: understanding the cell's functional organization. *Nat Rev Genet*. (2004) 5:101–13. doi: 10.1038/nrg1272
51. Horvath S, Zhang B, Carlson M, Lu KV, Zhu S, Felciano RM, et al. Analysis of oncogenic signaling networks in glioblastoma identifies ASPM as a molecular target. *Proc Natl Acad Sci USA*. (2006) 103:17402–7. doi: 10.1073/pnas.0608396103
52. Wang Z, Wan L, Zhong J, Inuzuka H, Liu P, Sarkar FH, et al. Cdc20: a potential novel therapeutic target for cancer treatment. *Curr Pharm Des*. (2013) 19:3210–4. doi: 10.2174/1381612811319180005
53. Nicolau-Neto P, Palumbo A, De Martino M, Esposito F, de Almeida Simao T, Fusco A, et al. UBE2C is a transcriptional target of the cell cycle regulator FOXM1. *Genes (Basel)*. (2018) 9:188. doi: 10.3390/genes9040188
54. Fang Y, Yu H, Liang X, Xu J, Cai X. Chk1-induced CCNB1 overexpression promotes cell proliferation and tumor growth in human colorectal cancer. *Cancer Biol Ther*. (2014) 15:1268–79. doi: 10.4161/cbt.29691
55. Guo C, Wu G, Chin JL, Bauman G, Moussa M, Wang F, et al. Bub1 up-regulation and hyperphosphorylation promote malignant transformation in SV40 tag-induced transgenic mouse models. *Mol Cancer Res*. (2006) 4:957–69. doi: 10.1158/1541-7786.MCR-06-0168
56. Ullmannova V, Popescu NC. Expression profile of the tumor suppressor genes DLC-1 and DLC-2 in solid tumors. *Int J Oncol*. (2006) 29:1127–32. doi: 10.3892/ijo.29.5.1127
57. Yuan BZ, Jefferson AM, Millecchia L, Popescu NC, Reynolds SH. Morphological changes and nuclear translocation of DLC1 tumor suppressor protein precede apoptosis in human non-small cell lung carcinoma cells. *Exp Cell Res*. (2007) 313:3868–80. doi: 10.1016/j.yexcr.2007.08.009
58. Goodison S, Yuan J, Sloan D, Kim R, Li C, Popescu NC, et al. The RhoGAP protein DLC-1 functions as a metastasis suppressor in breast cancer cells. *Cancer Res*. (2005) 65:6042–53. doi: 10.1158/0008-5472.CAN-04-3043
59. Lu J, Xu W, Qian J, Wang S, Zhang B, Zhang L, et al. Transcriptome profiling analysis reveals that CXCL2 is involved in anlotinib resistance in human lung cancer cells. *BMC Med Genomics*. (2019) 12(Suppl. 2):38. doi: 10.1186/s12920-019-0482-y
60. Wei S, Dunn TA, Isaacs WB, De Marzo AM, Luo J. GOLPH2 and MYO6: putative prostate cancer markers localized to the Golgi apparatus. *Prostate*. (2008) 68:1387–95. doi: 10.1002/pros.20806
61. Zhang F, Gu Y, Li X, Wang W, He J, Peng T. Up-regulated Golgi phosphoprotein 2 (GOLPH2) expression in lung adenocarcinoma tissue. *Clin Biochem*. (2010) 43:983–91. doi: 10.1016/j.clinbiochem.2010.05.010
62. Litovchick L, Sadasivam S, Florens L, Zhu X, Swanson SK, Velmurugan S, et al. Evolutionarily conserved multisubunit RBL2/p130 and E2F4 protein complex represses human cell cycle-dependent genes in quiescence. *Mol Cell*. (2007) 26:539–51. doi: 10.1016/j.molcel.2007.04.015
63. Fischer M, Grossmann P, Padi M, DeCaprio JA. Integration of TP53, DREAM, MMB-FOXM1 and RB-E2F target gene analyses identifies cell cycle gene regulatory networks. *Nucleic Acids Res*. (2016) 44:6070–86. doi: 10.1093/nar/gkw523
64. Iness AN, Felthousen J, Ananthapadmanabhan V, Sesay F, Saini S, Guiley KZ, et al. The cell cycle regulatory DREAM complex is disrupted by high expression of oncogenic B-Myb. *Oncogene*. (2019) 38:1080–92. doi: 10.1038/s41388-018-0490-y
65. Zhang J, Zhang J, Cui X, Yang Y, Li M, Qu J, et al. FoxM1: a novel tumor biomarker of lung cancer. *Int J Clin Exp Med*. (2015) 8:3136–40.
66. Wang IC, Meliton L, Ren X, Zhang Y, Balli D, Snyder J, et al. Deletion of Forkhead Box M1 transcription factor from respiratory epithelial cells inhibits pulmonary tumorigenesis. *PLoS ONE*. (2009) 4:e6609. doi: 10.1371/journal.pone.0006609
67. McClendon AK, Osheroff N. DNA topoisomerase II, genotoxicity, and cancer. *Mutat Res*. (2007) 623:83–97. doi: 10.1016/j.mrfmmm.2007.06.009
68. Barger CJ, Branick C, Chee L, Karpf AR. Pan-cancer analyses reveal genomic features of FOXM1 overexpression in cancer. *Cancers (Basel)*. (2019) 11:E251. doi: 10.3390/cancers11020251

69. Gentles AJ, Newman AM, Liu CL, Bratman SV, Feng W, Kim D, et al. The prognostic landscape of genes and infiltrating immune cells across human cancers. *Nat Med*. (2015) 21:938–45. doi: 10.1038/nm.3909
70. Fan X, Wang Y, Jiang T, Cai W, Jin Y, Niu Y, et al. B-Myb mediates proliferation and migration of non-small-cell lung cancer via suppressing IGFBP3. *Int J Mol Sci.* (2018) 19:1479. doi: 10.3390/ijms19051479
71. Souza RF, Yin J, Smolinski KN, Zou TT, Wang S, Shi YQ, et al. Frequent mutation of the E2F-4 cell cycle gene in primary human gastrointestinal tumors. *Cancer Res.* (1997) 57:2350–3.
72. Schwemmle S, Pfeifer GP. Genomic structure and mutation screening of the E2F4 gene in human tumors. *Int J Cancer.* (2000) 86:672–7.
73. Shimo A, Tanikawa C, Nishidate T, Lin ML, Matsuda K, Park JH, et al. Involvement of kinesin family member 2C/mitotic centromere-associated kinesin overexpression in mammary carcinogenesis. *Cancer Sci.* (2008) 99:62–70. doi: 10.1111/j.1349-7006.2007.00635.x
74. Narayan G, Bourdon V, Chaganti S, Arias-Pulido H, Nandula SV, Rao PH, et al. Gene dosage alterations revealed by cDNA microarray analysis in cervical cancer: identification of candidate amplified and overexpressed genes. *Genes Chromosomes Cancer*. (2007) 46:373–84. doi: 10.1002/gcc.20418
75. Castillo A, Morse HC, 3rd, Godfrey VL, Naeem R, Justice MJ. Overexpression of Eg5 causes genomic instability and tumor formation in mice. *Cancer Res.* (2007) 67:10138–47. doi: 10.1158/0008-5472.CAN-07-0326
76. Mazumdar M, Sundareshan S, Misteli T. Human chromokinesin KIF4A functions in chromosome condensation and segregation. *J Cell Biol.* (2004) 166:613–20. doi: 10.1083/jcb.200401142
77. Yu Y, Feng YM. The role of kinesin family proteins in tumorigenesis and progression: potential biomarkers and molecular targets for cancer therapy. *Cancer*. (2010) 116:5150–60. doi: 10.1002/cncr.25461
78. Xu Z, Zhang Q, Luh F, Jin B, Liu X. Overexpression of the ASPM gene is associated with aggressiveness and poor outcome in bladder cancer. *Oncol Lett.* (2019) 17:1865–76. doi: 10.3892/ol.2018.9762
79. Zhong X, Guan X, Dong Q, Yang S, Liu W, Zhang L. Examining Nek2 as a better proliferation marker in non-small cell lung cancer prognosis. *Tumour Biol.* (2014) 35:7155–62. doi: 10.1007/s13277-014-1935-8
80. Fang Y, Zhang X. Targeting NEK2 as a promising therapeutic approach for cancer treatment. *Cell Cycle*. (2016) 15:895–907. doi: 10.1080/15384101.2016.1152430
81. Hu P, Shangguan J, Zhang L. Downregulation of NUF2 inhibits tumor growth and induces apoptosis by regulating lncRNA AF339813. *Int J Clin Exp Pathol.* (2015) 8:2638–48.
82. O'Brien SL, Fagan A, Fox EJ, Millikan RC, Culhane AC, Brennan DJ, et al. CENP-F expression is associated with poor prognosis and chromosomal instability in patients with primary breast cancer. *Int J Cancer.* (2007) 120:1434–43. doi: 10.1002/ijc.22413
83. Kim HE, Kim DG, Lee KJ, Son JG, Song MY, Park YM, et al. Frequent amplification of CENPF, GMNN and CDK13 genes in hepatocellular carcinomas. *PLoS ONE*. (2012) 7:e43223. doi: 10.1371/journal.pone.0043223
84. Ahmed F, Raghava GP. Designing of highly effective complementary and mismatch siRNAs for silencing a gene. *PLoS ONE*. (2011) 6:e23443. doi: 10.1371/journal.pone.0023443
85. Li J, Wang Y, Luo J, Fu Z, Ying J, Yu Y, et al. miR-134 inhibits epithelial to mesenchymal transition by targeting FOXM1 in non-small cell lung cancer cells. *FEBS Lett*. (2012) 586:3761–5. doi: 10.1016/j.febslet.2012.09.016
86. Ke Y, Zhao W, Xiong J, Cao R. miR-149 inhibits non-small-cell lung cancer cells EMT by targeting FOXM1. *Biochem Res Int.* (2013) 2013:506731. doi: 10.1155/2013/506731
87. Zhu X, Li D, Yu F, Jia C, Xie J, Ma Y, et al. miR-194 inhibits the proliferation, invasion, migration, and enhances the chemosensitivity of non-small cell lung cancer cells by targeting forkhead box A1 protein. *Oncotarget.* (2016) 7:13139–52. doi: 10.18632/oncotarget.7545
88. Wang P, Lv HY, Zhou DM, Zhang EN. miR-204 suppresses non-small-cell lung carcinoma (NSCLC) invasion and migration by targeting JAK2. *Genet Mol Res.* (2016) 2:gmr.15026415. doi: 10.4238/gmr.15026415
89. Wang Y, Chen J, Lin Z, Cao J, Huang H, Jiang Y, et al. Role of deregulated microRNAs in non-small cell lung cancer progression using fresh-frozen and formalin-fixed, paraffin-embedded samples. *Oncol Lett.* (2016) 11:801–8. doi: 10.3892/ol.2015.3976
90. Zhang G, Jiang G, Wang C, Zhong K, Zhang J, Xue Q, et al. Decreased expression of microRNA-320a promotes proliferation and invasion of non-small cell lung cancer cells by increasing VDAC1 expression. *Oncotarget.* (2016) 7:49470–80. doi: 10.18632/oncotarget.9943
91. Chen W, Wang J, Liu S, Wang S, Cheng Y, Zhou W, et al. MicroRNA-361-3p suppresses tumor cell proliferation and metastasis by directly targeting SH2B1 in NSCLC. *J Exp Clin Cancer Res.* (2016) 35:76. doi: 10.1186/s13046-016-0357-4
92. Chen MJ, Wu DW, Wang GC, Wang YC, Chen CY, Lee H. MicroRNA-630 may confer favorable cisplatin-based chemotherapy and clinical outcomes in non-small cell lung cancer by targeting Bcl-2. *Oncotarget*. (2018) 9:13758–67. doi: 10.18632/oncotarget.24474
93. Geng GJ, Yang YT, Jiang J, Yu XY, Fa XE. MicroRNA-30a suppresses non-small-cell lung cancer by targeting Myb-related protein B. *Exp Ther Med.* (2018) 15:1633–9. doi: 10.3892/etm.2017.5559
94. Huang R, Gao L. Identification of potential diagnostic and prognostic biomarkers in non-small cell lung cancer based on microarray data. *Oncol Lett.* (2018) 15:6436–42. doi: 10.3892/ol.2018.8153
95. Piao J, Sun J, Yang Y, Jin T, Chen L, Lin Z. Target gene screening and evaluation of prognostic values in non-small cell lung cancers by bioinformatics analysis. *Gene.* (2018) 647:306–11. doi: 10.1016/j.gene.2018.01.003
96. Ahmed F, Benedito VA, Zhao PX. Mining functional elements in messenger RNAs: overview, challenges, and perspectives. *Front Plant Sci.* (2011) 2:84. doi: 10.3389/fpls.2011.00084
97. Ahmed F, Kumar M, Raghava GP. Prediction of polyadenylation signals in human DNA sequences using nucleotide frequencies. *In Silico Biol.* (2009) 9:135–48.
98. Landi D, Gemignani F, Landi S. Role of variations within microRNA-binding sites in cancer. *Mutagenesis*. (2012) 27:205–10. doi: 10.1093/mutage/ger055

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

ORIGINAL RESEARCH
published: 29 October 2019
doi: 10.3389/fonc.2019.01139

Detection of Circulating Tumor Cell Molecular Subtype in Pulmonary Vein Predicting Prognosis of Stage I–III Non-small Cell Lung Cancer Patients

*Jingsi Dong, Daxing Zhu, Xiaojun Tang, Xiaoming Qiu, Dan Lu, Bingjie Li, Dan Lin and Qinghua Zhou**

Department of Lung Cancer Center, West China Hospital, Sichuan University, Chengdu, China

Background: There was rare studies on prognosis of pulmonary venous CTC and early or advanced NSCLC patients. We want to investigate whether CTCs and the subtype of it can predict the prognosis of NSCLC patients.

Patients and Methods: One hundred and fourteen patients with stage I-III NSCLC were included CanPatrol^{TM} CTC analysis. PD-L1 expression level were detected in CTC of pulmonary vein. PD-L1, number of CTC in pulmonary, CTC's subtype, clinical characteristics, prognosis of patients were analyzed.

Results: 110/114 (96.5%) patients could be found CTCs in pulmonary vein, 58/114 (50.9%) patients had CTC≥15/ml in pulmonary vein, 53/110 patients (48.2%) were defined as having MCTC subtype and 56/110 patient were found have PD-L1 (+) CTC in pulmonary vein. Multivariate analyses showed that PVCTC, MCTC, and stage were independent factors of DFS ($P < 0.05$). No OS difference was found between number of CTC ($P = 0.33$) and other CTC factors ($P > 0.05$), only stage was independent factor of OS ($P = 0.019$). There were decreases of CTC number and MCTC number in EGFR mutant subgroup ($P = 0.0009$ and $P = 0.007$). There were increases of CTC ($P = 0.0217$), MCTC ($P = 0.0041$), and PD-L1 (+) CTC ($P = 0.0002$) number in KRAS mutant subgroup. There was increase of MCTC (P =0.0323) number in BRAF mutant. There were fewer CTCs in pulmonary vein for patients with EGFR mutant than in patients with full wild-type gene (P = 0.0346). There were more PD-L1 positive CTCs in pulmonary vein for patients with ALK rearrangement, KRAS mutant, BRAF mutant, or ROS1 mutant than in patients with full wild-type gene ($P = 0.0610$, $P = 0.0003$, $P = 0.032$, and $P = 0.0237$). There were more mesenchymal CTCs in pulmonary vein for patients with KRAS mutant and BRAF mutant than in patients with full wild-type gene ($P = 0.073$ and $P = 0.0381$). There were fewer mesenchymal CTCs in pulmonary vein for patients with EGFR mutant than in patients with full wild-type gene ($P = 0.0898$).

Conclusions: The patients with high number of CTCs, MCTCs, or PD-L1 (+) CTCs in pulmonary vein experienced poor prognosis of DFS. There are obvious correlations between the CTC subtype of NSCLC and the gene subgroups of tumor tissue.

Keywords: pulmonary vein, CTC, non-small cell lung cancer, PD-L1, EMT

OPEN ACCESS

Edited by:
Etienne Giroux Leprieur, Service de Pneumologie et d'Oncologie Thoracique, Hôpital Ambroise-Paré, France

Reviewed by:
Laura Mezquita, Institut Gustave Roussy, France
Jessica Desiree Menis, Istituto Oncologico Veneto (IRCCS), Italy

****Correspondence:***
Qinghua Zhou
zhouqh135@163.com

Specialty section:
This article was submitted to Thoracic Oncology, a section of the journal Frontiers in Oncology

Received: *18 July 2019*
Accepted: *11 October 2019*
Published: *29 October 2019*

Citation:
Dong J, Zhu D, Tang X, Qiu X, Lu D, Li B, Lin D and Zhou Q (2019) Detection of Circulating Tumor Cell Molecular Subtype in Pulmonary Vein Predicting Prognosis of Stage I–III Non-small Cell Lung Cancer Patients. Front. Oncol. 9:1139. doi: 10.3389/fonc.2019.01139

INTRODUCTION

Lung cancer, known as a public health problem in the world, is the leading cause of death caused by malignant tumors worldwide. According to Cancer Statistics published in CA, the estimated deaths caused by lung cancer in 2018 number 83,550 for males and 70,500 for females (1). In China, Lung cancer is the leading cause of death of male and female malignancies (2). Non-small cell lung cancer (NSCLC) accounts for 85% of all lung cancers, and post-treatment recurrence and metastasis are the leading cause of death. Despite many advances in treatment, the overall 5-year survival rate for lung cancer is <20% (3).

In 2002, Dunn et al. proposed the theory of immune editing (4). In recent years, pd-l1 inhibitors have been approved for the treatment of advanced non-small cell lung cancer and achieved remarkable results (5, 6). At present, the recognized detection of programmed cell death ligand 1 (PD-L1) was still at the level of tissue samples, and there was rear research attention to PD-L1 detection of circulating tumor cell (CTC), which was the root of tumor metastasis. Many clinical studies on the CTCs of NSCLC have shown the reliability of CTCs as a prognostic indicator (7–9). A study of Europe confirm CTCs as an independent prognostic indicator of progression-free survival and overall survival in advanced NSCLC (10). Wang's study indicates that CTC detection is mainly related to tumor stage, lymph node metastasis and prognosis, and CTC detection is significantly associated with the shortening of progression-free survival (PFS) and overall survival (OS) in NSCLC (11). CTCs had already been considered the leading causes and markers for tumor recurrence and metastasis (7). Study on the relationship between postoperative disease-free survival (DFS) and CTC in pulmonary vein of NSCLC patient is rear. A study showed that CTCs isolated from early stages of lung cancer are predictive of poor prognosis and can be interrogated to determine biomarkers predictive of recurrence (12), but this study included only 36 patient of lung cancer (NSCLC = 35, SCLC = 1) and the stages of patients was not all early (19 patients were stage I, seven were stage II, eight were stage III, and one patient was stage IV disease).

During the dissemination of cancer cells, epithelial cells frequently exhibit a downregulation of epithelial markers and a loss of intercellular junctions (13). The loss of epithelial features is often accompanied by increased expression of mesenchymal genes. This process, described as epithelial-mesenchymal transition (EMT), endows cancer cells with migratory and invasive properties and promotes cancer recurrence (14–16). Although the number of CTC in pulmonary veins is the largest, current studies have not focused on the relationship between molecular subtypes of CTC in pulmonary and prognosis of cancer therapy.

Here in this study, we focused on the relationship between pulmonary venous CTCs (including different CTC molecular subtypes) and postoperative prognosis of patients with stage I-III NSCLC. The relationship between different CTC molecular subtypes in pulmonary veins and tumor molecular subgroups (EGFR, KRAS, ALK, and BRAF) was studied by translational medicine methods.

Abbreviations: CTC, circulating tumor cell; NSCLC, non-small cell lung cancer; EMT, epithelial-mesenchymal transition; WT, wild-type; MCTC, mesenchymal circulating tumor cell; ECTC, epithelial circulating tumor cell; PVCTC, pulmonary vein circulating tumor cell; PFS, progression-free survival; DFS, disease-free survival; OS, overall survival; PD-L1, programmed cell death ligand 1; EGFR, epidermal growth factor receptor; KRAS, kirsten rat sarcoma viral oncogene; BRAF, v-raf murine sarcoma viral oncogene homolog B; ALK, anaplastic lymphoma kinase.

METHODS

Study Design

One hundred and sixty-four non-small cell lung cancer (NSCLC) patients with stage I-III who could receive surgical resection at the West China Hospital of Sichuan University were included in this study from February 2017 to January 2019. One hundred and fourteen patients eventually met the inclusion criteria. The study was approved by the medical ethics committee of Sichuan University. The patient flow is show in **Figure 1**. All the patients were informed of the procedure and signed informed consent. Our report adheres to the REMARK criteria (17). Inclusion criteria: (a) NSCLC patients who received surgery in Lung Cancer Center of West China Hospital; (b) age of more than 18 years old; (c) postoperative pathological stages were stage I to III; (d) patients have complete clinical data and follow-up data.

In this study, all of the patients with lung cancer received conventional thoracotomy. During surgery, the roots of the pulmonary vein were ligated at the proximal end of the heart, and then 5 ml of blood was extracted at the distal end of the pulmonary vein which was shown in **Figure S1**. After the blood was extracted from the pulmonary vein, it was immediately injected into the blood vessel containing EDTA. Next, the proximal end of the pulmonary vein was ligated a second time, and the distal end of the pulmonary vein was ligated finally (**Figure S1**). Routine laboratory analyses were also performed on all patients, with data prospectively collected for age, sex, histological subtype, genotype, ECOG performance status, smoking status, sites of metastasis, treatment received, stage, date of progression, date of death as the previous studies have been published (18).

Mutations in EGFR exons 18 through 21 were examined using a DxS EGFR mutation test kit (Amoy Diagnostics, China). KRAS mutation was analyzed by Sanger sequencing as described. ALK rearrangement was detected by FISH using the Vysis LSI ALK Break Apart FISH Probe (Abbott Molecular, USA) according to the manufacturer's instructions as described (19). Somatic mutation analysis of BRAF and ROS1 was analyzed by SurPlex-xTAG70plex (Surexam, China).

CTC Analysis

CanPatrol^{TM} (Surexam Biotech, Guangzhou, China) was used to identify CTCs in lung adenocarcinoma patients, as previously described (**Figure S2**). PD-L1, EpCAM, CK8, CK18, and CK19, vimentin, and twist gene expression levels from these different cell types were also detected by RNA *in situ* hybridization. The detection method of CTC has been described in detail

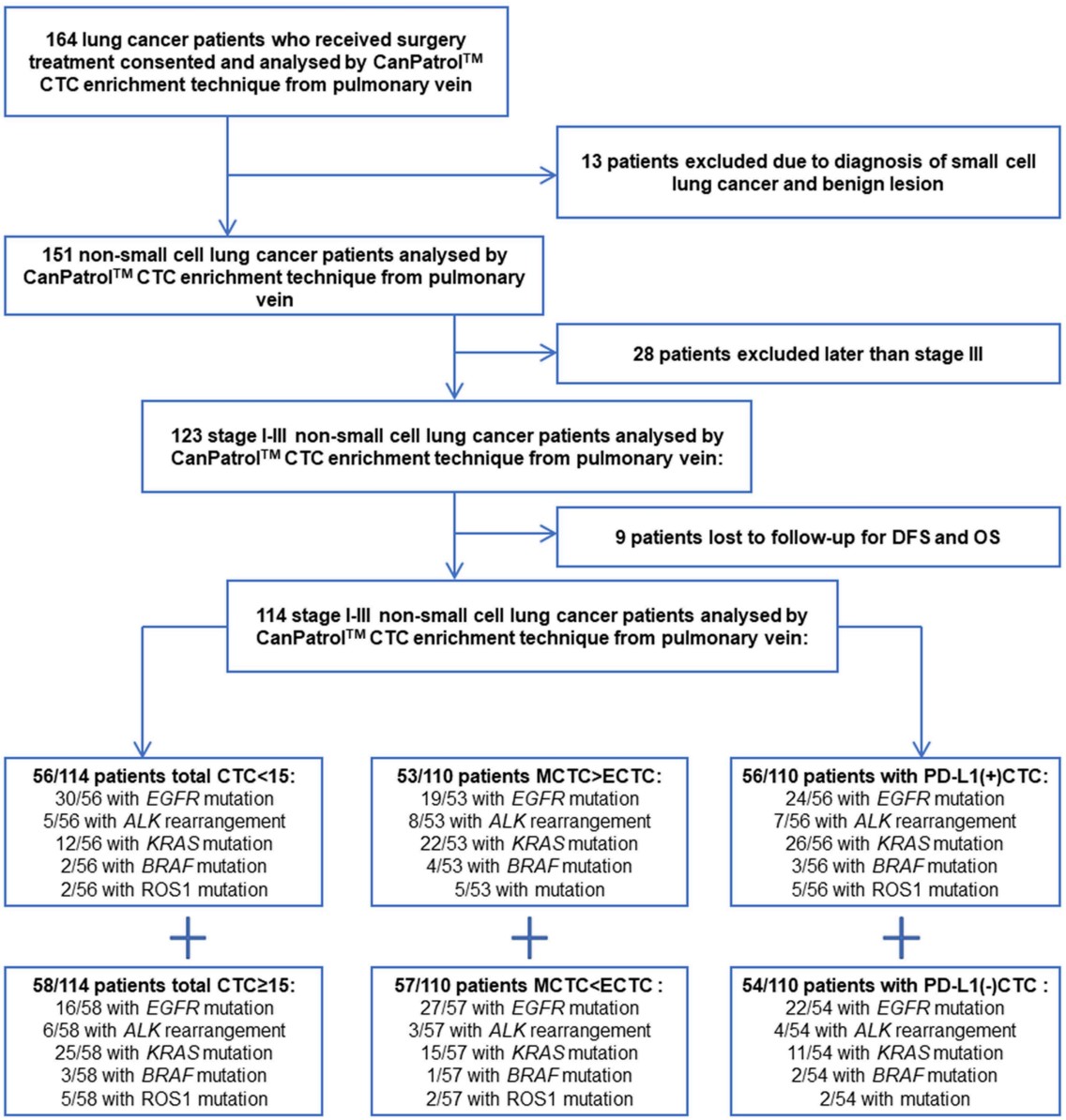


FIGURE 1 | Patient flow.

in our published articles (20). The above markers were used to help distinguish among epithelial, mesenchymal, and hybrid phenotype CTCs. Detection and classification of CTCs using multiple epithelial markers, including EpCAM, CK8/18/19 (red fluorescence) and mesenchymal markers such as Vimentin and Twist (green fluorescence) which were shown in **Figure 2**. The PD-L1 mRNA expression level in CTCs was detected by RNA-ISH (purple fluorescence) which were shown in **Figure 2**.

Statistical Analysis

REMARK guidelines were followed in planning, analysis and reporting of this study. SPSS Statistics 19 software (IBM Deutschland GmbH, Germany) was used for statistical analysis. A $P < 0.05$ was considered a statistically significant difference. GraphPad Prism 6.02 was used for image processing. The survival curve of the OS and DFS of NSCLC patients were plotted by the Kaplan–Meier method after the log-rank test. OS was the time from surgery to death. DFS was the time from surgery to the time of diagnosis of local recurrence, distant metastasis or death, whichever occurred first. The Cox regression model was used for multivariate analysis of all independent influence factors, including the CTC results and other factors, on OS and DFS. Kaplan–Meier curves were computed using GraphPad Prism 6.02. *T*-test was used to compare and analyze continuous variable factors in this study. Chi-squared test was used to analyze the factors of categorical variables. Significance was indicated by the *P*-values of two-tailed tests <0.05.

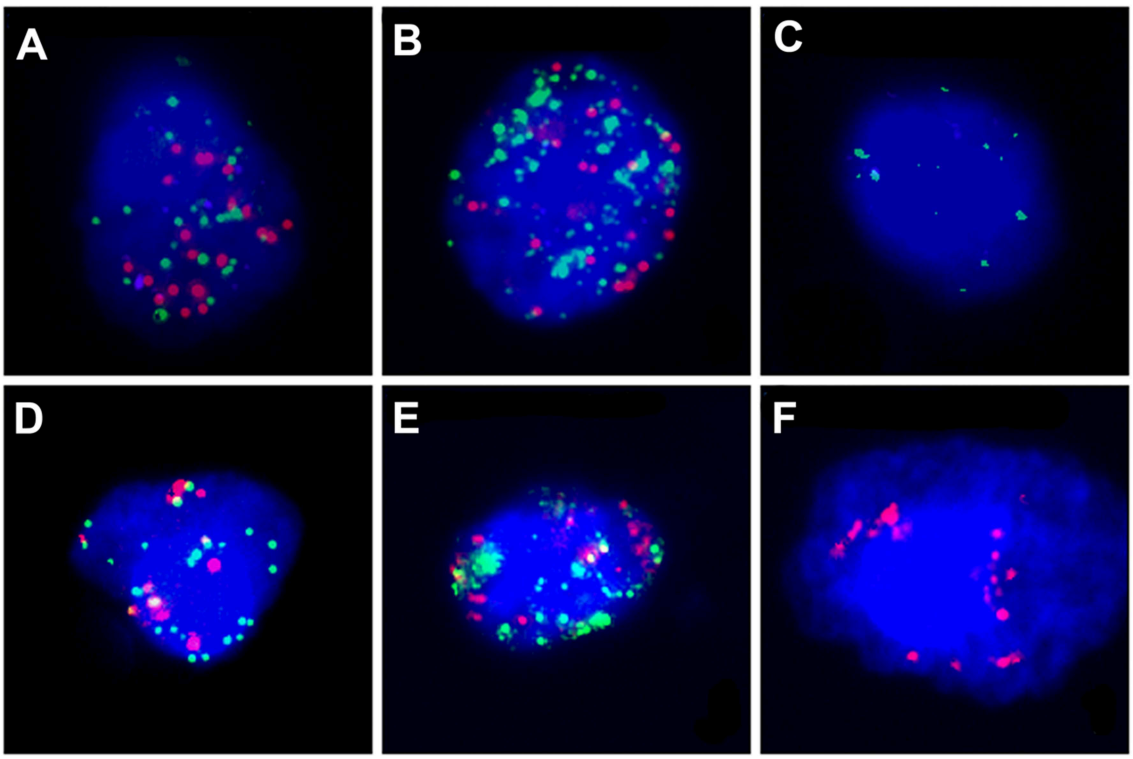


FIGURE 2 | The PD-L1 mRNA expression level and CTC subtypes in CTCs was detected by RNA-ISH. PD-L1: purple fluorescence, epithelial markers: red fluorescence, mesenchymal markers: green fluorescence. **(A)** Hybrid type CTC with PD-L1(++); **(B)** Hybrid type CTC with PD-L1(+); **(C)** Mesenchymal Type CTC with PD-L1(–); **(D)** Mixed type CTC with PD-L1(+); **(E)** Hybrid type CTC with PD-L1(–); **(F)** Epithelial type CTC with PD-L1(–).

RESULTS

Patient Characteristics

One hundred and fourteen patients were included in this study finally as shown in **Figure 1**. Forty-nine patients were female (42.98%), and 65 were male (57.02%). Forty squamous lung cancer patients (35.08%) and 68 patients with adenocarcinoma lung cancer (59.65%) were included in this study, only six patients with other histological types of NSCLC. Fifty-one stage I patients, 21 stage II patients, and 42 stage III patients were included in this study.

CTCs were found in 110 patients' pulmonary veins, only four patients with no CTC can be found. According to the median number of CTC in patients' pulmonary veins, patients were divided into group with CTC≥15 (58/114, 50.9%) and group with CTC<15 (56/114, 49.1%). The Baseline Clinical characteristics was shown in **Table 1.** According to the CTC subtype in patients' pulmonary veins, patients were divided into group with mesenchymal CTC (53/110, 48.2%) and group with non- mesenchymal CTC (epithelial and hybrid subtypes, 57/110, 51.8%). The Baseline Clinical characteristics was shown in **Table S1**. According to whether express PD-L1 in CTC of patients' pulmonary veins or not, patients were divided into group with PD-L1 positive CTC (56/110, 50.1%) and group with PD-L1 negative CTC (54/110, 49.9%). The Baseline Clinical characteristics was shown in **Table S2**.

Clinical Relevance of Total CTCs and CTC Subtype

The median DFS time of the whole group was 20.6 (CI: 18.7–22.5) months (follow-up range: 0–30 months). The median overall survival time of the whole group was 24.3 (CI: 22.5–25.8) months (follow up range: 0–30 months). After follow-up, univariate analyses showed significant reductions in median DFS in CTCs≥15 patients group. The median postoperative DFS was 15.3 (CI: 12.3–18.4) months (range: 0–30 months) in the group with pulmonary vein CTCs≥15 patients, and 24.7 (CI: 22.7–26.7) months (range: 0–30 months) in the group with pulmonary vein CTCs<15 patients ($P < 0.001$). The median postoperative overall survival (OS) was 20.2 (CI: 16.3–24.1) months (range: 0–30 months) in the group with pulmonary vein CTCs≥15 patients, and 25.4 (CI: 23.6–27.3) months (range: 0–24 months) in the group with pulmonary vein CTCs<15 patients ($P = 0.0093$). And there was significant reductions in median DFS in MCTC patients group. The median postoperative DFS was 18.4 (CI: 15.4–21.4) months in the MCTC group, and 22.5 (CI: 20.2–24.8) months in the Non-MCTC group ($P = 0.0168$). And there was no significant difference in OS between the two groups of CTC subtype ($P = 0.4864$). The

TABLE 1 | Baseline clinical characteristics and CTC status of enrolled non-small cell patients.

Characteristic	Total CTC<15 (%)	Total CTC≥15 (%)	*P*-value
Total patient numbers	56/114 (49.1)	58/114 (50.9)	0.791
Age (mean)	59.7	59.2	0.931
Gender			0.685
Female	23/56 (41.1)	26/58 (44.8)	
Male	33/56 (58.9)	32/58 (55.2)	
Smoking status (piece*year)	317.9	432.7	
Histology			
Squamous	13/56 (23.2)	15/58 (25.9)	0.743
Adenocarcinoma	42/56 (75.0)	41/58 (70.7)	0.605
Others*	1/56 (1.8)	2/58 (3.4)	0.579
Surgical method			
Lobectomy	39/56 (69.6)	41/58 (70.7)	0.903
Segmentectomy	12/56 (21.4)	7/58 (12.1)	0.180
Sleeve lobectomy	4/56 (7.1)	8/58 (13.8)	0.247
Pneumonectomy	1/56 (1.8)	2/58 (3.4)	0.579
Stage (AJCC 8)			0.361
Stage I-II	42/56 (75.0)	39/58 (67.2)	
Stage III	14/56 (25.0)	19/58 (32.8)	
Performance status (EGOG)			0.262
0–1	54/56 (96.4)	53/58 (91.4)	
2	2/56 (3.6)	5/58 (8.6)	
Adjuvant chemotherapy			0.420
Yes	20/56 (35.7)	25/58 (43.1)	
No	36/56 (64.3)	33/58 (56.9)	
Adjuvant radiotherapy			0.216
Yes	10/56 (17.9)	16/58 (27.6)	
No	46/56 (82.1)	42/58 (72.4)	

Large cell carcinoma.

Kaplan–Meier survival curve of above data was shown in **Figure 3**.

Clinical Relevance of PD-L1 Expression in CTCs and Stage of Patients

After follow-up, univariate analyses showed significant reductions in median DFS in PD-L1 positive CTC patients group. The median postoperative DFS was 16.7 (CI: 13.9–19.3) months (range: 0–30 months) in the group with PD-L1 positive CTC, while 24.5 (CI: 22.4–26.5) months (range: 0–30 months) in the group with PD-L1 negative CTC ($P = 0.0003$). There was no significant difference in OS between the two groups ($P = 0.09$).

For the group of EGFR negative and ALK negative, there were also significant reductions in median postoperative DFS and OS in PD-L1 positive CTC patients group (showed in **Figures 4C,D**, $P < 0.05$). And there was significant reductions in median DFS and OS in stage I-II patients group. The median postoperative DFS was 24.1 (CI: 22.1–26.1) months in the stage I-II patients group, and 14.7 (CI: 11.8–17.5) months in the stage III group ($P < 0.001$). The median postoperative overall survival (OS) was 26.4 (CI: 24.9–27.9) months in the stage I-II patients group, and 19.1 (CI: 16.0–22.0) months in the stage III group ($P = 0.034$).

TABLE 2 | Baseline CTC characteristics of patients with advanced NSCLC according to total, EMT and PD-L1+ CTC status.

Group	*N* (%) total = 114
No. of patient with no CTC in PV	4/114 (0.035)
No. of patient with CTC≥1 in PV	
Total	110/114 (0.965)
MCTC>ECTC	53/114 (46.5)
PD-L1+CTC	56/114 (49.1)
MCTC>ECTC with PD-L1+CTC	27/114 (23.7)
No. of patient with CTC≥15 in PV	58/114 (50.9)
MCTC>ECTC	26/114 (22.8)
PD-L1+CTC	24/114 (21.1)
MCTC>ECTC with PD-L1+CTC	9/114 (7.9)

The Kaplan–Meier survival curve of above data was shown in **Figure 4.**

In multivariate survival analysis, patients' clinical data, tumor's gene type, number of pulmonary venous CTC, and CTC subtype were included stage, mesenchymal CTCs, and the number of pulmonary vein CTCs were the independent factors of DFS, and only stage was independent factors of OS (showed in **Table S3**).

Relationship Between CTC Subtype and NSCLC Gene Subgroup

All of the 114 patients received gene test after surgery using the tissue samples. 46/114 (40.4%) patients have found EGFR mutations, 11/114 (9.6%) patients have found ALK rearrangements, 37/114 (32.5%) patients have found KRAS mutations and 5/114 (4.4%) patients have found BRAF mutations. As shown in **Table 2**, for the patient with PVCTC≥1, there was 53 of 114 patients with MCTC>ECTC, there was 56/114 patients with PD-L1 (+) CTC. For the patient with PVCTC≥15, there was 26 of 114 patients with MCTC>ECTC, there was 24/114 patients with PD-L1 (+) CTC.

For the detection of the EGFR mutant subgroup, there were fewer CTCs in pulmonary vein for patients with EGFR mutant than in patients with wild-type EGFR (EGFR mutant vs. WT: mean 15.3 vs. 23.2, $P = 0.0009$). There were fewer mesenchymal CTCs in pulmonary vein for patients with EGFR mutant than in patients with wild-type EGFR (EGFR mutant vs. WT: mean 6.9 vs. 12.6, $P = 0.0007$). And there were little fewer PD-L1 positive CTCs in pulmonary vein for patients with EGFR mutant than in patients with wild-type EGFR, but the difference was not statistically significant (EGFR mutant vs. WT: median 6.4 vs. 8.8, $P = 0.67$). The data above was shown in **Figures 5A–C**.

For the detection of the ALK rearrangement subgroup, there was no difference of CTCs number in pulmonary vein between patients with ALK rearrangement and patients with wild-type ALK (ALK rearrangement vs. WT: mean 16.8 vs. 19.9, $P = 0.8885$). There was no difference of mesenchymal CTCs number in pulmonary vein between patients with ALK rearrangement and patients with wild-type ALK (ALK rearrangement vs. WT: mean 9.4 vs. 10.1, $P = 0.7459$). And there was also no difference

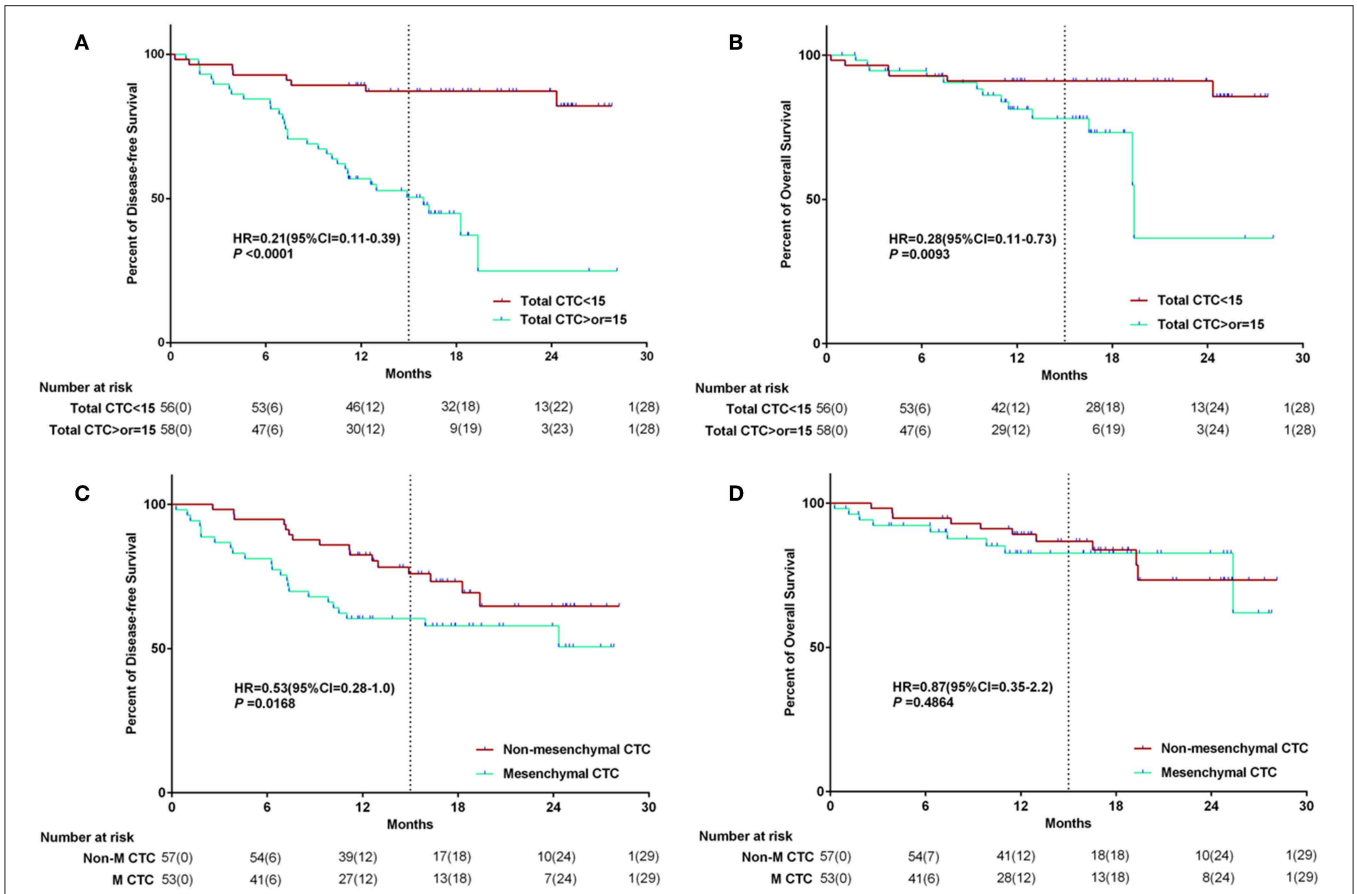


FIGURE 3 | The Kaplan–Meier survival curve of CTC number and MCTC number association with survival. The relationship between CTC quantity and survival **(A,B)**; The relationship between CTC subtype and survival **(C,D)**.

of PD-L1 positive CTCs number in pulmonary vein between patients with ALK rearrangement and patients with wild-type ALK (ALK rearrangement vs. WT: mean 10.1 vs. 7.6, $P = 0.576$). The data above was shown in **Figures 5D–F**.

For the detection of the KRAS mutant subgroup, there were more CTCs in pulmonary vein for patients with KRAS mutant than in patients with wild-type KRAS (KRAS mutant vs. WT: mean 28.1 vs. 16.0, $P = 0.0217$). There were more mesenchymal CTCs in pulmonary vein for patients with KRAS mutant than in patients with wild-type KRAS (KRAS mutant vs. WT: mean 15.2 vs. 7.8, $P = 0.0041$). And there were more PD-L1 positive CTCs in pulmonary vein for patients with KRAS mutant than in patients with wild-type KRAS (KRAS mutant vs. WT: mean 14.8 vs. 4.4, $P = 0.0002$). The data above was shown in **Figures 5G–I**.

For the detection of the BRAF mutant subgroup, there were more CTCs in pulmonary vein for patients with BRAF mutant than in patients with wild-type BRAF, but the difference was not statistically significant (BRAF mutant vs. WT: mean 55.2 vs. 18.6, $P = 0.2229$). There were more mesenchymal CTCs in pulmonary vein for patients with BRAF mutant than in patients with wild-type BRAF, but the difference was not statistically significant (BRAF mutant vs. WT: mean 30.8 vs. 9.5 $P = 0.0323$). And there were more PD-L1 positive CTCs in pulmonary vein for patients with BRAF mutant than in patients with wild-type BRAF, but the difference was not statistically significant (BRAF mutant vs. WT: mean 27.7 vs. 7.5, $P = 0.1934$). The data above was shown in **Figures 5J–L**.

For the detection of the ROS1 mutant subgroup, there were more CTCs in pulmonary vein for patients with ROS1 mutant than in patients with wild-type ROS1, but the difference was not statistically significant (ROS1 mutant vs. WT: mean 25.0 vs. 18.7, $P = 0.2877$). There were more mesenchymal CTCs in pulmonary vein for patients with ROS1 mutant than in patients with wild-type ROS1, but the difference was not statistically significant (ROS1 mutant vs. WT: mean 14.5 vs. 9.9, $P = 0.1154$). And there were more PD-L1 positive CTCs in pulmonary vein for patients with ROS1 mutant than in patients with wild-type ROS1, but the difference was not statistically significant (ROS1 mutant vs. WT: mean 11.3 vs. 7.4, $P = 0.1934$). The data above was shown in **Figures 5M–O**.

Furthermore, we compared the each molecular subgroup with full WT group. As shown in **Figure 6**, for the detection of the EGFR mutant subgroup, there were fewer CTCs in pulmonary vein for patients with EGFR mutant than in patients with full wild-type gene ($P = 0.0346$); There were more PD-L1 positive CTCs in pulmonary vein for patients with ALK rearrangement,

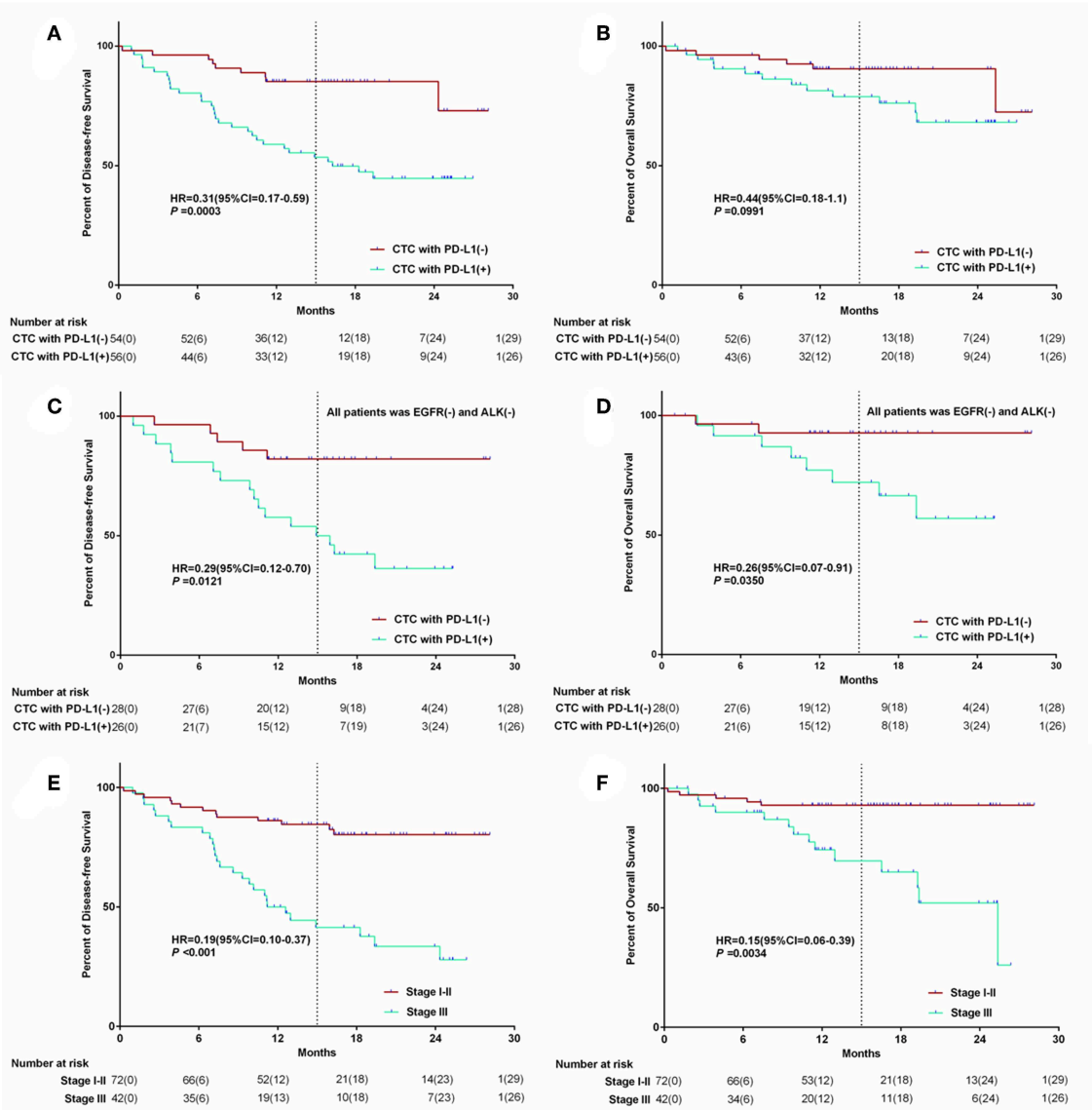


FIGURE 4 | The Kaplan–Meier survival curve of PD-L1 (+) CTC number and stage of patient association with survival. The relationship between PD-L1 expreesion in CTC and survival **(A,B)**; The relationship between PD-L1 expreesion in CTC for the wild type EGFR&ALK patients and survival **(C,D)**; The relationship between Stage and survival **(E,F)**.

KRAS mutant, BRAF mutant, or ROS1 mutant than in patients with full wild-type gene ($P = 0.0610$, $P = 0.0003$, $P = 0.032$, and $P = 0.0237$); There were more mesenchymal CTCs in pulmonary vein for patients with KRAS mutant and BRAF mutant than in patients with full wild-type gene ($P = 0.073$ and $P = 0.0381$); There were fewer mesenchymal CTCs in pulmonary vein for patients with EGFR mutant than in patients with full wild-type gene ($P = 0.0898$).

DISCUSSION

Pulmonary veins are the closest reflux vessels to tumors. Pulmonary veins are the main route for tumor cells to enter the blood from tumor tissues of NSCLC. Okumura et al. found that CTCs in the pulmonary vein were significantly higher than those in peripheral blood (21). Lindsay et al. (18) had reported that PFS and OS are shorter in patients with advanced non-small cell lung cancer whose peripheral venous CTC is >5/ml. A study showed that CTC monitoring after SBRT for presumed early stage NSCLC may give lead-time notice of disease recurrence or progression (22). A study in 2012 demonstrated that CTC test has high sensitivity in early and advanced lung cancer (23). So far, studies that have focused entirely on the relationship between CTC and survival in patients with early non-small cell lung cancer have not been published. In our point of view, the number of tumor cells in peripheral venous blood is not

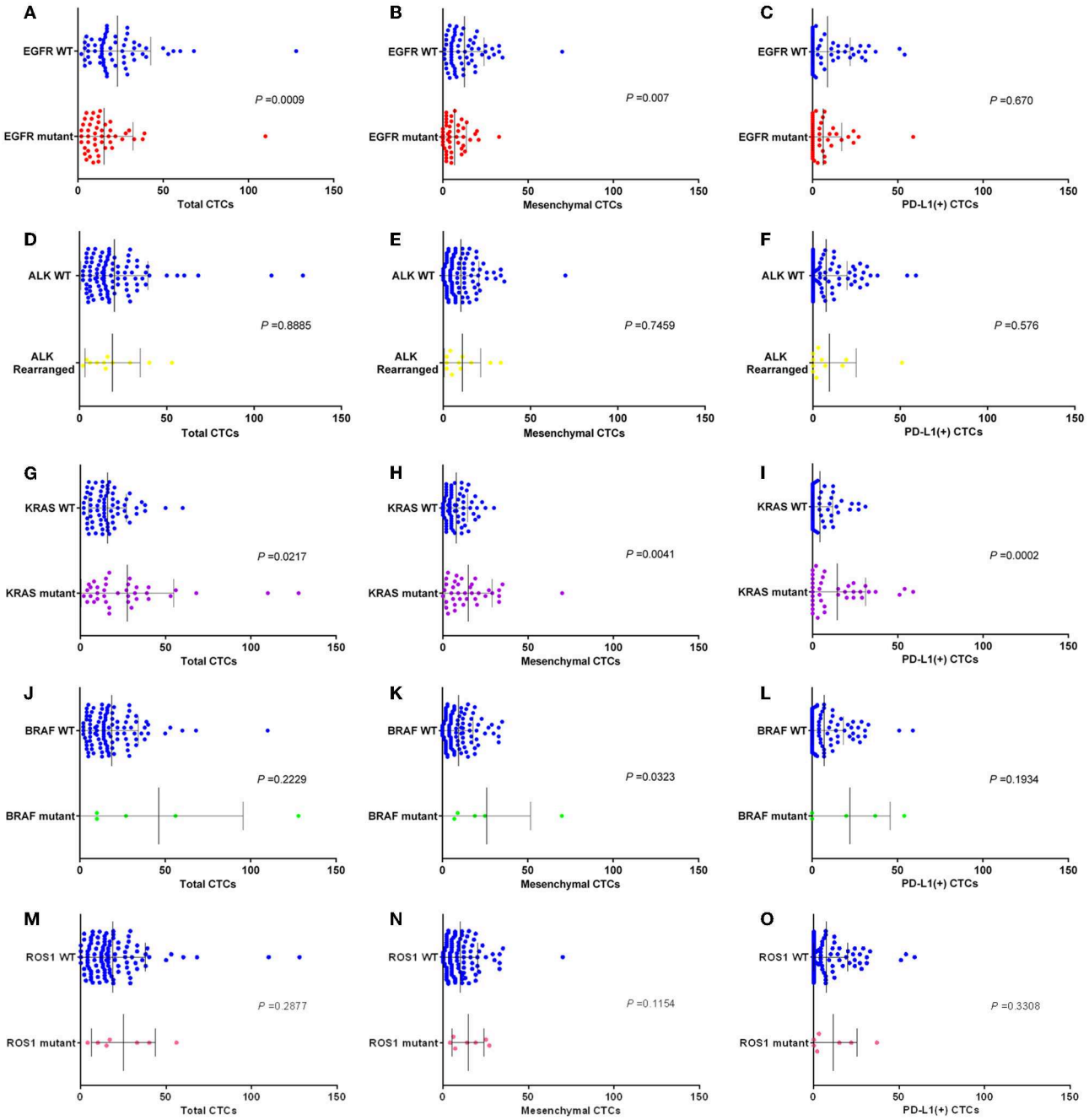


FIGURE 5 | Box plots assessing differences in relative numbers of total CTCs, MCTCs, and PD-L1 (+) CTCs in EGFR. **(A–C)**, ALK **(D–F)**, KRAS **(G–I)**, BRAF **(J–L)**, and ROS1 **(M–O)** subgroups. *P*-values obtained by Mann–Whitney tests.

accurate because peripheral blood first flows through tissue cells and then flows back into venous blood. In this study, we confirmed that the patients with the number of CTC≥15/5 ml in pulmonary veins had the significant shorter postoperative DFS (**Figure 3A**, $P < 0.0001$). Although patients had received multiple comprehensive treatments after postoperative relapse, the OS of patients included in this study was relatively determined by the amount of CTC in pulmonary veins (**Figure 3B**, $P = 0.0093$). All the patients enrolled in this study underwent open-chest surgery, while there was no patient received thoracoscopic surgery. The reason is that open-chest surgery is appropriate for NSCLC patients with stage I to III, while thoracoscopic surgery is commonly considered appropriate for NSCLC patients with stage I to II all over the world.

Although the Cellsearch System has been used in the majority of published studies, it depends on tumor epithelial cell expression of EpCAM, the presence of an intact nucleus, and the absence of CD45 (4, 7, 25–31) (16, 18). However,

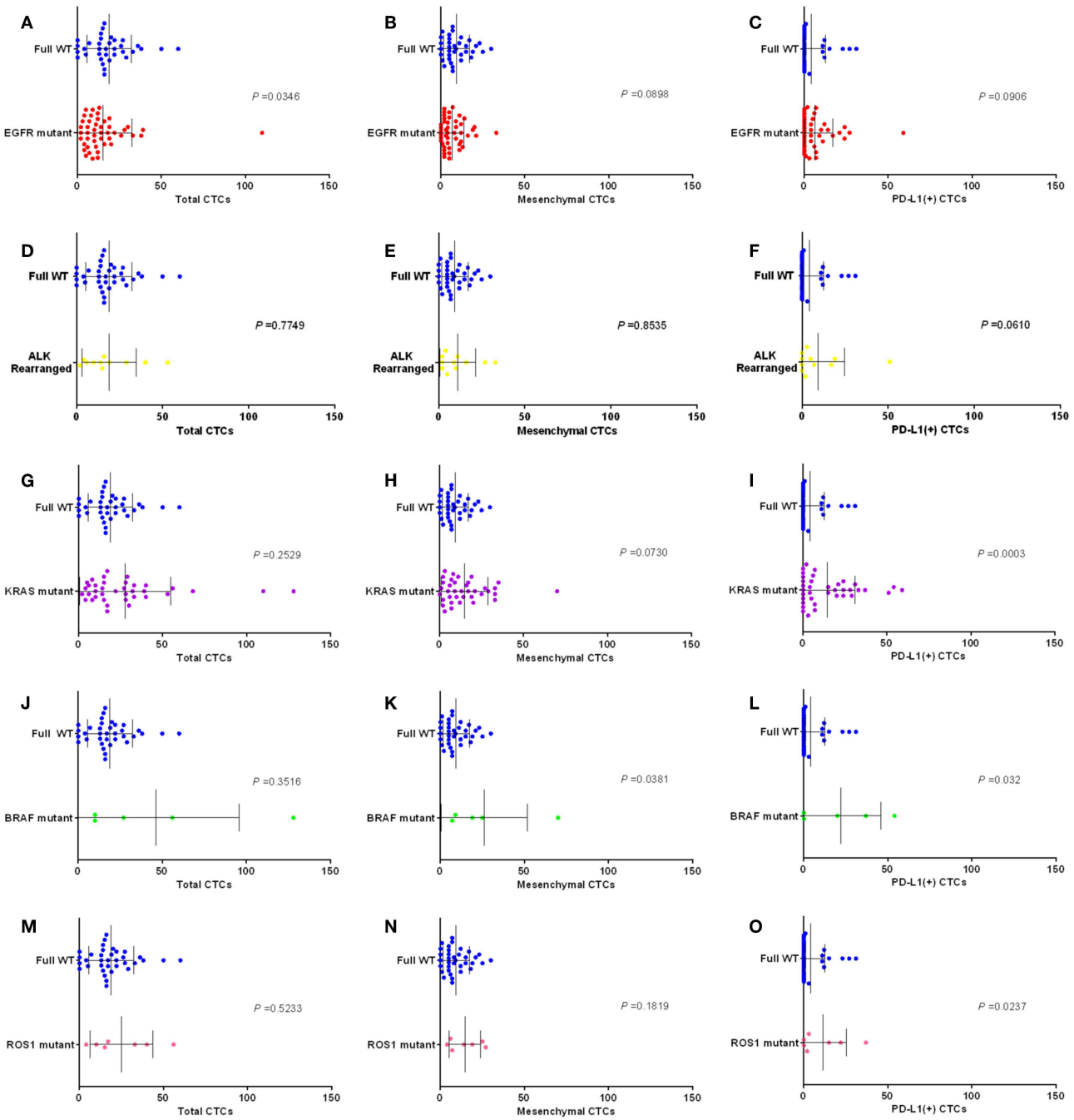


FIGURE 6 | Box plots assessing differences in relative numbers of total CTCs, MCTCs and PD-L1 (+) CTCs in EGFR **(A–C)**, ALK **(D–F)**, KRAS **(G–I)**, BRAF **(J–L)**, and ROS1 **(M–O)** subgroups (mutant vs. full WT). *P*-values obtained by Mann–Whitney tests.

this method lacks the detection of mesenchymal CTC, and the EMT of CTC is easy to be ignored. Recently, Canpatrol™ CTC analysis system was developed to detect CTC and classify EMT phenotypes via multiple mRNA *in situ* hybridization assay, by which revealing that CTC count and EMT classification are correlated with clinical stage and prognosis in many kinds of cancers (24, 25). In this study, we found that postoperative DFS was shorter in patients with mesenchymal CTC predominance than in patients with epithelial and hybrid type (**Figure 3C**, $P = 0.0168$). Different from the effect of tumor cell count on OS, the OS of patients included in this study was not completely determined by subtype of CTC in pulmonary veins (**Figure 3D**, $P = 0.48$).

The PD-L1 will downregulate T-cell activation and promote immune escape when binding with programmed death 1 (PD-1) protein expressed on the T-cell surface (26). Many studies have

found that the expression of PD-L1 in tumors is significantly correlated with the treatment response of nivolumab (27, 28). PD-L1 expression correlated with benefit to immune checkpoint inhibitors, not only nivolumab. Although there were some differences reported in the results of clinical trials in which different PD-1/PD-L1 inhibitors were involved, the benefit of immunotherapy was significantly higher in those with high expression of PD-L1 than in those with low expression (29, 30). Until now, there is no study on detection of PD-L1 in CTC has been performed expect us. Detection of expression of PD-L1 in CTC will be an important supplement to liquid biopsy and of great significance to guide the treatment of cancer. In this study, we not only successfully detected the expression of PD-L1 in CTC, but also found that whether the express of PD-L1 or not was correlated with postoperative DFS and OS in NSCLC patients (**Figure 4A**, $P = 0.0003$; **Figure 4B**, $P = 0.0991$). For the patients with EGFR(–) and ALK(–), the effect of PD-L1 positive in CTC to survival (DFS & OS) was more obvious, as shown in **Figures 4C,D**. This phenomenon may be related to PD-L1-mediated tumor immune avoidance in NSCLC. It is well-known that UICC stage is an important factor determining the prognosis of patients in NSCLC. In this study, we confirmed that postoperative pathological stage of NSCLC patients is obviously related to postoperative DFS and OS (**Figures 4E,F**, $P < 0.05$).

The growth of NSCLC is related to many driver genes, including EGFR, KRAS, BRAF, ALK, and so on. NSCLC tumors have different genetic subgroups, even with the same pathological type (29, 31). In this study, we found that there were differences in the number of CTC, the number of mesenchymal CTC, and the number of PD-L1 positive CTC among patients with different NSCLC genotypes (**Figure 5**). The relationship between information of PD-L1 in tissue and PD-L1 in CTC was published by us in previous study (20). And we found that positive PD-L1 in CTC was positively correlated with positive PD-L1 in tissues. So in this study, we included new group of patients to do many more survival analyses and focus on the relationship between information of PD-L1 in CTC and gene type of tissues. Within tumor cells, the network of multiple driver genes plays an important role in tumor growth and immune escape. This just shows that different genotypes of tumor surface lead to different types of CTC into the blood.

CONCLUSIONS

In conclusion, we found that pulmonary venous examination is a more reliable method for analyzing CTC in NSCLC patients receiving surgical treatment. In addition, the detection of PD-L1 expression in CTC may provide an important decision for post-operative adjuvant immunotherapy. In subsequent studies, we will focus on the impact of different conditions of CTC on postoperative adjuvant therapy effect, so as to provide more information for individualized treatment.

DATA AVAILABILITY STATEMENT

All datasets for this study are included in the manuscript/**Supplementary Material**.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the medical ethics committee of Sichuan University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

JD and QZ designed the study. JD, BL, and DLi wrote the manuscript. DZ and XT supported the methods of the study. XQ and DLu supported the literature and statistical methods.

FUNDING

This work was supported by Health Commission of Sichuan Province (18PJ432) and National Natural Science Foundation of China (81572288). The National Key Research and Development Program is the Key Project of International Scientific and Technological Innovation Cooperation between Governments (2016YEE0103400).

ACKNOWLEDGMENTS

This study was partly supported by grants from the Key Project of International Cooperation of Science and Technology Innovation between Governments, the National Key Research and Development Plan of China (No. 2016YEE0103400), and the National Natural Science Foundation of China (No. 81572288). The authors would like to thank all the patients who took the time to participate in this research at a time when they had other priorities.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2019.01139/full#supplementary-material

REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. *CA Cancer J Clin*. (2018) 68:7–30. doi: 10.3322/caac.21442
2. Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, et al. Cancer statistics in China, 2015. *CA Cancer J Clin*. (2016) 66:115–32. doi: 10.3322/caac.21338
3. Torre LA, Siegel RL, Jemal A. Lung cancer statistics. *Adv Exp Med Biol*. (2016) 893:1–19. doi: 10.1007/978-3-319-24223-1_1

4. Dunn GP, Bruce AT, Ikeda H, Old LJ, Schreiber RD. Cancer immunoediting: from immunosurveillance to tumor escape. *Nat Immunol*. (2002) 3:991–8. doi: 10.1038/ni1102-991
5. Giroux Leprieur E, Dumenil C, Julie C, Giraud V, Dumoulin J, Labrune S, et al. Immunotherapy revolutionises non-small-cell lung cancer therapy: results, perspectives and new challenges. *Eur J Cancer*. (2017) 78:16–23. doi: 10.1016/j.ejca.2016.12.041
6. Gandhi L, Rodriguez-Abreu D, Gadgeel S, Esteban E, Felip E, De Angelis F, et al. Pembrolizumab plus chemotherapy in metastatic non-small-cell lung cancer. *N Engl J Med*. (2018) 378:2078–92. doi: 10.1056/NEJMoa1801005
7. Krebs MG, Sloane R, Priest L, Lancashire L, Hou JM, Greystoke A, et al. Evaluation and prognostic significance of circulating tumor cells in patients with non-small-cell lung cancer. *J Clin Oncol*. (2011) 29:1556–63. doi: 10.1200/JCO.2010.28.7045
8. Aggarwal C, Wang X, Ranganathan A, Torigian D, Troxel A, Evans T, et al. Circulating tumor cells as a predictive biomarker in patients with small cell lung cancer undergoing chemotherapy. *Lung Cancer*. (2017) 112:118–25. doi: 10.1016/j.lungcan.2017.08.008
9. Syrigos K, Fiste O, Charpidou A, Grapsa D. Circulating tumor cells count as a predictor of survival in lung cancer. *Crit Rev Oncol Hematol*. (2018) 125:60–8. doi: 10.1016/j.critrevonc.2018.03.004
10. Lindsay CR, Blackhall FH, Carmel A, Fernandez-Gutierrez F, Gazzaniga P, Groen HJM, et al. EPAC-lung: pooled analysis of circulating tumour cells in advanced non-small cell lung cancer. *Eur J Cancer*. (2019) 117:60–8. doi: 10.1016/j.ejca.2019.04.019
11. Wang X, Ma K, Yang Z, Cui J, He H, Hoffman AR, et al. Systematic correlation analyses of circulating tumor cells with clinical variables and tumor markers in lung cancer patients. *J Cancer*. (2017) 8:3099–104. doi: 10.7150/jca.18032
12. Murlidhar V, Reddy RM, Fouladdel S, Zhao L, Ishikawa MK, Grabauskiene S, et al. Poor prognosis indicated by venous circulating tumor cell clusters in early-stage lung cancers. *Cancer Res.* (2017) 77:5194–206. doi: 10.1158/0008-5472.CAN-16-2072
13. Chen Y, Li S, Li W, Yang R, Zhang X, Ye Y, et al. Circulating tumor cells undergoing EMT are poorly correlated with clinical stages or predictive of recurrence in hepatocellular carcinoma. *Sci Rep*. (2019) 9:7084. doi: 10.1038/s41598-019-43572-1
14. Kalluri R, Weinberg RA. The basics of epithelial-mesenchymal transition. *J Clin Investig*. (2009) 119:1420–8. doi: 10.1172/JCI39104
15. Thiery JP, Acloque H, Huang RY, Nieto MA. Epithelial-mesenchymal transitions in development and disease. *Cell*. (2009) 139:871–90. doi: 10.1016/j.cell.2009.11.007
16. Qi LN, Xiang BD, Wu FX, Ye JZ, Zhong JH, Wang YY, et al. Circulating tumor cells undergoing EMT provide a metric for diagnosis and prognosis of patients with hepatocellular carcinoma. *Cancer Res*. (2018) 78:4731–44. doi: 10.1158/0008-5472.CAN-17-2459
17. Sauerbrei W, Taube SE, McShane LM, Cavenagh MM, Altman DG. Reporting recommendations for tumor marker prognostic studies (REMARK): an abridged explanation and elaboration. *J Natl Cancer Inst*. (2018) 110:803–11. doi: 10.1093/jnci/djy088
18. Lindsay CR, Faugeroux V, Michiels S, Pailler E, Facchinetti F, Ou D, et al. A prospective examination of circulating tumor cell profiles in non-small-cell lung cancer molecular subgroups. *Ann Oncol*. (2017) 28:1523–31. doi: 10.1093/annonc/mdx156
19. Wang Z, Yang JJ, Huang J, Ye JY, Zhang XC, Tu HY, et al. Lung adenocarcinoma harboring EGFR T790M and in Trans C797S responds to combination therapy of first- and third-generation EGFR TKIs and shifts allelic configuration at resistance. *J Thorac Oncol*. (2017) 12:1723–7. doi: 10.1016/j.jtho.2017.06.017
20. Dong J, Zhu D, Tang X, Lu D, Qiu X, Li B, et al. Circulating tumor cells in pulmonary vein and peripheral arterial provide a metric for PD-L1 diagnosis and prognosis of patients with non-small cell lung cancer. *PLoS ONE*. (2019) 14:e0220306. doi: 10.1371/journal.pone.0220306
21. Okumura Y, Tanaka F, Yoneda K, Hashimoto M, Takuwa T, Kondo N, et al. Circulating tumor cells in pulmonary venous blood of primary lung cancer patients. *Ann Thorac Surg*. (2009) 87:1669–75. doi: 10.1016/j.athoracsur.2009.03.073
22. Frick MA, Kao GD, Aguarin L, Chinniah C, Swisher-McClure S, Berman AT, et al. Circulating tumor cell assessment in presumed early stage non-small cell lung cancer patients treated with stereotactic body radiation therapy: a prospective pilot study. *Int J Radiat Oncol Biol Phys*. (2018) 102:536–42. doi: 10.1016/j.ijrobp.2018.06.041
23. Wendel M, Bazhenova L, Boshuizen R, Kolatkar A, Honnatti M, Cho EH, et al. Fluid biopsy for circulating tumor cell identification in patients with early-and late-stage non-small cell lung cancer: a glimpse into lung cancer biology. *Phys Biol*. (2012) 9:016005. doi: 10.1088/1478-3967/9/1/016005
24. Liu YK, Hu BS, Li ZL, He X, Li Y, Lu LG. An improved strategy to detect the epithelial-mesenchymal transition process in circulating tumor cells in hepatocellular carcinoma patients. *Hepatol Int*. (2016) 10:640–6. doi: 10.1007/s12072-016-9732-7
25. Si Y, Lan G, Deng Z, Wang Y, Lu Y, Qin Y, et al. Distribution and clinical significance of circulating tumor cells in nasopharyngeal carcinoma. *Jpn J Clin Oncol*. (2016) 46:622–30. doi: 10.1093/jjco/hyw046
26. Pardoll DM. The blockade of immune checkpoints in cancer immunotherapy. *Nat Rev Cancer*. (2012) 12:252–64. doi: 10.1038/nrc3239
27. Borghaei H, Brahmer J, Horn L, Ready N, Steins M, Felip E, et al. P2.35: nivolumab vs. docetaxel in advanced NSCLC: CheckMate 017/057 2-Y update and exploratory cytokine profile analysis: track: immunotherapy. *J Thorac Oncol*. (2016) 11:S237–S8. doi: 10.1016/j.jtho.2016.08.106
28. Horn L, Spigel DR, Vokes EE, Holgado E, Ready N, Steins M, et al. Nivolumab versus docetaxel in previously treated patients with advanced non-small-cell lung cancer: two-year outcomes from two randomized, open-label, phase III trials (CheckMate 017 and CheckMate 057). *J Clin Oncol*. (2017) 35:3924–33. doi: 10.1200/JCO.2017.74.3062
29. Dong J, Li B, Lin D, Zhou Q, Huang D. Advances in targeted therapy and immunotherapy for non-small cell lung cancer based on accurate molecular typing. *Front Pharmacol*. (2019) 10:230. doi: 10.3389/fphar.2019.00230
30. Dong J, Li B, Zhou Q, Huang D. Advances in evidence-based medicine for immunotherapy of non-small cell lung cancer. *J Evid Based Med*. (2018) 11:278–87. doi: 10.1111/jebm.12322
31. Duruisseaux M, Mc Leer-Florin A, Moro-Sibilot D, Cadranel J. Are ALK rearrangement variants promising predictive biomarker of ALK tyrosine kinase inhibitors efficacy? *Ann Oncol*. (2017) 28:1401. doi: 10.1093/annonc/mdx116

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

ORIGINAL RESEARCH
published: 05 November 2019
doi: 10.3389/fonc.2019.01123

Prognostic Impact of Metabolism Reprogramming Markers Acetyl-CoA Synthetase 2 Phosphorylation and Ketohexokinase-A Expression in Non-Small-Cell Lung Carcinoma

Xueying Yang[1], Fei Shao[1,2,3], Susheng Shi[4], Xiaoli Feng[4], Wei Wang[1], Yalong Wang[1], Wei Guo[1], Juhong Wang[1], Shugeng Gao[1], Yibo Gao[1], Zhimin Lu[5]* and Jie He[1]**

[1] Department of Thoracic Surgery, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China, [2] Cancer Institute, The Affiliated Hospital of Qingdao University, Qingdao, China, [3] Qingdao Cancer Institute, Qingdao, China, [4] Department of Pathology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China, [5] Zhejiang Provincial Key Laboratory of Pancreatic Disease, The First Affiliated Hospital, Institute of Translational Medicine, Zhejiang University School of Medicine, Hangzhou, China

OPEN ACCESS

Edited by:
Etienne Giroux Leprieur,
Hôpital Ambroise-Paré, France

Reviewed by:
Cancan Zhou,
First Affiliated Hospital of Xi'an Jiaotong University, China
Yuping Sun,
Jinan Central Hospital, China

***Correspondence:**
Jie He
prof.jiehe@gmail.com
Zhimin Lu
zhiminlu@zju.edu.cn
Yibo Gao
gaoyibo@cicams.ac.cn

Specialty section:
This article was submitted to Thoracic Oncology, a section of the journal Frontiers in Oncology

Received: 19 July 2019
Accepted: 09 October 2019
Published: 05 November 2019

Citation:
Yang X, Shao F, Shi S, Feng X, Wang W, Wang Y, Guo W, Wang J, Gao S, Gao Y, Lu Z and He J (2019) Prognostic Impact of Metabolism Reprogramming Markers Acetyl-CoA Synthetase 2 Phosphorylation and Ketohexokinase-A Expression in Non-Small-Cell Lung Carcinoma. Front. Oncol. 9:1123. doi: 10.3389/fonc.2019.01123

Background: The identification of prognostic markers for non-small-cell lung carcinoma (NSCLC) is needed for clinical practice. The metabolism-reprogramming marker ketohexokinase (KHK)-A and acetyl-CoA synthetase 2 (ACSS2) phosphorylation at S659 (ACSS2 pS659) play important roles in tumorigenesis and tumor development. However, the clinical significance of KHK-A and ACSS2 pS659 in NSCLC is largely unknown.

Methods: The expression levels of KHK-A and ACSS2 pS659 were assessed by immunohistochemistry analyses of surgical specimens from 303 NSCLC patients. The prognostic values of KHK-A and ACSS2 pS659 were evaluated by Kaplan–Meier methods and Cox regression models.

Results: The expression levels of KHK-A and ACSS2 pS659 were significantly higher in NSCLC tissues than those in adjacent non-tumor tissues ($P < 0.0001$). KHK-A or ACSS2 pS659 alone and the combination of KHK-A and ACSS2 pS659 were inversely correlated with overall survival in NSCLC patients ($P < 0.001$). The multivariate analysis indicated that KHK-A or ACSS2 pS659 and KHK-A/ACSS2 pS659 were independent prognostic biomarkers for NSCLC ($P = 0.008$ for KHK-A, $P < 0.001$ for ACSS2 pS659, and $P < 0.001$ for KHK-A/ACSS2 pS659). Furthermore, the combination of KHK-A and ACSS2 pS659 can be used as a prognostic indicator for all stages of NSCLC.

Conclusions: KHK-A or ACSS2 pS659 alone and the combination of KHK-A and ACSS2 pS659 can be used as prognostic markers for NSCLC. Our findings highlight the important role of metabolic reprogramming in NSCLC progression.

Keywords: metabolism reprogramming, KHK-A, ACSS2 pS659, non-small-cell lung carcinoma, prognosis, immunohistochemistry

INTRODUCTION

Lung cancer is becoming the leading cause of cancer-related death worldwide (1). Non-small-cell lung carcinoma (NSCLC) is the most common type of lung cancer, accounting for ∼85% of all cases (2). Due to the high rates of metastasis, recurrence, and drug resistance of NSCLC, its prognosis remains poor (3). There is an urgent need for novel biomarkers to identify a subset of patients with poor survival outcomes.

The reprogramming of energy metabolism is one of the hallmarks of cancer (4). During reprograming, cancer cell metabolism and other cellular activities are integrated and mutually regulated (5). Recent studies have shown that metabolic enzymes, such as ketohexokinase (KHK)-A and acetyl-CoA synthetase 2 (ACSS2), are moderated spatially and temporally in cancer cells so that these enzymes not only have changes in metabolic activities but also gain non-canonical functions (5).

KHK initiates fructose catabolism by catalyzing the transfer of a phosphate group from adenosine triphosphate (ATP) to fructose to produce AMP and fructose 1-phosphate (F1P) (6). F1P is then metabolized to dihydroxyacetone phosphate and glyceraldehyde-3-phosphate to bypass important glycolytic regulatory steps in glycolysis and enter the latter stages of glycolysis (7). KHK-A and KHK-C are splicing isoforms of KHK with one exon difference (8). KHK-C is expressed mainly in the liver, intestines, and kidney, whereas KHK-A is ubiquitously expressed at low levels (9). Although fructose can be metabolized by both KHK-C and KHK-A, the primary enzyme involved in fructose metabolism is considered to be KHK-C rather than KHK-A due to its K_m (10). We recently reported that a splicing switch from KHK-C to KHK-A occurs in hepatocellular carcinoma (HCC), leading to fructose metabolism reduction, and KHK-A phosphorylates and activates phosphoribosyl pyrophosphate synthetase 1 (PRPS1), resulting in increased *de novo* nucleic acid synthesis for HCC development (6). Under oxidative stress, KHK-A dissociates from PRPS1 and phosphorylates p62 to activate Nrf2, and activated Nrf2 induces gene expression to counteract oxidative stress and promote HCC development in mice (11). Notably, high KHK-A expression predicted a poor prognosis for HCC patients (6). Thus, KHK-A reprograms HCC cell metabolism and other cellular activities by reducing fructose metabolism and increasing *de novo* nucleic acid synthesis and the antioxidative stress response by the protein kinase activity of KHK-A. An important remaining question is whether KHK-A plays an important role in cancers other than HCC.

Histone lysine acetylation is essential for regulating chromatin architecture and promoting transcription (12). In mammalian cells, acetyl coenzyme A (acetyl-CoA) is a necessary acetyl donor for lysine acetylation and can be produced by three enzymes: ATP-citrate lyase (ACL), the pyruvate dehydrogenase complex (PDC), and acetyl-CoA synthetase (ACSS) (13–15). In nutrient-rich environments, acetyl-CoA is primarily produced by ACL (13), and growth signals promote PDC-dependent acetyl-CoA production (14). In tumors, metabolic stress frequently occurs. Our previous study revealed that AMP-activated protein kinase (AMPK) can mediate ACSS2 phosphorylation at S659 (ACSS2 pS659) to induce its nuclear translocation in a glucose-deficient environment, and the binding of ACSS2 to the promoter regions of lysosomal and autophagy genes can promote acetyl-CoA production to support histone acetylation and gene expression to promote tumor development (16). Collectively, these results suggest that ACSS2 pS659 plays an important role in tumor metabolism reprogramming through its nuclear function. However, whether ACSS2 pS659 expression is a biomarker for the clinical features and prognosis of cancer is unknown.

In this study, we examined the expression of KHK-A and ACSS2 pS659 in human NSCLC specimens and the relationship between their abundance and clinical relevance in a large cohort of surgically resected NSCLCs. We found that both KHK-A and ACSS2 pS659 are independent prognostic factors for NSCLC patients after surgery, and the combination of KHK-A and ACSS2 pS659 can be used as a prognostic indicator for all stages of NSCLC.

MATERIALS AND METHODS

Patients and Specimens

We enrolled a total of 303 consecutive patients diagnosed with NSCLC, including 227 with lung adenocarcinoma (LUAD) and 76 with lung squamous cell carcinoma (LUSC), by pathological examination at the National Cancer Center/Cancer Hospital in Chinese Academy of Medical Sciences. Patients were diagnosed

TABLE 1 | Patient characteristics ($N = 303$).

Characteristics	Number of patients (%)
Gender	
Male	206 (68.0%)
Female	97 (32.0%)
Age (years)	
≤60	132 (43.6%)
>60	171 (56.4%)
Histology	
Adenocarcinoma	227 (74.9%)
Squamous cell carcinoma	76 (25.1%)
T stage	
I	63 (20.8%)
II	141 (46.5%)
III	62 (20.5%)
IV	37 (12.2%)
Node metastasis	
No	153 (50.5%)
Yes	150 (49.5%)
TNM stage	
I	84 (27.7%)
II	89 (29.4%)
III	122 (40.3%)
IV	8 (2.6%)

TNM stage, tumor-node-metastasis stage.

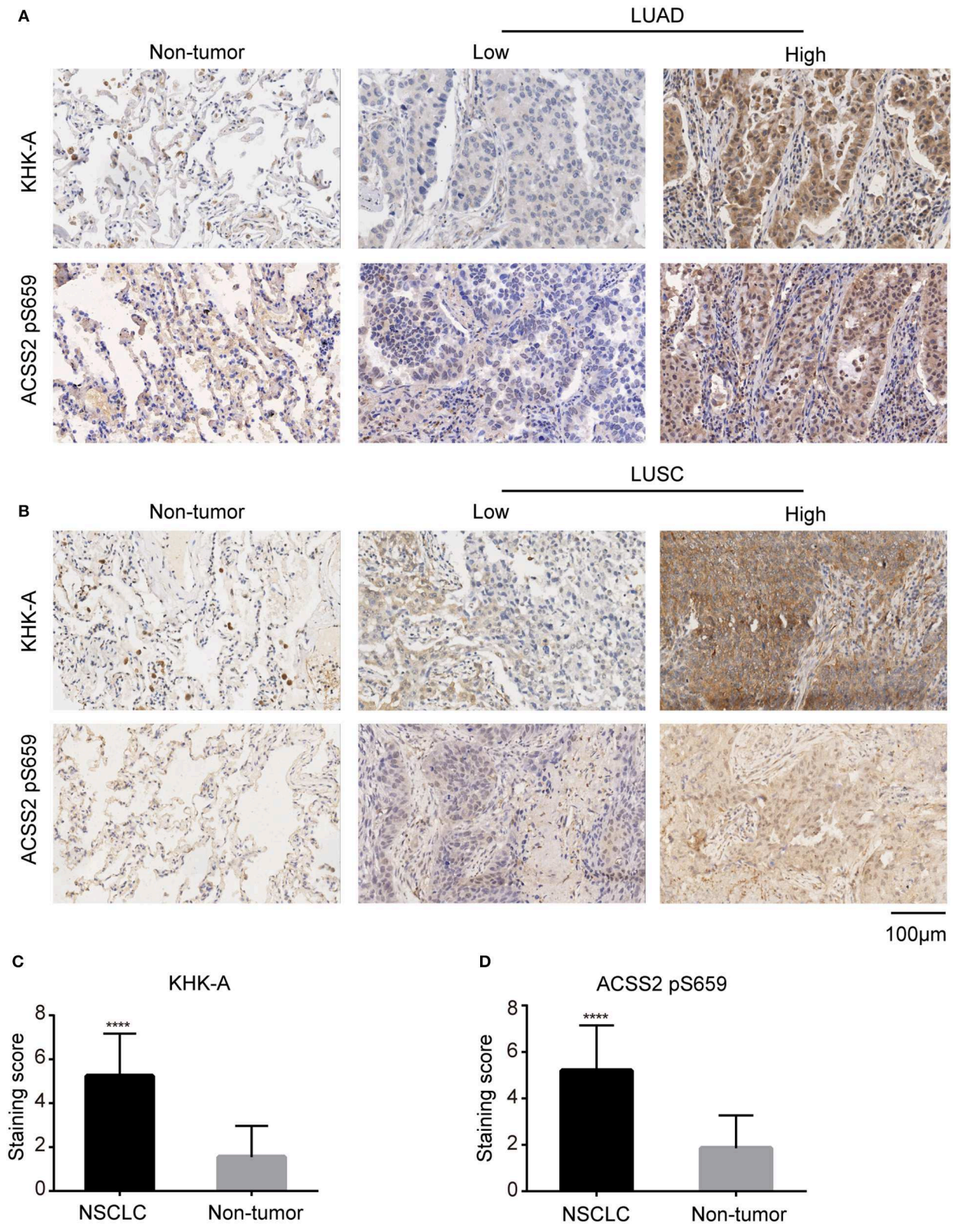


FIGURE 1 | NSCLC specimens have increased KHK-A and ACSS2 pS659 expression levels. **(A,B)** Representative IHC staining of low and high expression of KHK-A and ACSS2 pS659 in NSCLC tissues and adjacent non-tumor tissues ($N = 303$). **(A)** LUAD, **(B)** LUSC. Scale bar, 200×, 100 μm. **(C,D)** The expression levels of KHK-A **(C)** and ACSS2 pS659 **(D)** in NSCLC and adjacent non-tumor tissues were compared by IHC staining. ****Correlation is significant at the 0.0001 level (two-tailed).

with NSCLC and were without preoperative chemotherapy, radiotherapy, and distant metastasis. All paired tumor and adjacent non-tumor tissues used in this study were collected in compliance with an informed consent policy. This study was approved by the Ethics Committee of the National Cancer Center/Cancer Hospital, Chinese Academy of Medical Sciences, and Peking Union Medical College.

We obtained clinical data by reviewing the patients' medical histories, which are summarized in **Table 1**. Pathological staging was assessed by the 8th edition of the American Joint Committee

TABLE 2 | Relationship of KHK-A and ACSS2 pS659 expression levels with patient characteristics.

Characteristics	KHK-A expression level		*P*-value	ACSS2 pS659 expression level		*P*-value
	Low	High		Low	High	
Gender						
Male	149	57	0.191	114	92	0.034
Female	63	34		41	56	
Age (years)						
≤60	95	37	0.504	78	54	0.015
>60	117	54		77	94	
T stage						
I+II	146	58	0.383	110	94	0.167
III+IV	66	33		45	54	
Node metastasis						
No	114	39	0.081	82	71	0.391
Yes	98	52		73	77	
TNM stage						
I+II	127	46	0.131	97	76	0.048
III+IV	85	45		58	72	
Histology						
Adenocarcinoma	147	80	0.001	108	119	0.031
Squamous cell carcinoma	65	11		47	29	

TNM stage, tumor-node-metastasis stage.

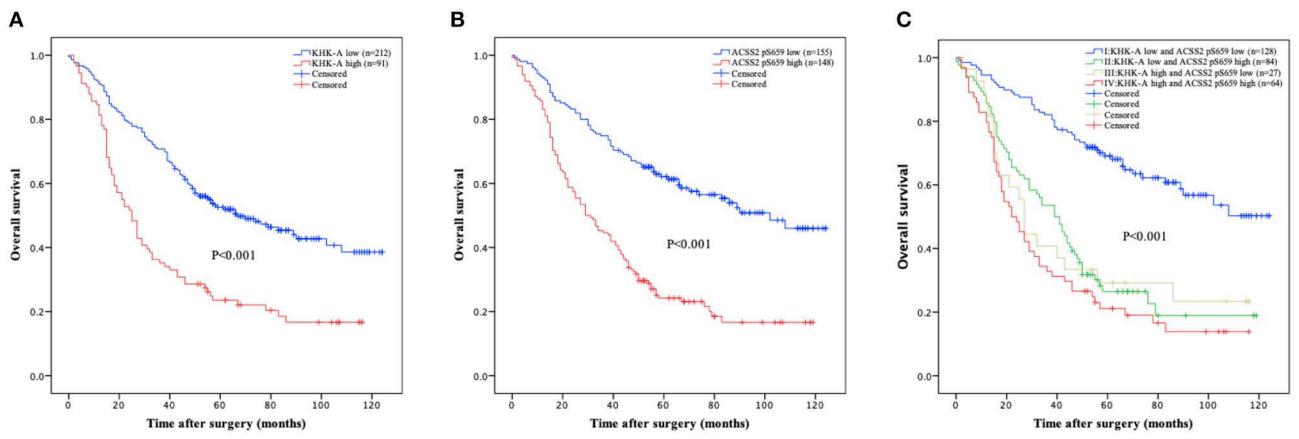


FIGURE 2 | Prognostic value of KHK-A and ACSS2 pS659 expression in NSCLC. **(A,B)** The value of the IHC score was used to divide the indicated NSCLC patients into two groups with high and low levels of KHK-A **(A)** and ACSS2 pS659 **(B)** expression. Kaplan–Meier survival curves were compared using the log-rank test. All statistical tests were two-sided. Crosses represent censored data from patients who were alive at the last clinical follow-up. **(C)** IHC scores of KHK-A and ACSS2 pS659 were used to divide the NSCLC patients into four subgroups (I, KHK-A low expression, and ACSS2 pS659 low expression; II, KHK-A low expression, and ACSS2 pS659 high expression; III, KHK-A high expression, and ACSS2 pS659 low expression; IV, KHK-A high expression, and ACSS2 pS659 high expression). Kaplan–Meier survival curves were compared using the log-rank test. All statistical tests were two-sided. Crosses represent censored data from patients who were alive at the last clinical follow-up.

on Cancer/Union for International Cancer Control TNM classification system (17). We obtained completed follow-up information for all patients, and the time from the date of diagnosis to death or the last known date of follow-up was defined as overall survival (OS).

Tissue Microarray and Immunohistochemistry

Prior to tissue microarray (TMA) construction, hematoxylin, and eosin (H&E)-stained slides were evaluated by two pathologists independently, and high tumor/stroma ratio areas were marked in the paraffin-embedded specimens. Two core regions (1 mm in diameter) were extracted from each marked tumor tissue and peritumoral tissue to construct the TMA slides, and consecutive sections measuring 4 μm were placed on adhesive slides (18). Immunohistochemistry (IHC) analyses of the paraffin sections were performed as previously described (19). A rabbit polyclonal antibody recognizing KHK-A was obtained from Signalway Biotechnology (Pearland, TX, Cat. #21708) and was diluted to 1:50. An anti-ACSS2 pS659 rabbit polyclonal antibody was obtained from Jiaxing Xinda Biological Technology and was diluted to 1:50. The specificities of these antibodies were previously validated (6, 16).

Evaluation of the Immunohistochemical Findings

IHC scoring of KHK-A and ACSS2 pS659 was based on the percentage of positive cells and the staining intensity. The proportion scores were defined as follows: 0, 0% of positive cells; 1, 0–1%; 2, 2–10%; 3, 11–30%; 4, 31–70%; and 5, 71–100%. The staining intensity was rated as follows: 0, negative; 1, weak; 2, moderate; and 3, strong. A total score (range: 0, 2–8) was then obtained by combining the proportion and intensity scores, as described previously (20). IHC scoring was assessed independently by two experienced pathologists who did not know the clinical information of the patients. The cutoff values for high and low expression were based on the values of the IHC scores. In this study, the expression of KHK-A was classified as low (score ≤6) or high (score >6), and the expression of ACSS2-pS659 was classified as low (score ≤5) or high (score >5).

Statistical Analysis

We used SPSS 21.0 (IBM, Armonk, NY) to perform statistical analyses. The association between marker expression and clinical factors was analyzed by the chi-squared (χ^2) test. The survival analyses were performed by the Kaplan–Meier method (log-rank test). We used univariate Cox regression to calculate the risk factors for progress, and the risk factors were then included in a multivariate Cox regression model to identify the independent prognostic factors. $P < 0.05$ was considered statistically significant. All statistical tests were two-sided.

RESULTS

NSCLC Specimens Have Increased KHK-A and ACSS2 pS659 Expression Levels

We performed immunohistochemical (IHC) staining of NSCLC specimens ($N = 303$), including LUAD (**Figure 1A**) and LUSC (**Figure 1B**) tissues. We showed that KHK-A was primary in the cytoplasm of the NSCLC cells and that ACSS2 pS659 was observed in both nucleus and cytoplasm of the NSCLC cells (**Figures 1A,B**). In addition, the expression levels of KHK-A and ACSS2 pS659 were significantly higher in NSCLC tissues than those in adjacent non-tumor tissues (**Figures 1A–D**), indicating

TABLE 3 | Univariate and multivariate analyses of overall survival for 303 NSCLC.

Factor	Univariate *P*-value	Multivariate		
		Hazard ratio	95% CI	*P*-value
Gender (female vs. male)	0.930			NA
Age (>60 vs. ≤60)	0.065			NA
T stage (I vs. II vs. III vs. IV)	0.017			0.811
Node metastasis (yes vs. no)	<0.001	2.003	1.486–2.700	<0.001
Histology (squamous vs. adenocarcinoma)	0.001	0.65	0.434–0.975	0.037
KHK-A (high vs. low)	<0.001	1.533	1.120–2.099	0.008
ACSS2 pS659 (high vs. low)	<0.001	2.313	1.687–3.172	<0.001

CI, confidence interval; NA, not adopted.

TABLE 4 | Univariate and multivariate analyses of overall survival for 303 NSCLC.

Factor	Univariate *P*-value	Multivariate		
		Hazard ratio	95% CI	*P*-value
Gender (female vs. male)	0.930			NA
Age (>60 vs. ≤60)	0.065			NA
T stage (I vs. II vs. III vs. IV)	0.017			0.823
Node metastasis (yes vs. no)	<0.001	1.953	1.446–2.637	<0.001
Histology (squamous vs. adenocarcinoma)	0.001	0.661	0.441–0.992	0.046
Combination of KHK-A and ACSS2 pS659	<0.001			<0.001
II vs. I	<0.001	2.803	1.920–4.094	<0.001
III vs. I	<0.001	2.319	1.366–3.936	0.002
IV vs. I	<0.001	3.587	2.413–5.331	<0.001

CI, confidence interval; NA, not adopted; I, KHK-A^{Low}/ACSS2 pS659^{Low}; II, KHK-A^{Low}/ACSS2 pS659^{High}; III, KHK-A^{High}/ACSS2 pS659^{Low}; IV, KHK-A^{High}/ACSS2 pS659^{High}.

that NSCLC specimens have increased KHK-A and ACSS2 pS659 expression levels.

KHK-A and ACSS2 pS659 Expression Levels Are Correlated With Clinical Features in Patients With NSCLC

We examined the relationship between KHK-A and ACSS2 pS659 expression levels and clinical features in NSCLC. KHK-A expression levels were higher in LUAD than those in LUSC ($P = 0.001$). However, there were no statistical correlations between KHK-A and other clinical features, including gender, age, T classification, N classification, or TNM stage (all $P > 0.05$, **Table 2**). Notably, ACSS2 pS659 expression levels were higher in LUAD than those in LUSC ($P = 0.031$), and higher in female patients than in males ($P = 0.034$). In addition, high ACSS2 pS659 expression was associated with older age ($P = 0.015$) and advanced TNM stage ($P = 0.048$).

KHK-A and ACSS2 pS659 Expression Levels Are Inversely Correlated With the OS of Patients With NSCLC

We next analyzed whether KHK-A and ACSS2 pS659 expression levels in NSCLC patients have any relationship with disease prognosis. As shown in **Figure 2A**, high KHK-A expression levels predicted a poor 5-years OS rate (high vs. low: 23.5 vs. 52.5%, $P < 0.001$). In addition, high ACSS2 pS659 expression levels were also correlated with a poor 5-years OS rate (**Figure 2B**; high vs. low: 24.2 vs. 62.1%, $P < 0.001$). To analyze the prognostic value of KHK-A and ACSS2 pS659 co-expression, we divided the patients into four subgroups using the IHC scores of KHK-A and ACSS2 pS659: I, KHK-A low expression and ACSS2 pS659 low expression; II, KHK-A low expression, and ACSS2 pS659 high expression; III, KHK-A high expression and ACSS2 pS659 low expression; IV, KHK-A high expression, and ACSS2 pS659 high expression. We observed that OS was significantly different among the four subgroups (**Figure 2C**, $P < 0.001$) and that high expression of both KHK-A and ACSS2 pS659 appeared to have the worst prognosis with the lowest survival rates (5-years OS: 21.1%). In contrast, patients with low tumor expression of both KHK-A and ACSS2 pS659 had the best prognosis with the highest OS rate (5-years OS: 69.1%).

A univariate analysis was performed to investigate the risk factors for OS. As shown in **Table 3**, T stage, node metastasis, histology, KHK-A, and ACSS2 pS659 were significantly associated with OS ($P < 0.05$). Statistically significant variables were added to multivariate survival analyses. In the Cox regression model, node metastasis, histology, KHK-A expression, and ACSS2 pS659 levels were independent prognostic factors of OS [hazard ratio (HR) = 2.003, 95% confidence interval (95% CI) = 1.486–2.700 for node metastasis; HR = 0.65, 95% CI = 0.434–0.795 for histology; HR = 1.533, 95% CI = 1.120–2.099 for KHK-A expression; and HR = 2.313, 95% CI = 1.687–3.172 for ACSS2 pS659]. To evaluate the prognostic value of combined KHK-A and ACSS2 pS659 expression in NSCLC, we considered the expression of KHK-A and ACSS2 pS659 as a single factor for a separate multivariate analysis. **Table 4** shows that the combination of KHK-A and ACSS2 pS659 was an independent predictor of OS (HR = 2.803, 95% CI = 1.920–4.094 for II vs. I; HR = 2.319, 95% CI = 1.366–3.936 for III vs. I; and HR = 3.587, 95% CI = 2.413–5.331 for IV vs. I).

Furthermore, we analyzed the relationship between combination of KHK-A and ACSS2 pS659 and prognosis from TNM stage I to IV. Low combined values of KHK-A and ACSS2 pS659 expression were associated with improved OS in stages I (**Figure 3A**, $P < 0.001$), II (**Figure 3B**, $P = 0.003$), and III–IV (**Figure 3C**, $P < 0.001$). These results indicate that the combined expression of KHK-A and ACSS2 pS659 is inversely correlated with OS in all stages of NSCLC.

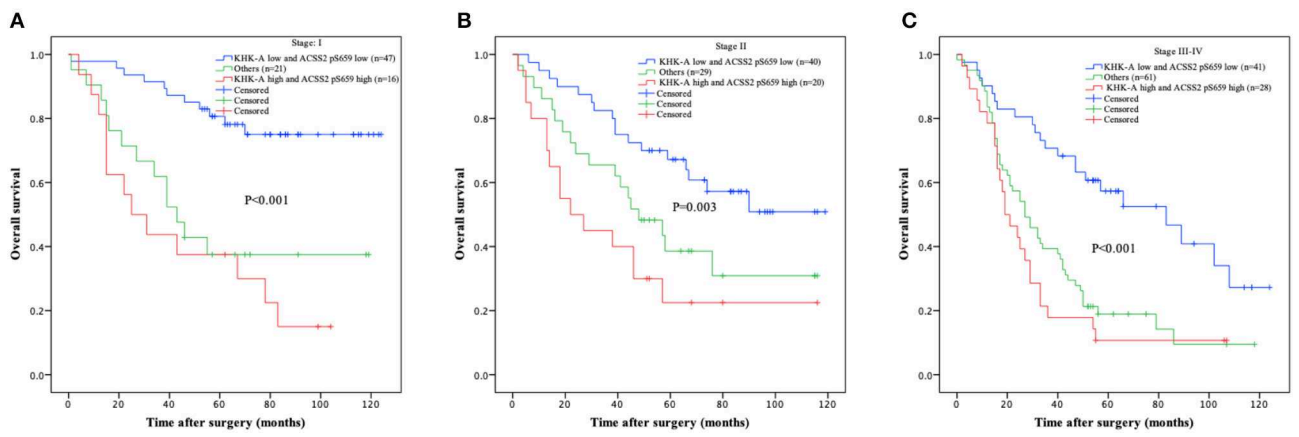


FIGURE 3 | Prognostic value of combined KHK-A and ACSS2 pS659 expression in NSCLC based on TNM stage. **(A–C)** IHC scores of KHK-A and ACSS2 pS659 were used to divide the NSCLC patients into three subgroups (KHK-A low expression and ACSS2 pS659 low expression, KHK-A high expression and ACSS2 pS659 high expression, and others: KHK-A low expression and ACSS2 pS659 high expression or KHK-A high expression and ACSS2 pS659 low expression) in stage I **(A)**, stage II **(B)**, and stage III–IV **(C)**. Kaplan–Meier survival curves were compared using the log-rank test. All statistical tests were two-sided. Crosses represent censored data from patients who were alive at the last clinical follow-up.

DISCUSSION

Although tremendous clinical success in the treatment of NSCLC during the past few decades has been achieved, the prognosis of patients with NSCLC remains unsatisfactory. We still need to explore biological characteristics that can reflect NSCLC behavior. In this study, we showed that the expression levels of metabolism reprogramming marker KHK-A and ACSS2 pS659 were significantly higher in NSCLC tissues than those in adjacent non-tumor tissues, and KHK-A expression levels were higher in LUAD than those in LUSC. In addition, increased ACSS2 pS659 expression was correlated with older age, advanced TNM stage, LUAD, and female patients. Of note, we showed that high expression levels of KHK-A and ACSS2 pS659 were correlated with reduced OS in NSCLC. In the multivariate analysis, both KHK-A and ACSS2 pS659 expression levels were identified as independent prognostic biomarkers for NSCLC. These findings suggest that KHK-A and ACSS2 pS659 are potential prognostic indicators for NSCLC.

Our previous work demonstrated that KHK-A and ACSS2 are two key metabolic enzymes that have important roles in tumor development. KHK-A can act as a protein kinase to directly phosphorylate PRPS1 to increase *de novo* nucleic acid synthesis for hepatocellular tumorigenesis (6). In addition to its role in nucleic acid synthesis, KHK-A phosphorylates p62, leading to the activation of Nrf2 and the expression of its downstream genes to counteract oxidative stress and support HCC development (11). Reports on the role of ACSS2 expression in cancer have been controversial. High ACSS2 expression predicted a poor prognosis in patients with renal cell carcinoma and bladder cancer (21, 22), whereas low ACSS2 expression predicted a poor prognosis in gastric cancer and hepatocellular carcinoma (23, 24). Our previous study showed that the phosphorylation of ACSS2 at S659 rather than its expression level is critical for GBM cells to counteract energy stress. S659-phosphorylated ACSS2 translocates to the nucleus, leading to the binding to downstream genes and acetyl-CoA production, which induces histone acetylation and gene expression for glioma development (16, 25). Although KHK-A and ACSS2 pS659 have been studied in HCC and GBM, respectively, their roles in NSCLC had not yet been explored. We showed here that the combination of KHK-A and ACSS2 pS659 was an independent prognostic biomarker for NSCLC and was inversely related with OS in all stages of NSCLC. Given that multiple randomized clinical trials have suggested that postoperative NSCLC patients with a poor prognosis are more likely to benefit from adjuvant therapy (26), the combination of KHK-A and ACSS2 pS659 expression levels can be taken into consideration for decisions regarding the use of adjuvant therapy for NSCLC, especially in the early stage.

CONCLUSIONS

In summary, KHK-A or ACSS2 pS659 alone and the combination of KHK-A and ACSS2 pS659 can be used as prognostic markers for NSCLC. Our findings highlight the importance of metabolic reprogramming in the clinical behavior of NSCLC and reveal that KHK-A and ACSS2 pS659 can be targeted for NSCLC treatment.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/supplementary material.

ETHICS STATEMENT

All samples used in this study were collected with signed informed consent, and this study was approved by the Ethics Committee of the National Cancer Center/Cancer Hospital, Chinese Academy of Medical Sciences, and Peking Union Medical College.

AUTHOR CONTRIBUTIONS

XY, YG, ZL, and JH conceived and designed the study and carried out data interpretation. XY wrote the manuscript with help from YG, ZL, and JH. SG, WG, and JW collected tissue samples. XY, WW, and JW performed the immunohistochemical assays. SS and XF evaluated and scored the staining. XY, FS, and YW acquired data and performed the statistical analysis. All authors read and approved the final manuscript.

FUNDING

This study was supported by the National Key R&D Program of China (2017YFC1308702, 2018YFC1312100, and 2017YFC1311000), the CAMS Initiative for Innovative Medicine (2017-I2M-1-005, 2017-I2M-2-003, and 2019-I2M-2-002), the Beijing Municipal Science & Technology Commission (Z181100001918002), and the Non-profit Central Research Institute Fund of Chinese Academy of Medical Sciences (2017PT32001 and 2017PT32017).

REFERENCES

1. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA Cancer J Clin*. (2015) 65:87–108. doi: 10.3322/caac.21262
2. Molina JR, Yang P, Cassivi SD, Schild SE, Adjei AA. Non-small cell lung cancer: epidemiology, risk factors, treatment, and survivorship. *Mayo Clinic Proceedings*. (2008) 83:584–94. doi: 10.1016/S0025-6196(11)60735-0
3. Liu M, Zhang Y, Zhang J, Cai H, Zhang C, Yang Z, et al. MicroRNA-1253 suppresses cell proliferation and invasion of non-small-cell lung carcinoma by targeting WNT5A. *Cell Death Dis*. (2018) 9:189. doi: 10.1038/s41419-017-0218-x
4. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell*. (2011) 144:646–74. doi: 10.1016/j.cell.2011.02.013
5. Wang Y, Xia Y, Lu Z. Metabolic features of cancer cells. *Cancer Commun*. (2018) 38:65. doi: 10.1186/s40880-018-0335-7
6. Li X, Qian X, Peng LX, Jiang Y, Hawke DH, Zheng Y, et al. A splicing switch from ketohexokinase-C to ketohexokinase-A drives hepatocellular carcinoma formation. *Nat Cell Biol*. (2016) 18:561–71. doi: 10.1038/ncb3338

7. Ishimoto T, Lanaspa MA, Le MT, Garcia GE, Diggle CP, Maclean PS, et al. Opposing effects of fructokinase C and A isoforms on fructose-induced metabolic syndrome in mice. *Proc Natl Acad Sci USA*. (2012) 109:4320–5. doi: 10.1073/pnas.1119908109
8. Hayward BE, Bonthron DT. Structure and alternative splicing of the ketohexokinase gene. *Eur J Biochem*. (1998) 257:85–91. doi: 10.1046/j.1432-1327.1998.2570085.x
9. Diggle CP, Shires M, Leitch D, Brooke D, Carr IM, Markham AF, et al. Ketohexokinase: expression and localization of the principal fructose-metabolizing enzyme. *J Histochem Cytochem*. (2009) 57:763–74. doi: 10.1369/jhc.2009.953190
10. Asipu A, Hayward BE, O'Reilly J, Bonthron DT. Properties of normal and mutant recombinant human ketohexokinases and implications for the pathogenesis of essential fructosuria. *Diabetes*. (2003) 52:2426–32. doi: 10.2337/diabetes.52.9.2426
11. Xu D, Li X, Shao F, Lv G, Lv H, Lee JH, et al. The protein kinase activity of fructokinase A specifies the antioxidant responses of tumor cells by phosphorylating p62. *Sci Adv*. (2019) 5:eaav4570. doi: 10.1126/sciadv.aav4570
12. Li B, Carey M, Workman JL. The role of chromatin during transcription. *Cell*. (2007) 128:707–19. doi: 10.1016/j.cell.2007.01.015
13. Wellen KE, Hatzivassiliou G, Sachdeva UM, Bui TV, Cross JR, Thompson CB. ATP-citrate lyase links cellular metabolism to histone acetylation. *Science*. (2009) 324:1076–80. doi: 10.1126/science.1164097
14. Sutendra G, Kinnaird A, Dromparis P, Paulin R, Stenson TH, Haromy A, et al. A nuclear pyruvate dehydrogenase complex is important for the generation of acetyl-CoA and histone acetylation. *Cell*. (2014) 158:84–97. doi: 10.1016/j.cell.2014.04.046
15. Zachary Schug T, Peck B, Dylan Jones T, Zhang Q, Grosskurth S, Israt Alam S, et al. Acetyl-CoA synthetase 2 promotes acetate utilization and maintains cancer cell growth under metabolic stress. *Cancer Cell*. (2015) 27:57–71. doi: 10.1016/j.ccell.2014.12.002
16. Li X, Yu W, Qian X, Xia Y, Zheng Y, Lee JH, et al. Nucleus-translocated ACSS2 promotes gene transcription for lysosomal biogenesis and autophagy. *Mol Cell*. (2017) 66:684–97 e9. doi: 10.1016/j.molcel.2017.04.026
17. Detterbeck FC. The eighth edition TNM stage classification for lung cancer: what does it mean on main street? *J Thor Cardiov Surg*. (2018) 155:356–9. doi: 10.1016/j.jtcvs.2017.08.138
18. Kononen J, Bubendorf L, Kallioniemi A, Barlund M, Schraml P, Leighton S, et al. Tissue microarrays for high-throughput molecular profiling of tumor specimens. *Nat Med*. (1998) 4:844–7. doi: 10.1038/nm0798-844
19. Guo W, Zhang F, Shao F, Wang P, Li Z, Yang X, et al. PD-L1 expression on tumor cells associated with favorable prognosis in surgically resected esophageal squamous cell carcinoma. *Hum Pathol*. (2019) 84:291–8. doi: 10.1016/j.humpath.2018.09.014
20. Yang W, Xia Y, Ji H, Zheng Y, Liang J, Huang W, et al. Nuclear PKM2 regulates beta-catenin transactivation upon EGFR activation. *Nature*. (2011) 480:118–22. doi: 10.1038/nature10598
21. Zhang S, He J, Jia Z, Yan Z, Yang J. Acetyl-CoA synthetase 2 enhances tumorigenesis and is indicative of a poor prognosis for patients with renal cell carcinoma. *Urol Oncol*. (2018) 36:243.e9–43.e20. doi: 10.1016/j.urolonc.2018.01.013
22. Lee MY, Yeon A, Shahid M, Cho E, Sairam V, Figlin R, et al. Reprogrammed lipid metabolism in bladder cancer with cisplatin resistance. *Oncotarget*. (2018) 9:13231–43. doi: 10.18632/oncotarget.24229
23. Hur H, Kim YB, Ham IH, Lee D. Loss of ACSS2 expression predicts poor prognosis in patients with gastric cancer. *J Surg Oncol*. (2015) 112:585–91. doi: 10.1002/jso.24043
24. Sun L, Kong Y, Cao M, Zhou H, Li H, Cui Y, et al. Decreased expression of acetyl-CoA synthase 2 promotes metastasis and predicts poor prognosis in hepatocellular carcinoma. *Cancer Sci*. (2017) 108:1338–46. doi: 10.1111/cas.13252
25. Zlotorynski E. Gene expression: ACSS2 boosts local histone acetylation. *Nat Rev Mol Cell Biol*. (2017) 18:405. doi: 10.1038/nrm.2017.61
26. Douillard Y-J, Rosell R, De Lena M, Carpagnano F, Ramlau R, Gonzáles-Larriba JL, et al. Adjuvant vinorelbine plus cisplatin versus observation in patients with completely resected stage IB–IIIA non-small-cell lung cancer (Adjuvant Navelbine International Trialist Association [ANITA]): a randomised controlled trial. *Lancet Oncol*. (2006) 7:719–27. doi: 10.1016/S1470-2045(06)70804-X

Conflict of Interest: ZL owns shares in Signalway Biotechnology (Pearland, TX), which supplied the rabbit antibody that recognizes KHK-A. ZL's interest in this company had no bearing on its being chosen to supply this reagent.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

ORIGINAL RESEARCH
published: 13 December 2019
doi: 10.3389/fonc.2019.01372

A Strong Decrease in *TIMP3* Expression Mediated by the Presence of *miR-17* and *20a* Enables Extracellular Matrix Remodeling in the NSCLC Lesion Surroundings

Karolina H. Czarnecka[1,2], Bartosz Szmyd[1], Magda Barańska[1], Marcin Kaszkowiak[1], Jacek Kordiak[3], Adam Antczak[4], Dorota Pastuszak-Lewandoska[5] and Ewa Brzeziańska-Lasota[1]*

[1] Department of Biomedicine and Genetics, Medical University of Lodz, Łódz, Poland, [2] Quantitative Genomic Medicine Laboratories, S.L., Esplugues de Llobregat, Barcelona, Spain, [3] Department of Chest Surgery, General and Oncological Surgery, University Teaching Hospital No. 2, Medical University of Lodz, Łódz, Poland, [4] Department of General and Oncological Pulmonology, Medical University of Lodz, Łódz, Poland, [5] Department of Microbiology and Laboratory Medical Immunology, Medical University of Lodz, Łódz, Poland

OPEN ACCESS

Edited by:
Etienne Giroux Leprieur, Hôpital Ambroise-Paré, France

Reviewed by:
Jose Gilberto Bazan, Ohio State University, United States
Elena Levantini, Beth Israel Deaconess Medical Center, Harvard Medical School, United States

***Correspondence:**
Karolina H. Czarnecka
karolina.czarnecka@umed.lodz.pl

Specialty section:
This article was submitted to Thoracic Oncology, a section of the journal Frontiers in Oncology

Received: *10 July 2019*
Accepted: *21 November 2019*
Published: *13 December 2019*

Citation:
Czarnecka KH, Szmyd B, Barańska M, Kaszkowiak M, Kordiak J, Antczak A, Pastuszak-Lewandoska D and Brzeziańska-Lasota E (2019) A Strong Decrease in TIMP3 Expression Mediated by the Presence of miR-17 and 20a Enables Extracellular Matrix Remodeling in the NSCLC Lesion Surroundings. Front. Oncol. 9:1372. doi: 10.3389/fonc.2019.01372

Background: Lung cancer is one of the most common causes of death worldwide with a relatively high fatality rate and a mean 5-years survival of about 18%. One of the hallmarks of cancer is the extracellular matrix (ECM) remodeling, which is crucial for metastasis. This process may be regulated by miRs targeting metalloproteinases (MMPs) associated with the ECM breakdown and metastatic process or blocking the action of tissue inhibitors of metalloproteinases (TIMPs). Search for early biomarkers is essential in detecting non-small cell lung cancer (NSCLC) and distinguishing its subtypes: Adenocarcinoma (AC) from Squamous Cell Carcinoma (SCC), enabling targeted chemotherapy.

Methods: *MiR-17* and *miR-20a* targeting *MMP2* and *TIMP3* were selected by TCGA data analysis with further validation using miRTarBase and literature. The study group comprised 47 patients with primary NSCLC (AC and SCC subtypes). RNA was isolated from the tumor and normal-looking neighboring tissue (NLNT) free of cancer cells. MiRs from peripheral blood exosomes were extracted on admission and 5–7 days after surgery. Gene and miRs expression were assessed in qPCR using TaqMan probes.

Results: The *MMP2* has been expressed on a similar level in NLNT, as in cancer. While, *TIMP3* expression was decreased both in cancer tissue and NLNT, with significantly lower expression in cancer. *TIMP3* downregulation in NLNT and in SCC subtype correlated negatively with *miR-20a*. The preoperative *miR-17* expression was significantly higher among patients with SCC compared to AC. Receiver operating characteristic (ROC) analysis of *miR-17* as AC subtype classifier revealed 90% specificity and 48% sensitivity in optimal cut-off point with area under ROC curve (AUC): 0.71 (95%CI: 0.55–0.87). Within NSCLC subtypes: a strong negative correlation between pack-years (PY) and *TIMP3* expression was observed for NLNT in the SCC group.

Conclusion: The *TIMP3* silencing observed in the NLNT and its negative correlation with presurgical expression of *miR-20a* (from serum exosomes), suggest that miRs can influence ECM remodeling at a distance from the center of the lesion. The *miRs* expression pattern in serum obtained before surgery significantly differs between AC and SCC subtypes. Moreover, decreased *TIMP3* expression in NLNT (in SCC group) negatively correlates with the amount of tobacco smoked in a lifetime in PY.

Keywords: NSCLC molecular diagnostic markers, miRNA regulation, microRNA, extracellular matrix remodeling, metalloproteinases, tissue inhibitors of metalloproteinases, exosomes

INTRODUCTION

Lung cancer (LC) is one of the most prevalent cancers, with 2 million new cases in 2018 and with a relatively high fatality rate: the overall ratio of mortality to incidence is 0.87 (1, 2). It is also one of the leading causes of cancer mortality in most developed countries, representing almost 20% of deaths due to cancer (2). Two main clinical types are distinguished: small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC). NSCLC accounts for almost 80% of LC, and comprised squamous cell carcinoma (SSC; 30–32% of LC), adenocarcinoma (AC; 10–35%), and large cell cancer (LCC; 10%) (3–5). Five-years survival rates drastically decrease with the cancer stage, from 83% for AJCC stage I (American Joint Committee on Cancer Staging), to ∼6.6% for detection at an advanced stage (AJCC stage III/IV) (6, 7). The molecular markers are needed to identify NSCLC at its early stages, predict cancer development, and treatment response. Potential biomarkers may be found among molecules responsible for extracellular matrix (ECM) remodeling: a process driven by matrix metalloproteinases (MMPs) countered by the endogenous tissue inhibitors of metalloproteinases (TIMPs) (8, 9).

MMPs are the proteolytic enzymes degrading the components of the basement membrane, acting in many physiological processes (embryogenesis, angiogenesis, apoptosis, wound healing) and in cancer development (8, 10). Elevated *MMP2* expression was observed in stromal fibroblasts, preneoplastic bronchial squamous lesions and pulmonary carcinoma (both in highly invasive and moderate growth areas) (11–13). In NSCLC, the *MMP2* upregulation has been associated with greater tumor size or distant metastasis (14, 15). The MMPs' action can be specifically inhibited by non-covalent binding of TIMPs, which leads to tumor growth suppression and apoptosis promotion (9, 16, 17). Decreased *TIMP3* expression has been observed in many human cancers, i.e., LC, gastric, hepatic, prostate, and endometrial cancer (18–20).

MMP2 and *TIMP3* expression is regulated through microRNAs (miRs), in a post-transcriptional epigenetic mechanism, leading to mRNA degradation, or translation inhibition (21, 22). MiRs are considered as promising molecular markers for the non-invasive early diagnosis of NSCLC (18, 21) and can be assessed in an inexpensive and patients-friendly way in the peripheral blood exosomes (23). Up to date, miRs have been described as potential biomarkers detecting early stages of NSCLC (*miR-182, miR-183, miR-210*, and *miR-126*) (24) or distinguishing SCC from AC (*miR-26a;* small *miR* panel—*205-5p, 944)* (15, 25). In our study, we focused on the *miR-17* and *miR-20a* targeting *MMP2* and *TIMP3*; these were selected based on TCGA data with further miRTarBase and literature validation (see section Selection of microRNA Molecules). *MiR-17* and *miR-20a* have a significant impact on the development of cancer throughout the body (26–28). Both miRs share the ability to stimulate cell proliferation and inhibit apoptosis (29). One of the well-characterized actions of *miR-17* is its ability to target *MMP2* and *TIMP3* genes. Elevated *miR-17* expression was correlated with a worse outcome (negative correlation with overall survival and disease-free survival) in hepatocellular and pancreatic cancers (30). *MiR-20* possesses tumor suppressor activity by blocking VEGF-induced endothelial cell migration (31). Moreover, decreased *miR-20a* expression was found to be associated with faster tumor growth and poor prognosis (32).

The present study evaluates the relative expression of selected genes (*MMP2, TIMP3*) and miRs (*miR-17, miR-20a*) engaged in ECM remodeling in histopathologically-confirmed NSCLC. Many studies showed that the elevated concentration of the exosomes detected in cancer patient's serum originated from the cancer (26–28, 33). The idea of our study was to look for preoperative circulating miR, that can be obtained in a patient-friendly way from the peripheral blood, as simple preoperative clinical biomarker, distinguishing histopathological subtypes of NSCLC, as well as pTNM, and AJCC stages.

MATERIALS AND METHODS

Patient Clinical Features and Lung Tissue Samples

The study material comprised lung tissue and serum samples obtained from 47 patients diagnosed with NSCLC admitted to the Department of Thoracic Surgery, General and Oncologic Surgery (University Teaching Hospital No. 2 in Lodz., Medical University of Lodz, Poland) between July 2014–March 2017. All patients underwent surgery, based on the results of preoperative assessment. The exclusion criteria included a history of other malignancies, active infectious disease and chemo-, or radiotherapy before the surgery. The study was performed in accordance with the Helsinki Declaration and was approved by the Ethical Committee of the Medical University of Lodz, no. RNN/140/10/KE. All participants provided written, informed consent to take part in the study. Detailed characterization of

TABLE 1 | Demographic characteristics of patients and histopathological verifications of NSCLC samples.

Clinical and pathological features			Current study
Entire group			47
Gender		Woman	18
		Men	29
Age group		≤60	10
		>60	37
Histopathological type		AC	24
		SCC	22
pTNM	Tumor size	pT1	12
		pT2	27
		pT3	6
		pT4	1
	Lymph nodes invasion	N0	33
		N1 and N2	13
AJCC		AJCC I	23
		AJCC II	17
		AJCC III	6
Pack-years (PYs)		≤30 PYs	15
		31–45 PYs	14
		>45 PYs	17
Type of the surgery		Lobectomy	33
		Pneumonectomy	9
		Bilobectomy	4
		Segmentectomy	1

pTNM, International System of Clinico-Morphological Classification of Tumors (TNM, Tumor Node Metastasis) according to the WHO Histological Typing of Lung Tumors; AJCC, American Joint Committee on Cancer Staging (AJCC staging) according to the IASCLC Staging Project 7th edn.

patients, postoperative histopathological verifications of NSCLC samples (according to the WHO Histological Typing of Lung Tumors and IASCLC Staging Project 7th ed.) and patient smoking status estimated in pack-years (PYs) are presented in **Table 1**.

The NSCLC tumor lung tissue samples were taken from the center of the lesion, while samples of normal-looking neighboring tissue (NLNT) were taken from the surgical margin: histopathological examination confirmed them to be free of cancer cells. After resection, tissue fragments were stored in RNAlater™ Stabilization Solution (Ambion, USA) at −80°C. The peripheral blood samples (∼5 ml) were obtained on admission and 5–7 days after surgical treatment. Serum separated by centrifugation was stored at −80°C until exosome isolation.

Selection of microRNA Molecules

MicroRNAs targeting *MMP2* and/or *TIMP3* in LC were chosen based on the TCGA datasets, containing RNAseq results of NSCLC patients with AC (LUAD project) and SCC (LUSC project) (34–36). Two datasets for AC and SCC, each containing cancer group and a control group, were downloaded using the TCGA biolinks R package. The datasets sizes are presented in **Table 2**. Further validation, using data

TABLE 2 | The size of the obtained dataset from the GDC database.

Sample type	LUAD project		LUSC project	
	Genes: *MMP2, TIMP3*	miRs: *miR-17, miR-20a*	Genes: *MMP2, TIMP3*	miRs: *miR-17, miR-20a*
Cancer[a]	539	519	502	478
Controls[b]	59	46	51	45

LUAD project, TCGA dataset, containing RNAseq results of NSCLC patients with AC; LUSC project, TCGA dataset, containing RNAseq results of NSCLC patients with SCC;
[a] "Solid state Tumor" origin selected for cancer samples;
[b] "Solid Tissue Normal" origin selected for controls.

retrieved from public microRNA databases (microRNA.org; mirtarbase.mbc.nctu.edu.tw), indicated that *miR-20a* silences *MMP2* expression and *miR-17* targets both *MMP2* and *TIMP3* (see **Supplementary Figure 1**). In the performed literature search (PubMed query on miR & ECM remodeling & cancer) many studies indicated that both miRs have a significant impact on the development of cancer throughout the body (26–32).

Gene and miR Expression in the Studied NSCLC Cohort—Laboratory Procedures

miR and RNA Isolation and Reverse Transcription

Exosomes were pre-isolated from serum using a standardized isolation kit enabling enrichment of intact exosomes (Total Exosome Isolation Reagent, Invitrogen, USA) and resuspended in PBS. The Total Exosome RNA and Protein Isolation Kit (Invitrogen, USA), standardized for the isolation of 30–120 nm diameter vesicles, was used to recover total RNA including the small RNA fraction (37, 38). The total RNA from tissues was isolated using the Qiagen RNeasy Mini Kit (QIAGEN, USA), according to the manufacturer's protocol. The quality and quantity of RNA was spectrophotometrically assessed. Only samples fulfilling the following requirements were selected for further use: miR concentration 1–5 ng/μl and RNA concentration over 50 ng/μl; with 260/280 nm ratio 1.8–2.0. Reverse transcription (RT) reactions were performed using a High-Capacity cDNA Reverse Transcription Kit with MultiScribe™ Reverse Transcriptase and additional RNase Inhibitor (both Applied Biosystem, USA) according to the manufacturer's protocol.

Real-Time Quantitative Polymerase Chain Reactions (Real-Time qPCR)

Real-time qPCRs were performed using the 7900HT Fast Real-Time PCR System (Applied Biosystems, USA). The qPCR mixes for miRs analysis contained: diluted RT product, TaqMan® Universal Master Mix II, without UNG and 1 μl of miR probes: hsa-miR-17-5p or hsa-mir-20a (Applied Biosystems, USA). The miRs expression analysis, performed in DataAssist v3.01, was based on the global normalization method. The qPCR mixes for gene analysis contained diluted cDNA, KAPA PROBE FAST qPCR kit (Kapa Biosystems, USA) and TaqMan assays for *TIMP3* (Gene ID: 7078), *MMP2* (Gene ID: 4313), and *ACTB* (*β-actin*,

Gene ID:60) as endogenous control. All reactions were run in triplicate. RNA isolated from normal lung tissue (Human Lung Total RNA, Ambion, USA) served as a calibrator, for which the RQ (relative quantification) value was considered equal to 1. Analysis was based on the comparative ΔΔCT method, according to the formula ΔΔCT = ΔCT test sample—ΔCT calibrator sample.

Statistical Analysis

Statistical analysis was performed using the Statistica 13.1 (StatSoft, Tulsa, USA). Non-parametric tests were used for statistical analyses, as miR and mRNA expression values did not follow a normal distribution (Shapiro-Wilk test). The results of relative expression analysis (RQ values) are presented as mean ± SD for normal distribution and median value (MV) with interquartile range (IQR) in other cases. When comparing independent groups, ANOVA Kruskal-Wallis, Mann-Whitney *U*-tests (UMW) and the Spearman's rank correlation (r_s) were used. Expression analysis regarding the pTNM classification was performed according to the tumor size (pT) and presence of node involvement (N). With only one patient in the pT4 group, pT4 was not included in statistical analysis. For the NSCLC subtype determination test, the receiver operating characteristic (ROC) chart was used and to compare the accuracy of various classifiers the area under the ROC curve (AUC) was calculated. TCGA data were analyzed using R software and R/Bioconductor package. $P < 0.05$ was considered as statistically significant.

RESULTS

Selection of microRNA Molecules With Genomic Data Commons and miRTarbase

The correlation analysis of miR and gene expression performed on LUAD and LUSC projects data (34–36) revealed statistically significant negative correlations between the examined genes: *MMP2* and *TIMP3* and both miRs (**Supplementary Figure 1**; **Table 3A**). Moreover, significant differences were found in miR expression between the cancer and control groups in both NSCLC subtypes. In the LUSC project (SCC), a significant increase of *miR-17* with *miR-20a* decrease of (3.64 times lower) was observed in the cancer group compared to controls (both $p < 0.001$, UMW test). The opposite was found for the AC subtype: a significant decrease of *miR-17* and an increase of *miR-20a* in cancer samples (2.29 times higher) compared to controls (**Table 3B**).

Genes Expression in Tumor and Normal-Looking Neighboring Tissues

The gene expression analysis for *MMP2* and *TIMP3* was performed on 43 pairs of primary NSCLC tissue and corresponding normal-looking neighboring tissue (see **Table 4**). In comparison to calibrator (RNA isolated from normal lung tissue), *TIMP3* expression was downregulated in both tissues, and *MMP2* was upregulated in both tissues. The relative distribution and symmetry of gene expression within the analyzed groups (cancer tissue vs. NLNT; AC vs. SCC subtype) are presented in section Materials and Methods in **Supplementary Materials** and **Supplementary Figure 2**. The *TIMP3* expression was significantly decreased in cancer tissue in comparison to NLNT ($p = 0.01$; Wilcoxon test; see **Supplementary Figure 4**). The *MMP2* was expressed on a comparable level in cancer and NLNT ($p = 0.372$; Wilcoxon test). No significant differences were found for gene expression according to age, gender, cancer subtypes, TNM staging (pT and N groups), or AJCC classification. Considering long-life tobacco intake (measured in pack-years) among SCC patients, a negative correlation with *TIMP3* expression in NLNT was found ($R = -0.68$, $p < 0.001$; r_s). Furthermore, the positive correlation between *TIMP3* and *MMP2* expression was observed in NLNT from surgical margin ($R = 0.482$; $p = 0.001$; r_s; see **Figure 1**). In the AC subtype, there were also observed the positive correlations between *TIMP3* in NLNT and *MMP2* levels, both in cancer and NLNT ($R = 0.699$, $p < 0.001$, and $R = 0.500$, $p = 0.021$, respectively; r_s).

The validation of expressional data with the Genomic Data Commons also revealed significantly decreased *TIMP3* level in cancer tissues in comparison to controls in both LUAD ($p < 0.001$; UMW) and LUSC projects ($p < 0.001$; UMW). Whereas, *MMP2* was expressed on a comparable level in both tissues in LUAD and LUSC projects. *TIMP3* expression was higher in the AC subtype in comparison to SCC (34–36), see **Table 3B**. In the present study, we have also compared the *TIMP3/MMP2* expression ratios in NLNT vs. healthy lung tissues. The experimentally-assessed expression of *TIMP3* in

TABLE 3 | Analysis of gene and miR expression in LUAD and LUSC projects: **(A)** Correlation of analyzed miR and gene expression in both projects; **(B)** Gene and miR expression levels in cancer and control groups in both projects.

(A) Correlation of analyzed miR and gene expression in LUAD and LUSC projects

Squamous cell carcinoma—LUSC		
	miR-17	*miR-20a*
MMP2	**−0.47***	**−0.40***
TIMP3	**−0.37***	**−0.34***
Adenocarcinoma—LUAD		
	miR-17	*miR-20a*
MMP2	**−0.33***	**−0.24***
TIMP3	**−0.30***	**−0.25***

(B) Analysis of gene and miR expression in cancer and control groups for LUAD and LUSC projects

	Gene/miR	No of transcripts cancer group [FPKM]	No of transcripts control group [FPKM]	*P*-value
LUSC	*MMP2*	6530.92	7945.22	0.095
	TIMP3	4780.33	33445.33	**<0.001**
	miR-17	958.16	717.86	**<0.001**
	miR-20a	79.61	288.56	**<0.001**
LUAD	*MMP2*	5966.76	4919.22	0.388
	TIMP3	7539.80	38223.79	**<0.001**
	miR-17	626.16	367.34	**<0.001**
	miR-20a	195.48	85.93	**<0.001**

*$p < 0.001$—in R coefficient analysis.
The values marked in bold represent the statistically significant correlations.

TABLE 4 | Clinical and pathological features: median expression level (RQ value) of tested genes.

Clinical and pathological features			*N*	*MMP2* NLNT	*MMP2* cancer tissue	*TIMP3* NLNT	*TIMP3* cancer tissue
Entire group			43	1.422 (IQR: 0.600–3.162)	0.900 (IQR: 0.387–2.392)	0.013 (IQR: 0.006–0.057)	0.006 (IQR: 0.002–0.069)
Gender		Women	14	1.287	0.760	0.021	0.009
		Men	29	1.561	1.213	0.013	0.005
Age group		≤60 years	9	1.330	1.748	0.023	0.010
		>60 years	34	1.446	0.830	0.013	0.005
Histopathological type		AC	21	1.575	1.252	0.013	0.006
		SCC	21	1.015	0.530	0.013	0.004
pTNM	Tumor size	pT1	11	1.471	0.900	0.013	0.004
		pT2	24	1.073	0.479	0.012	0.012
		pT3	6	2.752	1.414	0.019	0.002
		pT4	1	9.255	6.217	0.054	0.004
	Lymph nodes invasion	pN0	30	1.376	0.911	0.014	0.009
		pN1 & pN2	12	1.488	0.645	0.011	0.003
AJCC		AJCC I	21	1.681	0.784	0.020	1.681
		AJCC II	15	1.376	1.061	0.008	1.376
		AJCC III	6	1.295	1.082	0.019	1.295
Pack-years (PYs)		≤30 PYs	11	1.000	0.524	0.023	0.004
		31-45 PYs	13	2.197	1.952	0.036	0.008
		>45 PYs	17	1.132	0.760	0.006	0.006
Type of the surgery		Lobectomy	33	1.287	0.836	0.013	0.007
		Pneumonectomy	9	2.925	1.535	0.014	0.003
		Bilobectomy	4	0.466	1.252	0.009	0.004
		Segmentectomy	1	22.100	16.898	0.052	0.010

NLNT, normal-looking neighboring tissue; pTNM, International System of Clinico-Morphological Classification of Tumors (TNM, Tumor Node Metastasis) according to the WHO Histological Typing of Lung Tumors; AJCC, American Joint Committee on Cancer Staging (AJCC staging) according to the IASCLC Staging Project 7th edn.

NLNT was ∼109 times lower than the *MMP2*. In the RNA-seq normal tissues datasets (gtexportal.org, proteinatlas.org: controls for LUSC and LUAD projects) *TIMP3* expression was 3.35–7.77 higher than *MMP2* (see **Supplementary Figure 4**).

microRNA Expression

MiR expression was assessed among 43 patients (see **Table 5**). Both miRs were expressed at comparable levels before and after surgery ($p = 0.681$ and $p = 0.334$, respectively, Wilcoxon test). The relative distribution and symmetry of miRs expression within the analyzed groups (before vs. after surgery; AC vs. SCC subtype) are presented in section Materials and Methods in **Supplementary Materials** and **Supplementary Figure 3**). The type of surgery did not affect the expression of postoperative miRs ($p = 0.202$ for *miR-17* and $p = 0.202$ for miR-20a, Kruskal-Wallis test). No statistically significant observations were made for miR levels in terms of age, gender, or TNM tumor classification. Besides, no significant correlations were found between miRs and tobacco intake (measured in pack-years), neither before nor after tumor resection. There were observed positive correlations between the miR expression before and after surgery (see **Figure 1**).

Significant differences were found for preoperative miR expression among NSCLC subtypes (see **Supplementary Figure 3**). The preoperative *miR-17* expression was significantly higher among SCC patients compared to AC ($p = 0.02$, UMW test). For preoperative *miR-20a*, the opposite dependence was observed ($p = 0.02$, UMW test). The presurgical *miR-17* and *miR-20a* expression levels were used to establish the NSCLC subtype determination test. AUC for *miR-17* classifier was 0.710 (95% CI: 0.554–0.865). For the optimal cut-off point (≤0.189; Youden's J statistic), the specificity and positive predictive values (PPV) in detecting AC were equal to 90 and 83%; sensitivity and negative predictive value (NPV) were equal to 48 and 63%, respectively (see **Figure 2**). Moreover, presurgical *miR-20a* expression may be used to distinguish the NSCLC subtypes with AUC of 0.71 (95% CI: 0.554–0.865). For the optimal cut-off point (≥5.295 for AC; Youden's J statistic) the specificity and PPV in detecting AC were equal to 90 and 83%; sensitivity and NPV were 48 and 63%, respectively.

The validation of the miR expressional data with the Genomic Data Commons (34–36) revealed significantly decreased *miR-17* expression in controls, compared to cancer group in both LUSC and LUAD projects. The miR-20a expression in cancer vs. control groups was opposite in LUSC and LUAD projects (see **Table 3B**).

Correlation of Gene and microRNA Expression

The analysis of miR correlation with gene expression was performed on 39 patients with overlapping data. No correlations

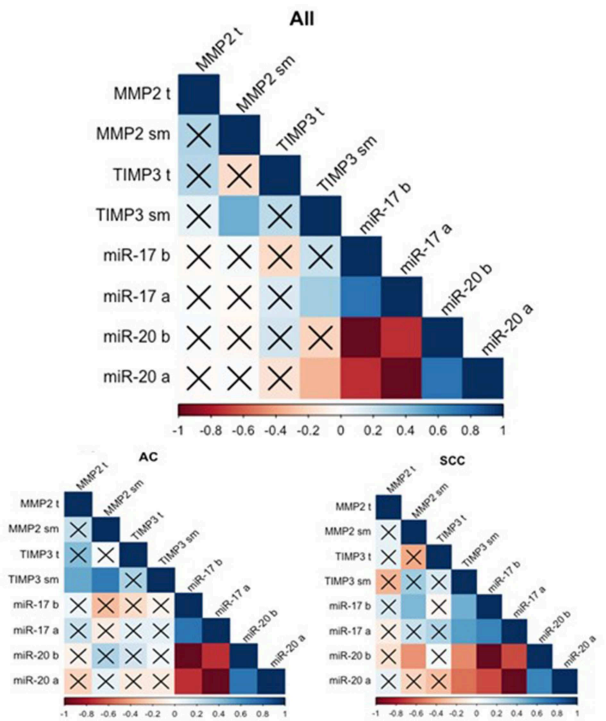


FIGURE 1 | Spearman's rank correlogram of dependencies of miR and gene expression. The lower part represents the dependencies of miR and gene expression in histotypes: Adenocarcinoma (AC) and Squamous Cell Carcinoma (SCC). Correlations marked with X are not statistically significant ($p > 0.05$). Color scale represents Spearman's correlation coefficient (r_s). t, tumor sample; sm, normal-looking neighboring tissue from surgical margin; b (before), preoperative expression level; a (after), postoperative expression level.

were found in cancer tissue for any of the samples. Weak correlations were discovered between *TIMP3* expression in the NLNT and postoperative miRs: a positive correlation for *miR-17* ($R = -0.334$, $p = 0.037$; r_s) and a negative correlation for *miR-20a* ($R = 0.330$, $p = 0.04$; r_s). Similarly in SCC subtype the *TIMP3* expression in NLNT correlated positively with *miR-17* (before: $R = 0.474$, $p = 0.03$; after: $R = 0.549$, $p = 0.01$; r_s), and negatively with *miR-20a* (before: $R = -0.474$, $p = 0.03$; after: $R = -0.549$, $p = 0.01$; r_s) (see **Figure 1**). No statistically significant correlations between miR and gene expression were observed within the AC group.

DISCUSSION

The high mortality and late detection rates associated with NSCLC create an urgent need for developing new biomarkers enabling early detection. Hence, it is essential to identify candidate biomarkers among genes and microRNAs regulating the ECM remodeling present in early stages of NSCLC, which can help distinguish subtypes. Our present study aimed to evaluate the relative expression of selected genes and miRs engaged in ECM remodeling in histopathologically-confirmed NSCLC, as candidates for molecular biomarkers.

Decreased *TIMP3* Expression in Normal-Looking Neighboring Tissue

Our analysis found *TIMP3* expression to be significantly lower in cancer tissue than in normal looking neighboring tissue. A similar tendency was also observed in the analysis of data from the Genomic Data Commons, where *TIMP3* was significantly decreased in cancer tissues in comparison to controls in both analyzed AC and SCC cohorts (34, 35). Under physiological conditions, any increase of metallopeptidase expression and its activity would be controlled by specific non-covalent binding of the TIMP3 protein, resulting in MMP inhibition (17). In our study, we have noted the positive correlation between *TIMP3* and *MMP2* expression in NLNT, confirmed in the AC subtype, but not in SCC. On the other hand, the observed in the present study, silencing of the metalloproteinase inhibitor *TIMP3* suggests that the proteins taking part in ECM remodeling can display intensified activity. Such remodeling, especially uncontrolled ECM proteolysis, may result in more significant cancer cell proliferation and migration. It was previously demonstrated that restoring *TIMP3* function by blocking the expression of its suppressor gene *EZH2* (using RNA interference) led to subsequent inhibition of cancer cell migration (39). Those findings are not entirely concordant with those of Kumaki et al. who report that increased immunoexpression of different metalloproteinase inhibitors correlated with tumor aggressiveness. The TIMP2 protein was significantly stronger in the invasive areas than the lepidic areas of Invasive Pulmonary Adenocarcinoma (11); this has been attributed to increased TIMP2 expression causing elevated ECM accumulation in the invasive tumor cells, resulting in fibrous scar formation (11).

Cigarette smoking is overwhelmingly tied with SCC rather than AC (40). Regarding smoking status in NSCLC subtypes, we have observed significantly lower *TIMP3* expression in NLNT in long term smokers in SCC subtype, but not in AC. Smoking-induced *TIMP3* downregulation may be one of the molecular causes of cancerogenesis among SCC patients. It was earlier demonstrated that molecular changes like microsatellite instability (MSI) or suppressor gene hypermethylation could occur in histologically-normal epithelia or macroscopically-unchanged tissue adjacent to the resected tumors in smoking patients with primary lung cancer. Also, in smokers without cancer, the loss of heterozygosity and MSI were detected in histologically-normal distal bronchial epithelium (41–44).

MMP2 Expressed on Comparable Level in Both Tissues

MMPs are typically expressed at moderate levels; their expression rapidly grows in response to tissue injury (8, 14) and the course of inflammation processes or cancer development. It was previously reported that MMP2 immunoexpression was significantly higher in the lung cancer group than among the healthy control group (45). On the other hand, ∼50% of the NSCLC patients revealed stronger upregulation of the *MMP2* in the fibroblasts neighboring the lesion, than in the tumor itself (12). MMP2 protein may act as an extracellular matrix modulator in fibroblast cells, enabling malignant transformation by ECM

TABLE 5 | Clinical and pathological features: median expression level (RQ value) of tested miRs.

Clinical and pathological features		*N*	*miR-17* before surgery	*miR-17* after surgery	*miR-20a* before surgery	*miR-20a* after surgery
Entire group		43	0.459 (IQR: 0.088–1.118)	0.667 (IQR: 0.135–1.102)	2.181 (IQR: 0.896–13.134)	1.499 (IQR: 0.907–7.410)
Gender	Women	17	0.496	0.876	2.017	1.141
	Men	26	0.408	0.542	2.489	1.893
Age group	≤60	8	0.081	0.295	4.919	4.856
	>60	35	0.496	0.868	2.017	1.152
Histopathological type	AC	21	0.348	0.593	2.871	1.687
	SCC	21	0.773	0.876	1.294	1.141
pTNM Tumor size	pT1	10	0.533	0.596	2.359	1.775
	pT2	25	0.396	0.667	2.523	1.499
	pT3	6	0.904	0.634	1.202	1.848
	pT4	1	0.500	0.587	9.516	7.619
Lymph node invasion	pN0	30	0.377	0.701	2.660	1.429
	pN1 & pN2	12	0.638	0.380	1.644	2.632
AJCC	AJCC I	20	0.020	0.006	0.427	0.701
	AJCC II	17	0.008	0.011	0.358	0.627
	AJCC III	5	0.019	0.003	1.160	0.391
Pack-years (PYs)	≤30 PYs	12	0.338	0.596	3.384	1.775
	31–45 PYs	15	0.776	0.895	1.289	1.117
	>45 PYs	15	0.292	0.137	3.424	7.300
Type of the surgery	Lobectomy	33	0.353	0.525	2.834	1.939
	Pneumonectomy	9	0.776	0.934	1.289	1.071
	Bilobectomy	4	0.459	0.348	2.181	2.877
	Segmentectomy	1	0.042	0.667	23.800	1.499

pTNM, International System of Clinico-Morphological Classification of Tumors (TNM, Tumor Node Metastasis) according to the WHO Histological Typing of Lung Tumors; AJCC, American Joint Committee on Cancer Staging (AJCC staging) according to the IASCLC Staging Project 7th edn.

degradation and the creation of a suitable microenvironment for vessel growth (12). In our study, performed on paired tissues from NSCLC patients, the *MMP2* expression was on comparable level in cancer tissue and NLNT. The lack of the *MMP2* expression differences among tissues, together with the observed *TIMP3/MMP2* ratio distortion in NLNT (strongly decreased *TIMP3* expression in comparison to *MMP2,* compared to the data from RNAseq studies), indicate that ECM remodeling may be observed some distance from the lesion center. Previous studies have linked the molecular changes in surrounding tissue, mediated by miRs, to the field cancerization effect (43). The phenomenon of tumor promoting (oncogenic) activity of the *MMP2* was previously demonstrated in colorectal cancer, where elevated *MMP2* mRNA levels in "healthy" tissue surrounding the lesion were significantly higher in patients with metastatic cancer than in non-metastatic lesions (46).

miR Expression From Serum Exosomes—Significantly Differs Between NSCLC Subtypes

Many studies report that both *miR-17* and *miR-20a* influence tumor formation and cancer progression in various organs and tissues throughout the body (47–50); however, the vast majority of such papers overwhelmingly describe experiments performed using various cancer cell lines. The present study assesses their expression in exosomes from peripheral blood samples of NSCLC patients, collected before and after surgical removal of the tumor, and compares our findings with data retrieved from TCGA. This approach to miR analysis is described in a few publications (51, 52). In the present study, the *miR-20a* expression levels in pre- and postsurgical samples correlated positively. The miRs produced by the tumor to control the process of matrix remodeling can be observed in patients serum some days after cancer removal. On the other hand Zhang et al. demonstrated that after surgery the lower expression of *miR-20a* and four other miRs after surgery might indicate the *miR-20a* originates from tumor tissue (52). While comparing *miR-17* expression before and after surgical treatment, no significant difference was found. Nevertheless, our study is the first to examine the change in *miR-17* expression among NSCLC patients before and after surgical treatment.

Our research indicates that the miR expression pattern in serum obtained before surgery significantly differs between NSCLC subtypes. *MiR-17* expression was higher among patients with SCC than those with AC, which is concordant with Molina-Pinelo et al. study (53) and was also confirmed by TCGA data (34, 35). Analysis of this dependency was used to create a classifier differentiating between NSCLC subtypes (**Figure 2**). Strongly decreased expression of *miR-17* can be treated as a hallmark

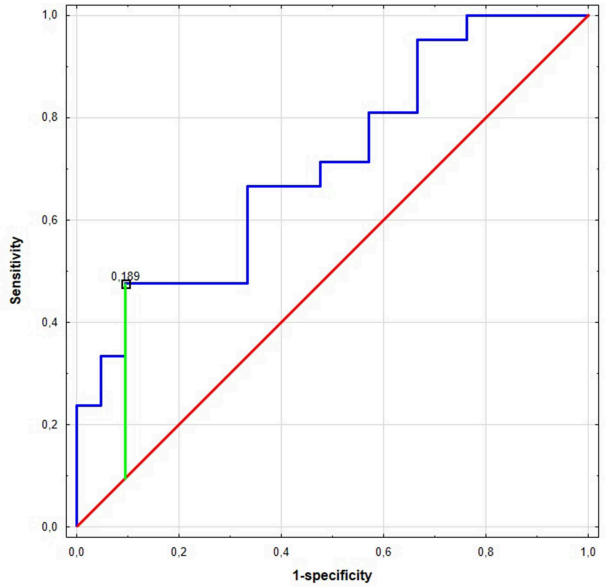


FIGURE 2 | The Receiver Operating Characteristic (ROC) curve of *miR-17* in the NCSLC subtype classification. The best cut-off point for AC detection was equal to ≤0.189 for *miR-17* (Youden's J statistic). The area under the curve (AUC) was equal to AUC 0.710 (95% CI: 0.554–0.865), its predictive value was as follows: specificity−90%, positive predictive value (PPV)−83%, sensitivity−48%, and negative predictive value (NPV)−63%.

of AC subtype with 90% specificity, where a mean RQ value over 0.189 FC can indicate a SCC subtype with 48% specificity. The opposite trend was observed for *miR-20a*: its expression was significantly increased in AC in comparison to the SCC group.

Our results are in accordance with the analysis of TCGA data, which revealed a higher number of *miR-20a* transcripts among AC patients than those with SCC. *MiR-20a* has previously been described as SCC subtype biomarker; however, these studies compared samples from SCC patients with those of healthy volunteers, or with postoperative samples from the same patient (only SCC subtype). Zhang et al. propose a panel of three microRNAs (including *miR-20a*) as a potential diagnostic marker for distinguishing male lung SCC patients from healthy volunteers (54). Aushev et al. propose a panel of five miRs (*miR-205,−19a,−19b,−30b,* and *-20a*) as a SCC biomarker; these five miRs were found to be upregulated in exosomes extracted from the peripheral blood of SCC patients before surgery compared to postoperative samples (55). However, neither study performed any assessment regarding other NSCLC subtypes.

As no literature comparing *miR-20a* expression among AC vs. SCC patients could be found, our study appears to be the first to propose *miR-20a* as biomarker for distinguishing NSCLC subtypes, following an analysis of miR expression in SCC and AC. This can be essential in case of advanced lung cancer (confirmed with low-invasive methods like bronchoscopy or lung needle biopsy) in AJCC clinical stage IIIA or IIIB, where chemotherapy can be used as induction treatment before surgery or as a complementary treatment for radiotherapy (3–5). Preoperative subtype diagnosis may help to select better first-line chemotherapy schemes in advanced NSCLC. Different outcomes were observed for squamous vs. non-squamous cell carcinomas: i.e., a better response was observed for pemetrexed treatment in non-squamous NSCLCs compared to SCC (56). As no correlations of miR expression with tumor grading and staging (TNM and AJCC) were found, those miRs probably would not be useful as preoperative circulating biomarkers of cancer stage.

The Correlations of Gene and miR Expression

The present study identified correlations regarding gene expression (*TIMP3* vs. *MMP2*) or genes and miRs (*TIMP3* vs. *miR-17* and *miR-20a*) in the normal-looking neighboring tissue, which were not detected in the center of the cancer lesion (**Figure 3**). Such changes in miR expression can be prompted by more complex regulation of ECM remodeling in the cancer neighborhood or could be characteristic of NSCLC subtype. Moreover, our results are in concordance with miR expressional data from the Genomic Data Commons, where the studied miRs presented different trends among subtypes: *miR-20a* was upregulated among tested LUAD patients, but not LUSC.

Due to the gene location (13q31) the analyzed miRs were observed to be expressed jointly as a 2 gene cluster or act as a part of larger *miR-17-92* cluster which has been overexpressed in LC, colorectal cancer, and hepatocellular carcinoma (30, 57). We did not observe any joint *miR-17/miR-20a* expression pattern: the preoperatively *miR-17* expression was higher in SCC, while *miR-20a* was higher in AC, and miRs correlated in opposite way with *TIMP3* in NLNT.

In physiological conditions, the *TIMP3* expression would be triggered by the ECM remodeling signals to counterpart the MMP activity (17). *TIMP3* silencing is associated with the intensified ECM remodeling and cancer cell migration (39). Such ECM remodeling in the tumor surrounding tissue can be forced by the rapid growth of the lesion and its rigidity: when the diameter of the tumor increases, it exerts pressure on neighboring tissues and blood vessels (48). Surprisingly we have observed the positive correlation of *TIMP3* in NLNT with *miR-17* expression (in the entire study cohort and SCC subtype, not detected in AC) as the increase of the *miR17* is linked to the angiogenesis, though we hypothesize that *TIMP3* expression may be activated by indirect miR action. On the other hand, we have observed a significant increase of the preoperative *miR17* expression in SCC compared to AC, which can partially explain the observed positive correlation with *TIMP3* decrease.

The observed in our study negative correlation of *TIMP3* with *miR-20a* expression (both preoperative and postoperative *miR20a* in SCC subtype, post *miR20a* in the entire study cohort), can be explained as epigenetic silencing of the genes controlling the ECM remodeling. The *miR-20a* action has been previously linked to the induction of vascular changes in invasive breast carcinomas (48) and metastasis in gastric cancer (50). *TIMP3* silencing mediated by *miR-20a* may be treated as a

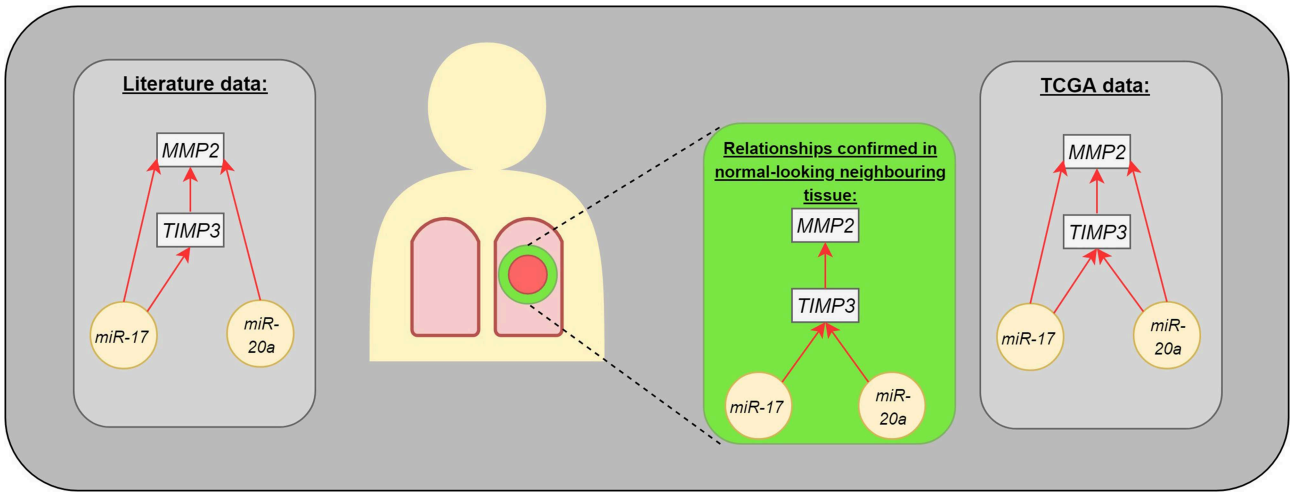


FIGURE 3 | The interaction between analyzed molecules (genes and microRNAs). The gray boxes present the dependencies published in medical literature (left) and TCGA data (right). The middle part presents the statistically significant differences observed in our study: microRNAs were assessed in postoperative blood sample, genes in normal-looking neighboring tissue from surgical margin.

hallmark of substantial ECM deregulation already present in the NLNT. Remodeling encourages further growth of the lesion and induces local hypoxia, which in turn can encourage the development of new blood vessels growth (48, 57). The observed strong downregulation of *TIMP3* and disturbed *MMP2/TIMP3* expression ratio can be explained as crucial activities in the ECM remodeling process, enabling creation of a microenvironment conducive to tumor growth (57).

CONCLUSIONS

- The *MMP2* expression on comparable level in both tissues, together with strong *MMP2* vs. *TIMP3* upregulation in NLNT (compared to the healthy controls data from HPA RNA-seq project or Genomic Data Commons), indicate the metalloproteinase mediated ECM remodeling can occur in the distance from the center of the lesion.
- The *TIMP3* silencing observed in the normal-looking neighboring tissue and its negative correlation with presurgical *miR-20a* expression from serum exosomes (in SCC subtype) suggest the role of miRs in ECM remodeling
- miR expression pattern in serum obtained before surgery significantly differs between NSCLC subtypes. Preoperative serum miR examination can be considered as a useful biomarker for the neoadjuvant therapy planning for patients with confirmed lung cancer and clinical AJCC stage III A/III B. Furthermore, the proposed miR-based classifier does not require the use of minimally invasive diagnostic methods, such as biopsy or broncho-alveolar lavage, just a simple extraction of miR from serum exosomes.
- The downregulation of *TIMP3* in long-term smokers and the decrease of presurgical *miR-17* expression, can be regarded as potential SCC subtype markers.

STRENGTHS AND WEAKNESSES OF THE STUDY

The strengths of this study are:

- the prospective design;
- inclusion of the NSCLC patients that were not treated with potentially mutagenic chemotherapy or radiotherapy before the surgery;
- analysis in the most common NSCLC subtypes: AC and SCC;
- analysis of *miR-17* and *miR-20a* expression before and after surgical treatment;
- analysis of *MMP2* and *TIMP3* expression in both cancer lesion and the normal-looking neighboring tissue (from surgical margin);
- validation of the obtained results with the available sequencing datasets from HPA RNA-seq project, or Genomic Data Commons.

The weaknesses of the study are:

- a relatively low number of studied patients, despite a proper distribution among NSCLC subtypes.
- Difficulty confirming the cancer origin of the analyzed miRs—miRs were extracted from exosomes circulating in blood.
- Expression analysis performed only at the mRNA level, not assessed on the protein level. IHC analysis was not provided either in our study nor in the TCGA validation dataset.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

The study was performed in accordance with the Helsinki Declaration and the ethical proceedings approved by the Ethical Committee of the Medical University of Lodz, Poland, no. RNN/140/10/KE. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

KC created the concept of the study and provided the final version. KC and DP-L design the study. KC, BS, MB, and MK performed laboratory procedures, conducted statistical analysis, and wrote the manuscript. JK, AA, DP-L, and EB-L revised the manuscript.

FUNDING

This study was supported by grant for Young Scientist (to KC) financed by Medical University of Lodz, number of subsidy 502-03/1-151-04/502-14-247.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2019.01372/full#supplementary-material

Supplementary Figure 1 | Diagram of the interactions between genes selected for the study (*TIMP3, MMP2*) and miRs targeting them. MiRs selected using the databases microRNA.org and mirtarbase.mbc.nctu.edu.tw.

Supplementary Figure 2 | Violin plot describing the distribution of *MMP2* and *TIMP3* gene expression in tumor and normal-looking neighboring tissue **(A,B)**, as well as in NSCLC subtypes **(C,D)**. The violin plot analysis represents the kernel density plot of the analysis of binary classified data. SCC, Squamous Cell Carcinoma; AC, Adenocarcinoma.

Supplementary Figure 3 | Violin plot describing distribution of *miR-17* and *miR-20a* expression in tumor and normal-looking neighboring tissue: **(A,B)** and in NSCLC subtypes **(C,D)**. The violin plot analysis represents the kernel density plot in the analysis of binary classified data. SCC, Squamous Cell Carcinoma; AC, Adenocarcinoma.

Supplementary Figure 4 | Box-and-whisker plot representing the expression of *TIMP3* in tumor and normal-looking neighboring tissue from surgical margin ($p = 0.01$; Wilcoxon test).

Supplementary Table 1 | Expression level of analyzed genes in control tissue from RNA-seq analysis and normal-looking neighboring tissue.

REFERENCES

1. 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. doi: 10.3322/caac.21492
2. International Agency for Research on Cancer. *Fact Sheets by Cancer (Colorectal). GLOBOCAN 2012*: Estimated *Cancer* Incidence, Mortality and *Prevalence* Worldwide in *2012*. (2012) Available online at: http://globocan.iarc.fr/Pages/fact_sheets_cancer.aspx (accessed May 17, 2018).
3. Travis WD, Brambilla E, Müller-Hermelink HK, Harris C. *World HEALTH ORGANIZATION Classification of Tumours; Tumours of Lung, Pleura, Thymus and Heart*. (2004). p. 9–122. Available online at: http://www.who.int/bookorders (accessed May 17, 2018).
4. Herbst RS, Morgensztern D, Boshoff C. The biology and management of non-small cell lung cancer. *Nature.* (2018) 553:446–54. doi: 10.1038/nature25183
5. Kalemkerian GP, Akerley W, Bogner P, Borghaei H, Chow LQ, Downey RJ, et al. Non-small cell lung cancer clinical practice guidelines in oncology. *J Natl Compr Cancer Netw.* (2010) 8:740–801. doi: 10.6004/jnccn.2010.0056
6. Kathuria H, Gesthalter Y, Spira A, Brody JS, Steiling K. Updates and controversies in the rapidly evolving field of lung cancer screening, early detection, and chemoprevention. *Cancers.* (2014) 6:1157–79. doi: 10.3390/cancers6021157
7. Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int J Cancer.* (2010) 127:2893–917. doi: 10.1002/ijc.25516
8. Trojanek J. Matrix metalloproteinases and their tissue inhibitors. *Postepy Biochem.* (2012) 58:353–62. Available online at: http://www.postepybiochemii.pl/pdf/3_2012/353-362.pdf
9. Anand-Apte B, Bao L, Smith R, Iwata K, Olsen BR, Zetter B, et al. A review of tissue inhibitor of metalloproteinases-3 (TIMP-3) and experimental analysis of its effect on primary tumor growth. *Biochem Cell Biol.* (1996) 74:853–62. doi: 10.1139/o96-090
10. Yan C, Boyd DD. Regulation of matrix metalloproteinase gene expression. *J Cell Physiol.* (2007) 211:19–26. doi: 10.1002/jcp.20948
11. Kumaki F, Matsui K, Kawai T, Ozeki Y, Yu ZX, Ferrans VJ, et al. Expression of matrix metalloproteinases in invasive pulmonary adenocarcinoma with bronchioloalveolar component and atypical adenomatous hyperplasia. *Am J Pathol.* (2001) 159:2125–35. doi: 10.1016/S0002-9440(10)63064-7
12. Ishikawa S, Takenaka K, Yanagihara K, Miyahara R, Kawano Y, Otake Y, et al. Matrix metalloproteinase-2 status in stromal fibroblasts, not in tumor cells, is a significant prognostic factor in non-small-cell lung cancer. *Clin Cancer Res.* (2004) 10:6579–85. doi: 10.1158/1078-0432.CCR-04-0272
13. Galateau-Salle FB, Luna RE, Horiba K, Sheppard MN, Hayashi T, Fleming M V, et al. Matrix metalloproteinases and tissue inhibitors of metalloproteinases in bronchial squamous preinvasive lesions. *Hum Pathol.* (2000) 31:296–305. doi: 10.1016/S0046-8177(00)80242-7
14. Osman NM, Osman WM. SDF-1 and MMP2 cross talk in cancer cells and tumor microenvironment in non-small cell lung cancer. *Egypt J Chest Dis Tuberc.* (2016) 65:517–25. doi: 10.1016/j.ejcdt.2016.01.001
15. Pastuszak-Lewandoska D, Kordiak J, Czarnecka KH, Migdalska-Sek M, Nawrot E, Domanska-Senderowska D, et al. Expression analysis of three miRNAs, miR-26a, miR-29b and miR-519d, in relation to MMP-2 expression level in non-small cell lung cancer patients: a pilot study. *Med Oncol.* (2016) 33:96. doi: 10.1007/s12032-016-0815-z
16. Kallio JP, Hopkins-Donaldson S, Baker AH, Kähäri V-M. TIMP-3 promotes apoptosis in nonadherent small cell lung carcinoma cells lacking functional death receptor pathway. *Int J Cancer.* (2011) 128:991–6. doi: 10.1002/ijc.25404
17. Hadler-Olsen E, Winberg JO, Uhlin-Hansen L. Matrix metalloproteinases in cancer: their value as diagnostic and prognostic markers and therapeutic targets. *Tumor Biol.* (2013) 34:2041–51. doi: 10.1007/s13277-013-0842-8
18. Lin PY, Yu SL, Yang PC. MicroRNA in lung cancer. *Br J Cancer.* (2010) 103:1144–8. doi: 10.1038/sj.bjc.6605901
19. Guan Z, Zhang J, Song S, Dai D. Promoter methylation and expression of TIMP3 gene in gastric cancer. *Diagn Pathol.* (2013) 8:110. doi: 10.1186/1746-1596-8-110
20. Shinojima T, Yu Q, Huang SK, Li M, Mizuno R, Liu ET, et al. Heterogeneous epigenetic regulation of TIMP3 in prostate cancer. *Epigenetics.* (2012) 7:1279–89. doi: 10.4161/epi.22333
21. Sohel MH. Extracellular/circulating MicroRNAs: release mechanisms, functions and challenges. *Achiev Life Sci.* (2016) 10:175–86. doi: 10.1016/j.als.2016.11.007
22. Mohr AM, Mott JL. Overview of MicroRNA biology. *Semin Liver Dis.* (2015) 35:3–11. doi: 10.1055/s-0034-1397344
23. Kosaka N, Iguchi H, Yoshioka Y, Takeshita F, Matsuki Y, Ochiya T. Secretory mechanisms and intercellular transfer of MicroRNAs in living Cells. *J Biol Chem.* (2010) 285:17442–52. doi: 10.1074/jbc.M110.107821

24. Zhu WY, Zhou KY, Zha Y, Chen DD, He JY, Ma HJ, et al. Diagnostic value of serum miR-182, miR-183, miR-210, and miR-126 levels in patients with early-stage non-small cell lung cancer. *PLoS ONE.* (2016) 11:e0153046. doi: 10.1371/journal.pone.0153046
25. Li H, Jiang Z, Leng Q, Bai F, Wang J, Ding X, et al. A prediction model for distinguishing lung squamous cell carcinoma from adenocarcinoma. *Oncotarget*. (2017) 8:50704–14. doi: 10.18632/oncotarget.17038
26. John B, Enright AJ, Aravin A, Tuschl T, Sander C, Marks DS. Human microRNA targets. *PLoS Biol.* (2004) 2:e363. doi: 10.1371/journal.pbio.0020363
27. Xu T, Jing C, Shi Y, Miao R, Peng L, Kong S, et al. MicroRNA-20a enhances the epithelial-to-mesenchymal transition of colorectal cancer cells by modulating matrix metalloproteinases. *Exp Ther Med.* (2015) 10:683–8. doi: 10.3892/etm.2015.2538
28. Yang X, Du WW, Li H, Liu F, Khorshidi A, Rutnam ZJ, et al. Both mature miR-17-5p and passenger strand miR-17-3p target TIMP3 and induce prostate tumor growth and invasion. *Nucleic Acids Res.* (2013) 41:9688–704. doi: 10.1093/nar/gkt680
29. Vasaikar S V, Straub P, Wang J, Zhang B. LinkedOmics: analyzing multi-omics data within and across 32 cancer types. *Nucleic Acids Res.* (2018) 46:D956–63. doi: 10.1093/nar/gkx1090
30. Kong W, Cheng Y, Liang H, Chen Q, Kun X, Huang Z, et al. Oncotargets and therapy dovepress prognostic value of mir-17-5p in cancers: a meta-analysis. *Onco Targets Ther.* (2018) 11:3541–9. doi: 10.2147/OTT.S150340
31. Pin A-L, Houle F, Guillonneau M, Paquet ER, Simard MJ, Huot J. miR-20a represses endothelial cell migration by targeting MKK3 and inhibiting p38 MAP kinase activation in response to VEGF. *Angiogenesis.* (2012) 15:593–608. doi: 10.1007/s10456-012-9283-z
32. Zhang L, Xiang P, Han X, Wu L, Li X, Xiong Z. Decreased expression of microRNA-20a promotes tumor progression and predicts poor prognosis of cutaneous squamous cell carcinoma. *Int J Clin Exp Pathol.* (2015) 8:11446–51. Available online at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4637689/
33. Wang H, Peng R, Wang J, Qin Z, Xue L. Circulating microRNAs as potential cancer biomarkers: the advantage and disadvantage. *Clin Epigenetics.* (2018) 10:59. doi: 10.1186/s13148-018-0492-1
34. *Lung Squamous Cell Carcinoma—TCGA*. Available online at: https://cancergenome.nih.gov/cancersselected/lungsquamouscell (accessed February 12, 2019).
35. *Lung Adenocarcinoma—TCGA*. Available online at: https://cancergenome.nih.gov/cancersselected/lungadenocarcinoma (accessed February 12, 2019).
36. Grossman RL, Heath AP, Ferretti V, Varmus HE, Lowy DR, Kibbe WA, et al. Toward a shared vision for cancer genomic data. *N Engl J Med.* (2016) 375:1109–12. doi: 10.1056/NEJMp1607591
37. Zeringer E. Methods for the extraction and RNA profiling of exosomes. *World J Methodol.* (2013) 3:11. doi: 10.5662/wjm.v3.i1.11
38. Tang YT, Huang YY, Zheng L, Qin SH, Xu XP, An TX, et al. Comparison of isolation methods of exosomes and exosomal RNA from cell culture medium and serum. *Int J Mol Med.* (2017) 40:834–44. doi: 10.3892/ijmm.2017.3080
39. Xu C, Hou Z, Zhan P, Zhao W, Chang C, Zou J, et al. EZH2 regulates cancer cell migration through repressing TIMP-3 in non-small cell lung cancer. *Med Oncol.* (2013) 30:713. doi: 10.1007/s12032-013-0713-6
40. Barbone F, Bovenzi M, Cavallieri F, Stanta G. Cigarette smoking and histologic type of lung cancer in men. *Chest.* (1997) 112:1474–9. doi: 10.1378/chest.112.6.1474
41. Mao L, Lee JS, Kurie JM, Fan YH, Lippman SM, Broxson A, et al. Clonal genetic alterations in the lungs of current and former smokers. *JNCI J Natl Cancer Inst.* (1997) 89:857–62. doi: 10.1093/jnci/89.12.857
42. Wistuba II, Virmani AK, Gazdar AF, Lam S, LeRiche J, Behrens C, et al. Molecular damage in the bronchial epithelium of current and former smokers. *JNCI J Natl Cancer Inst.* (1997) 89:1366–73. doi: 10.1093/jnci/89.18.1366
43. Kadara H, Wistuba II. Field cancerization in non–small cell lung cancer. *Proc Am Thorac Soc.* (2012) 9:38–42. doi: 10.1513/pats.201201-004MS
44. Czarnecka KH, Migdalska-Sek M, Domanska D, Pastuszak-Lewandoska D, Dutkowska A, Kordiak J, et al. FHIT promoter methylation status, low protein and high mRNA levels in patients with non-small cell lung cancer. *Int J Oncol.* (2016) 49:1175–84. doi: 10.3892/ijo.2016.3610
45. Kanoh Y, Abe T, Masuda N, Akahoshi T. Progression of non-small cell lung cancer: diagnostic and prognostic utility of matrix metalloproteinase-2, C-reactive protein and serum amyloid A. *Oncol Rep.* (2013) 29:469–73. doi: 10.3892/or.2012.2123
46. Mook OR, Frederiks WM, Van Noorden CJ. The role of gelatinases in colorectal cancer progression and metastasis. *Biochim Biophys Acta—Rev Cancer.* (2004) 1705:69–89. doi: 10.1016/j.bbcan.2004.09.006
47. Hayashita Y, Osada H, Tatematsu Y, Yamada H, Yanagisawa K, Tomida S, et al. A polycistronic MicroRNA cluster, miR-17-92, is overexpressed in human lung cancers and enhances cell proliferation. *Cancer Res.* (2005) 65:9628–32. doi: 10.1158/0008-5472.CAN-05-2352
48. Luengo-Gil G, Gonzalez-Billalabeitia E, Perez-Henarejos SA, Navarro Manzano E, Chaves-Benito A, Garcia-Martinez E, et al. Angiogenic role of miR-20a in breast cancer. *PLoS ONE.* (2018) 13:1–17. doi: 10.1371/journal.pone.0194638
49. Yan Y, Eliason S, Ries R, Hong L, Cao H, Amendt BA. Inhibition of microRNA-17 promotes the progression of thyroid cancer in a xenograft murine cancer model. *FASEB J.* (2017) 31:934.26. Available online at: https://www.fasebj.org/doi/abs/10.1096/fasebj.31.1_supplement.934.26
50. Yang R, Fu Y, Zeng Y, Xiang M, Yin Y, Li L, et al. Serum miR-20a is a promising biomarker for gastric cancer. *Biomed Rep.* (2017) 6:429–34. doi: 10.3892/br.2017.862
51. Guz M, Rivero-Müller A, Okon E, Stenzel-Bembenek A, Polberg K, Słomka M, et al. MicroRNAs-role in lung cancer. *Dis Markers.* (2014) 2014:218169. doi: 10.1155/2014/218169
52. Zhang H, Mao F, Shen T, Luo Q, Ding Z, Qian L, et al. Plasma miR-145, miR-20a, miR-21 and miR-223 as novel biomarkers for screening early-stage non-small cell lung cancer. *Oncol Lett.* (2017) 13:669–76. doi: 10.3892/ol.2016.5462
53. Molina-Pinelo S, Pastor MD, Suarez R, Romero-Romero B, De La Peña MG, Salinas A, et al. MicroRNA clusters: dysregulation in lung adenocarcinoma and COPD. *Eur Respir J.* (2014) 43:1740–49. doi: 10.1183/09031936.00091513
54. Zhang L, Shan X, Wang J, Zhu J, Huang Z, Zhang H, et al. A three-MicroRNA signature for lung squamous cell carcinoma diagnosis in Chinese male patients. *Oncotarget.* (2017) 8:86897–907. doi: 10.18632/oncotarget.19666
55. Aushev VN, Zborovskaya IB, Laktionov KK, Girard N, Cros MP, Herceg Z, et al. Comparisons of MicroRNA patterns in plasma before and after tumor removal reveal new biomarkers of lung squamous cell carcinoma. *PLoS ONE.* (2013) 8:e78649. doi: 10.1371/journal.pone.0078649
56. Selvaggi G, Scagliotti GV. Histologic subtype in NSCLC: does it matter? *Oncology.* (2009) 23:1133–40.
57. Wentz-Hunter KK, Potashkin JA. The role of miRNAs as key regulators in the neoplastic microenvironment. *Mol Biol Int.* (2011) 2011:839872. doi: 10.4061/2011/839872

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

ORIGINAL RESEARCH
published: 18 December 2019
doi: 10.3389/fonc.2019.01394

Isolation and Identification of Cancer Stem-Like Cells in Adenocarcinoma and Squamous Cell Carcinoma of the Lung: A Pilot Study

Valentina Masciale[1†], Giulia Grisendi[2,3†], Federico Banchelli[4], Roberto D'Amico[4], Antonino Maiorana[5], Pamela Sighinolfi[5], Alessandro Stefani[1], Uliano Morandi[1], Massimo Dominici[2‡] and Beatrice Aramini[1‡]*

[1] Division of Thoracic Surgery, Department of Medical and Surgical Sciences for Children & Adults, University of Modena and Reggio Emilia, Modena, Italy, [2] Division of Oncology, Department of Medical and Surgical Sciences for Children & Adults, University of Modena and Reggio Emilia, Modena, Italy, [3] Rigenerand SRL, Modena, Italy, [4] Department of Medical and Surgical Sciences for Children & Adults, Center of Medical Statistic, University of Modena and Reggio Emilia, Modena, Italy, [5] Department of Medical and Surgical Sciences for Children & Adults, Institute of Pathology, University of Modena and Reggio Emilia, Modena, Italy

OPEN ACCESS

Edited by:
Etienne Giroux Leprieur,
Hôpital Ambroise-Paré, France

Reviewed by:
Jessica Desiree Menis,
Istituto Oncologico Veneto (IRCCS), Italy
Conor Steuer,
Emory University, United States

***Correspondence:**
Beatrice Aramini
beatrice.aramini@unimore.it

[†]These authors share first authorship

[‡]These authors share last authorship

Specialty section:
This article was submitted to Thoracic Oncology, a section of the journal Frontiers in Oncology

Received: 12 August 2019
Accepted: 26 November 2019
Published: 18 December 2019

Citation:
Masciale V, Grisendi G, Banchelli F, D'Amico R, Maiorana A, Sighinolfi P, Stefani A, Morandi U, Dominici M and Aramini B (2019) Isolation and Identification of Cancer Stem-Like Cells in Adenocarcinoma and Squamous Cell Carcinoma of the Lung: A Pilot Study. Front. Oncol. 9:1394. doi: 10.3389/fonc.2019.01394

Background: Lung cancer stem cells (CSCs) share many characteristics with normal stem cells, such as self-renewal and multipotentiality. High expression of aldehyde dehydrogenase (ALDH) has been detected in many tumors, particularly in the CSC compartment, and it plays an important role in tumor proliferation, metastasis, and drug resistance. CD44 is commonly used as a cell surface marker of cancer stem-like cells in epithelial tumors. The aim of this study was to isolate and analyze cancer stem-like cells from surgically removed specimens to compare lung adenocarcinoma (ADENO) and squamous (SQUAMO) cell carcinoma.

Methods: The ALDEFLUOR assay was used to identify and sort $ALDH^{high}$ and $ALDH^{low}$ human lung cancer cells following tissue digestion. Fluorescence-activated cell sorting analysis for CD44 was performed with tumor cells. Quantitative real-time PCR was performed to assess the expression of SOX2 and NANOG as stemness markers. ALDH1A1 expression was additionally determined by immunohistochemistry. Anchorage-independent $ALDH^{high}$ cell growth was also evaluated. $ALDH^{high}$ ADENO and SQUAMO cells were cultured to analyze spheroid formation.

Results: All specimens contained 0.5–12.5% $ALDH^{high}$ cells with 3.8–18.9% CD44-positive cells. SOX2 and NANOG relative expression in $ALDH^{high}$ compared to $ALDH^{low}$ cells in ADENO and SQUAMO was analyzed and compared between the histotypes. Immunohistochemistry confirmed the presence of ALDH1A1 in the sections. SOX2 and NANOG were expressed at higher levels in the $ALDH^{high}$ subpopulation than in the $ALDH^{low}$ subpopulation only in ADENO cells, and the opposite result was seen in SQUAMO cells. *In vitro* functional assays demonstrated that $ALDH^{high}$ cells exhibited migration capacity with distinct behaviors between $ALDH^{high}$ spheres in ADENO vs. SQUAMO samples.

Conclusions: Our results highlight the importance of a better characterization of cancer stem-like cells in ADENO and SQUAMO histotypes. This may suggest new differential approaches for prognostic and therapeutic purposes in patients with non-small-cell lung cancer.

Keywords: cancer stem-like cells, non-small-cell lung cancer, lung adenocarcinoma, lung squamous cell carcinoma, CSC marker, aldehyde dehydrogenase

BACKGROUND

Lung cancer is the most common cancer worldwide, accounting for 1.8 million new cases and 1.6 million deaths in 2012; the number of deaths worldwide is expected to grow to 3 million by 2035 (1, 2). Non-small-cell lung cancer (NSCLC) accounts for 85–90% of all lung cancers. The primary treatment is surgery for early stages (stages I and II) and chemotherapy, radiotherapy, and/or immunotherapy for advanced-stage disease (3–6). Chemotherapy drugs cannot differentiate between tumor cells and normal cells while functioning; the treatment-related adverse effects are noticeably strong and therefore feared by patients. It was not until the emergence of targeted therapy based on molecular typing that the survival period of patients with advanced NSCLC was improved to several years. Until 2013, immunotherapy was crowned as the first place scientific breakthrough (7). The efficacy of immunotherapy for those without targetable oncogene mutations was proven from second-line treatment (8–12) to first-line treatment (13, 14). Through long-term follow-up, immunotherapy has also shown that it has the greatest potential long-term clinical benefit (15, 16), even though the efficacy is not fully satisfactory (17–24). Indeed, similar to targeted therapy, patients may eventually develop resistance to immunotherapy (25, 26), and some may even suffer hyperprogression after immunotherapy (27, 28). The problem of resistance has not yet been studied; however, recent data suggest that cancer stem cells (CSCs) with characteristics of self-renewal may be resistant to these therapies (29). Understanding the role of CSCs in lung cancer may be very important and useful for identifying future targets. Indeed, the development of methods for the isolation and characterization of CSCs from primary tumors is a critical step in understanding the processes that mediate chemoresistance and for the development of therapeutic strategies to overcome this resistance, including promising immunotherapy approaches (29–31). To date, cancer cell lines have been the most frequently used tools to study lung CSCs (15, 16, 32). The identification of a specific marker for CSCs in the lung remains controversial (33, 34). Current studies provide increasing evidence for the existence of CSCs using several specific biomarkers (e.g., CD133, CD90, and CD44) translated from studies of human hematological malignancies (35–37) and solid tumors (38–47). In particular, aldehyde dehydrogenase (ALDH) activity is an important functional marker of normal and malignant stem/progenitor cells (47–51). In addition, CSCs possess high ALDH activity, especially for the predominant ALDH isozymes, ALDH1A1, and ALDH1A3. Cortes-Dericks et al. (51) showed that the flow cytometry-based ALDEFLUOR assay could be used to select $ALDH^{high}$ and $ALDH^{low}$ populations to discriminate the cancer stem-like cell population from non-cancer stem-like cells. An enrichment of CSCs in the $ALDH^{high}$ population was also described in NSCLC patients and cell lines (52). In addition, several key regulators have been described as essential for the maintenance of a progenitor cell state under both normal and cancerous conditions (e.g., SOX2 and the homeobox protein NANOG) (53, 54). Following these investigations, the aim of the present study was to identify cancer stem-like cells in primary human lung cancer cells obtained from surgical specimens and to assess the differences and similarities between adenocarcinoma (ADENO) and squamous (SQUAMO) cell carcinoma using a combination of ALDH and CD44.

Abbreviations: NSCLC, non-small cell lung cancer; SCLC, small cell lung cancer; ALDH, aldehyde dehydrogenase; FACS, fluorescence-activated cell sorting; CSCs, cancer stem cells; SSC, side scatter; FSC, forward scattered; ECM, extracellular matrix; ADENO, adenocarcinoma; SQUAMO, squamous cell carcinoma.

METHODS

The identification of cancer stem-like cells and the assessment of the differences and similarities between ADENO and SQUAMO were carried out by performing real-time PCR (RT-PCR), 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTS) assays, and sphere cultures.

Collection of Tumor Specimens

This study was approved by the Regional Ethical Committee of Modena University Hospital and performed according to the guidelines of the Helsinki Convention. Upon signed informed consent, human lung cancer tissues were obtained from four consecutive patients with ADENO and four consecutive patients with SQUAMO who underwent major surgical lung resection between October 2017 and January 2018 at the Division of Thoracic Surgery of the University Hospital of Modena for stage I, II, or IIIA NSCLC (8th TNM) (**Table 1**). The collection of tumor tissues was carried out during surgery and was set according to the availability of the pathologists involved in our study. The excision of the tumor tissue was performed only from the primary lung nodule. The microscopic features of the cancer cells and immunohistochemistry were used to assess the histological diagnosis.

Dissociation of Primary Tissues

Freshly obtained tumor tissues (within 1–2 h after surgical removal) were washed in sterile Dulbecco's phosphate-buffered

TABLE 1 | Patients characteristics, cellular yield from each sample, viability, and aldehyde dehydrogenase (ALDH) expression determined by fluorescence-activated cell sorting (FACS).

		Adenocarcinoma ($n = 4$)	Squamous cell carcinoma ($n = 4$)	All patients ($n = 8$)
PATIENTS CHARACTERISTICS				
Age (years)	Mean ± SD	66.7 ± 9.4	73.7 ± 7.9	71 ± 8.5
Gender (male)	*n* (%)	3 (75.0%)	3 (75.0%)	6 (75.0%)
Smoker (yes)	*n* (%)	4 (100.0%)	4 (100.0%)	8 (100.0%)
Stage (8th TNM)				
IA3	*n* (%)	1 (25.0%)	1 (25.0%)	2 (25.0%)
IIA	*n* (%)	2 (50.0%)	0 (0.0%)	2 (25.0%)
IIB	*n* (%)	1 (25.0%)	1 (25.0%)	2 (25.0%)
IIIA	*n* (%)	0 (0.0%)	2 (50.0%)	2 (25.0%)
Neoadiuvant Chemotherapy	*n* (%)	1 (25.0%)	0 (0.0%)	1 (12.5%)
SAMPLE CHARACTERISTICS				
Weight (g)	Mean ± SD	1.0 ± 0.9	1.0 ± 0.6	1.0 ± 0.7
Cellular yield (million cells/g)	Mean ± SD	18.2 ± 8.6	20.4 ± 8.2	19.3 ± 7.9
FACS ANALYSIS				
7-AAD negative	Mean ± SD	94.3 ± 5.2%	90.5 ± 6.8%	92.4 ± 6.0%
$ALDH^{high}$	Mean ± SD	3.7 ± 5.9%	4.2 ± 3.9%	4.0 ± 4.6%

SD, standard deviation. 7-Amino-actinomycin D (7-AAD) negative cells are expressed as percentage of total number of sorted cells. $ALDH^{high}$ are expressed as percentage of 7-AAD negative cells.

saline (PBS) (L1825-BC—Merck Millipore) and mechanically minced into small pieces (2–4 mm). Minced samples were digested using a tumor dissociation kit in a disposable gentle MACS™ C-Tube (Miltenyi) according to the manufacturer's instructions. Samples were digested for 60 min at 37°C in a gentle MACS Octo dissociator and filtered through 70-μm sterile cell strainers, centrifuged at 300×*g* for 5 min, and resuspended in a mixture of Dulbecco's modified Eagle medium (DMEM) and Ham's F12 media (2:1) (Gibco) containing 50 IU/ml penicillin–streptomycin and 4 mM glutamine. Finally, viable cells were counted using an optic phase contrast microscope.

ALDEFLUOR Assay

Single-cell suspensions of the primary tumor cells from the surgical tumor specimens were diluted in ALDEFLUOR assay buffer containing BODIPY-aminoacetaldehyde (STEMCELL Technologies, Vancouver, BC). The assay was performed according to the manufacturer's protocol. Briefly, at least 5 million tumor cells were resuspended in ALDEFLUOR buffer (5 μl/10^6) and stained with ALDEFLUOR substrate. Immediately after, 5×10^5 cells were transferred to a control tube containing 5 μl diethylaminobenzaldehyde, which is a specific inhibitor of ALDH. Both control and test samples were incubated for 45 min at 37°C protected from light. Following incubation, the cells were centrifuged at 300×*g* for 5 min. The cell pellet was resuspended in 1 ml ALDEFLUOR assay buffer. Cell morphology was evaluated using side scatter (SSC) and forward scatter (FSC). Dead cells were excluded using 7-amino-actinomycin D (7-AAD) staining. Cell sorting and ALDH analysis were performed using a FACSAria III instrument (Becton Dickinson, Franklin Lakes, NJ). The results were analyzed using fluorescence-activated cell sorting (FACS) Diva software (Becton Dickinson). The gating strategy included the $ALDH^{high}$ gate, which was set at least one log apart from the $ALDH^{low}$ gate. Sorted cells were promptly lysed for gene expression analysis.

FACS Analyses

Primary tumor cell suspensions were stained with allophycocyanin-conjugated anti-CD45 (Becton Dickinson) and phycoerythrin-conjugated anti-CD44 (BioLegend, San Diego, CA). An isotype control sample for each condition was used to exclude the autofluorescence background. Dead cells were excluded using 7-AAD staining. The gate was set based on CD45-negative cells. Analyses were performed using a FACSAria III instrument (Becton Dickinson). Data were analyzed using the FACSDiva software.

RNA Isolation and Real-Time PCR

Total cellular RNA was extracted from $ALDH^{high}$ and $ALDH^{low}$ cells using the RNeasy Mini Kit (Qiagen) according to the manufacturer's instructions. Total RNA (500 ng) was reverse transcribed using the RevertAid™ First-Strand Complementary DNA (cDNA) Synthesis Kit (Thermo Scientific). Following cDNA synthesis, RT-PCR was performed in triplicate for each sample using FAST SYBR™ Green detection chemistry (Applied Biosystems) on Step One instrument. Human SOX2, NANOG, and GAPDH were amplified using gene-specific primers (GAPDH: forward primer 5′-ACATCGCTCAGACACCATG-3′, reverse primer 5′TGTAGTTGAGGTCAATGAAGGG-3′; SOX2: forward primer 5′-GGAAACTTTTGTCGGAGACG-3′, reverse primer 5′-GCAGCGTGTACTTATCCTTC-3′; NANOG: forward

primer 5′AGAAATACCTCAGCCTCCAG-3′, reverse primer 5′-CGTCACACCATTGCTATTCTT-3′). The cycling parameters consisted of denaturation at 95°C for 10 min; and 40 cycles of 15 s at 94°C, 30 s at 60°C, and 1 min at 72°C; followed by a continuous melting curve.

Immunohistochemistry

Slides were deparaffinized with xylene, rehydrated in a graded alcohol series, and washed in PBS twice for 5 min each. The sections were heated in 10 mM sodium citrate buffer, pH 6.0, for 15 min in a 95°C water bath for antigen retrieval. PBS washes (5 min each) were performed until the buffer cooled down. Endogenous peroxidase activity was blocked via incubation in 3% H_2O_2 at room temperature for 10 min. Blocking serum was added in a dropwise manner at room temperature for 20 min to reduce the non-specific background. Samples were incubated with the anti-ALDH1A1 monoclonal antibody (ab-134188; 1:100 dilution; Abcam, Cambridge, MA, USA) overnight at 4°C. Sections were washed in PBS three times for 2 min and then incubated with a biotinylated secondary antibody (PK-4001; Vector Labs, USA) for 30 min at room temperature. The slides were subsequently incubated with ABC-HRP (PK-4001; Vector Labs, USA) for another 30 min, washed in PBS, and stained with 3,3-diaminobenzidine. Finally, the sections were counterstained with Mayer's hematoxylin, dehydrated, and mounted. Images were collected using a Zeiss Axioskop microscope with a Zeiss Axiocam ICc3 High-Resolution Microscope Camera. The scoring of the ALDH1A1 staining was performed by two independent investigators who were blinded to the patients' clinicopathological characteristics. Sections were scored independently. Immunoreactivity was scored using a semiquantitative method based on the ALDH positivity of the tumor cells as follows: 0 (<5% positive), 1 (5–25% positive), 2 (>25–50% positive), 3 (>50–75% positive), and 4 (>75% positive) (55).

Cell Transformation Assay

Cell Biolabs CytoSelect™ 96-well cell transformation assay (cell recovery compatible, fluorometric) was used to analyze the anchorage-independent growth of $ALDH^{high}$ cells, and the MCF-7 cell line was used as a positive control. $ALDH^{high}$ cells were harvested and cultivated for a maximum of 48 h in appropriate serum-free medium, as described below. MCF-7 cells were cultivated in DMEM (Gibco) containing 50 IU/ml penicillin–streptomycin and 4 mM glutamine in the presence of 10% FBS (Euroclone). Cells were used at a concentration of 9,000 cells per well of the 96-well plate, and the growth kinetics on day 0 (T0) and day 8 (T8) were chosen to measure cell growth. This kit provided the soft agar material, solubilization solution, lysis buffer, and Cyquant® GR Dye. The dye binds nucleic acids, and the relative fluorescence units (RFUs) were quantified to reveal a relative quantity of cells based on nucleic content.

MTS Assay

The MTS Cell Proliferation Kit (Abcam) was used to measure the cell proliferation rate at six different time points after seeding (0, 1, 2, 7, 14, and 21 days). In a final volume of 200 μl of cell culture medium, 20 μl of MTS was added and incubated for 4 h at 37°C in standard culture conditions. After incubation, the optical density was measured at 490–500 nm by a Glomax Multi+ Detection System (Promega).

Tumor Sphere-Forming Assay

$ALDH^{high}$ and $ALDH^{low}$ tumor spheres were dissociated into single-cell suspensions, and 50,000 cells from four different patients, two ADENO and two SQUAMO, were transferred to 24 ultralow attachment well plates. Cells were cultured in a mixture of serum-free DMEM and Ham's F12 media (2:1) (Gibco) containing 50 IU/ml penicillin–streptomycin and 4 mM glutamine supplemented with 5 μg/ml insulin, 10 ng/ml epidermal growth factor (EGF), 20 ng/ml basic fibroblast growth factor, 0.18 nM adenine, and 2 nM triiodotironin. The cells were cultured in 5% CO_2 at 37°C for 2 weeks, and the media were replaced or supplemented with fresh growth factors twice a week. The entire well was digitally photographed using inverted phase-contrast microscopy (Zeiss Axioskop and Axiocam ICc3 color camera). All images were analyzed using the AxioVision software (Zeiss). The total number of spheres was counted, and sphere areas were manually measured at three different time points: 1, 2, and 3 weeks from seeding (56, 57).

RT-PCR Data Analysis

We included four patients with ADENO and four patients with SQUAMO in the analysis, for a total of eight patients. This study uses a three-factor full factorial experimental design with replications, with factors such as (1) $ALDH^{high}$ and $ALDH^{low}$ cells; (2) SOX2, NANOG, and GAPDH genes; and (3) ADENO and SQUAMO histotypes. Replications are represented by triplicates. The analysis was performed using a linear mixed-model approach (58), which allows formal statistical hypothesis testing of relative gene expression. All cycle threshold values ≥36 were set as equal to 36. We assessed the relative messenger RNA (mRNA) expression of SOX2 and NANOG genes, normalized to the expression of the housekeeping gene, GAPDH, in $ALDH^{high}$ cells compared to $ALDH^{low}$ cells by means of a linear mixed regression model. The dependent variable was cycle threshold (Ct), whereas the independent variables were ALDH (high vs. low), gene (SOX2 and NANOG vs. GAPDH, which is the reference category), histotype (ADENO vs. SQUAMO), and pairwise and three-way interactions, all of which were analyzed as fixed-effect factors. The model also included a random intercept and a random ALDH–histotype interaction term that was specific for each patient to take into account correlations among cycle threshold values. The following parameters of interest were examined: (1) relative expression of SOX2 and NANOG in $ALDH^{high}$ cells compared to $ALDH^{low}$ cells in adenocarcinoma; (2) relative expression of SOX2 and NANOG in $ALDH^{high}$ cells compared to $ALDH^{low}$ cells in squamous cell carcinoma; and (3) differences between ADENO and SQUAMO in SOX2 and NANOG relative expression in $ALDH^{high}$ cells compared to $ALDH^{low}$ cells. The relative expression of SOX2 and NANOG in $ALDH^{high}$ cells compared to $ALDH^{low}$ cells, using GAPDH as the housekeeping gene, is reported as the

fold change and the difference in cycle thresholds (equal to $-\log_2$ fold change) with 95% confidence intervals and *p*-values. Comparisons between ADENO and SQUAMO for SOX2 and NANOG relative expression in $ALDH^{high}$ cells compared to $ALDH^{low}$ cells are reported as *p*-values. All tests were two-sided *t*-tests using the Satterthwaite method for degrees of freedom.

Statistical Analysis

Continuous variables were expressed as the mean ± standard deviation (SD) and the range, and categorical variables were expressed as absolute and percent frequencies. Statistical analyses of RT-PCR data were described in the previous paragraph. All statistical analyses were performed with R 3.4.3 software (The R Foundation for Statistical Computing, Wien) with $p < 0.05$ as the significance level.

Sample Size

No formal sample size estimation was carried out in this pilot study. Therefore, the number of patients included in the analyses was only based on resource availability. The enrolled patients were those who underwent major surgical lung resection for stage I, II, or IIIA NSCLC in our division over a time period of 4 months. We included four consecutive patients with ADENO and four consecutive patients with SQUAMO to balance the two types of patients according to our full factorial experimental design.

Genomics

Genomics data were routinely recorded in our Hospital for ADENO. DNA was extracted from formalin-fixed and paraffin-embedded blocks of each tumor sample. Extraction was performed with the QIAamp DNA Mini Kit (Qiagen, Hilden, Germany), and DNA was quantified with Xpose-NGS (Trinean NV, Gentbrugge, Belgium). Mutations were detected in genome-amplified DNA using the high-throughput genotyping platform Sequenom MassARRAY System (Sequenom, San Diego, CA, USA) and the Myriapod Colon Status Kit (Diatech Pharmacogenetics, Italy) following the manufacturer's protocol. This molecular array allows for the identification of the most important mutations of the *KRAS*, *NRAS*, *BRAF*, *EGFR, PIK3CA*, and *ERBB2* genes.

RESULTS

Lung Cancer Tissues From Patients Yield Sufficient Numbers of Living Cells After Dissociation

All eight patients enrolled in the study (mean age, 71 ± 8.5 years old; range, 61–83; six male, two female, all smokers) underwent a lobectomy by lateral thoracotomy. Four patients were diagnosed with adenocarcinoma of the lung (mean age, 66.7 ± 9.4 years old; range, 61–79; three male, one female), and four patients (mean age, 73.7 ± 7.9 years old; range, 65–83; three male, one female) were diagnosed with squamous cell carcinoma (**Table 1**). Surgical tumor specimens (mean, 1.0 ± 0.7 g; range, 0.2–2.2) were obtained from each patient and used for the experiments, with similar specimen weights in the ADENO and SQUAMO groups (**Table 1**). A procedure combining mechanical dissociation with enzymatic degradation of the extracellular matrix that maintained tissue structural integrity was used to obtain single-cell suspensions from the surgical tumor specimens. The average cellular yield was a mean of 19.3 ± 7.9 million cells per gram (range, 10.0–30.3), with similar cellular yields in ADENO and SQUAMO (**Table 1**). Good cell viability was further confirmed by FACS analysis.

ALDH-Positive Stem-Like Cells Were Identified in Primary Lung Cancer Tissue

Tumor tissue dissociation efficiently released cancer cells characterized by a heterogeneous morphology, as illustrated by widespread FSC and SSC values (**Figure 1A**). The mean viability of the samples was 92.4 ± 6.0% (range, 82.9–99.4%) based on 7-AAD staining (**Table 1**). These data further confirmed that the developed dissociation procedure was a non-toxic approach to isolating cancer cells from tumor tissues (**Figure 1**). The putative CSCs were physically separated from the bulk parental tumor cells and recovered by FACS according to the following gating strategy: Tumor cells were first identified based on their morphological parameters (FSC/SSC, **Figure 1A**, gate P1), and ALDH activity was measured in the 7-AAD-negative cell population only (**Figure 1B**). $ALDH^{low}$ and $ALDH^{high}$ cells were selected and sorted (**Figure 1C**). An $ALDH^{high}$ subpopulation was identified in all samples (mean, 4.0 ± 4.6%; range, 0.5–12.5%, with two samples above 5%) of all viable lung cancer cells (**Table 1**), which indicates that it was possible to preserve lung cancer cells using a rapid dissociation protocol that allowed the identification of putative CSCs.

Primary Lung Tumor Cells Express CD44

The surface marker CD44 was investigated as a possible marker for cancer stem-like cells. 7-ADD was used to identify viable cells, and CD45 staining was used to exclude CD45-positive cells (**Figure 2**) (59).

We found that $ALDH^{high}$ and CD44-positive cells had comparable expression in our samples (4.0 ± 4.6 and 11.5 ± 7.7%, respectively), and there was a moderate positive correlation (Pearson correlation = 0.52).

SOX2 and NANOG in $ALDH^{high/low}$ Cells in Adenocarcinoma and Squamous Cell Carcinoma

A total of 143 Ct values were available. Two patients (one ADENO and one SQUAMO) had SOX2 and NANOG Ct values in triplicate above 36 cycles in $ALDH^{low}$ cells, and one ADENO patient had SOX2 and NANOG Ct values in triplicate above 36 cycles in $ALDH^{low}$ cells and NANOG Ct values in triplicate above 36 cycles in $ALDH^{high}$ cells. A total of 25 (17.5%) Ct values were above 36 and set equal to 36 for data analysis. The results from linear mixed-model analysis are reported in **Table 2**. The fold changes in ADENO were 20.72 (95% CI = 0.68; 635.58, $p = 0.0755$) and 25.49 (95% CI = 2.29; 283.44, $p = 0.0147$) for SOX2 and NANOG, respectively. The fold changes in SQUAMO were 0.14 (95% CI = 0.02; 1.13,

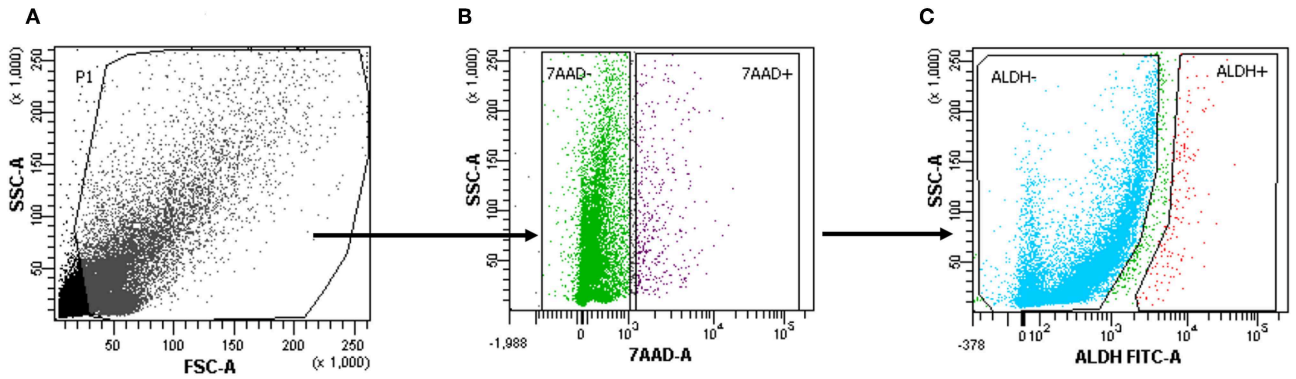


FIGURE 1 | Cells with high aldehyde dehydrogenase (ALDH) activity in one patient sample (1.2%). **(A–C)** The gating strategy of a representative fluorescence-activated cell sorting (FACS) analysis of a primary tumor cell suspension in one patient. 7-Amino-actinomycin D (7-AAD) was used to assess ALDH^{highorlow} on the live population of cells.

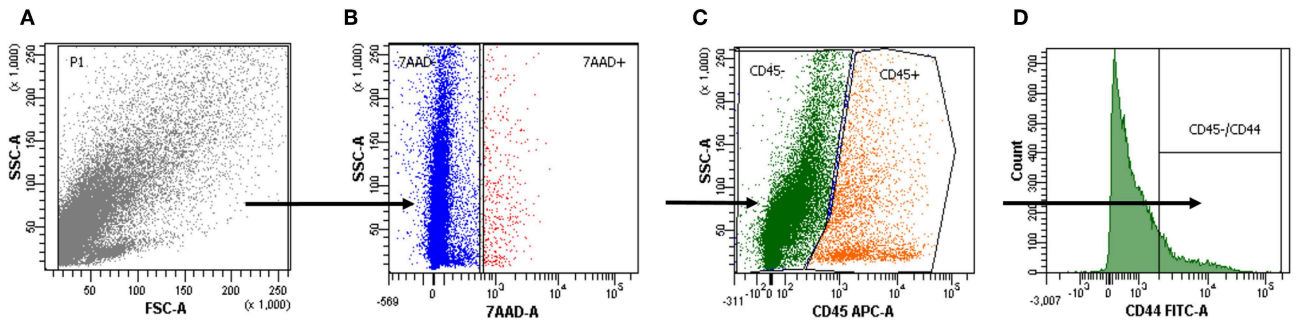


FIGURE 2 | Cell positivity for CD44. **(A–D)** The gating strategy of a representative cytofluorimetric analysis of a primary tumor cell suspension. 7-Amino-actinomycin D (7-AAD) was used to exclude the CD45-positive cells in the live population to further analyze tumor cell positivity for CD44.

$p = 0.1022$) and 0.07 (95% CI = 0.02; 0.31, $p = 0.0073$) for SOX2 and NANOG, respectively. These results are reported in **Figure 3** by both the fold change and the cycle threshold difference scales. The differences in relative expression between ADENO and SQUAMO were statistically significant ($p = 0.0101$ and $p = 0.0005$ for SOX2 and NANOG, respectively).

ALDH Positivity in Digested Samples Reflects Immunohistochemical Scoring in NSCLC

To further evaluate the ALDH expression pattern in the NSCLC samples, ALDH1A1 immunohistochemistry was scored as previously reported (55). Tissue sections were examined at 10× magnification to characterize the overall staining pattern and at 20× magnification for a more accurate evaluation of the cells to assign the appropriate values. As expected, normal bronchial epithelium and macrophages showed ALDH1A1 expression (49).

All eight patient samples showed a broadly similar intensity of ALDH1A1 staining in the cancerous fraction (i.e., all had scores of 0 with <5% positive tumor cells), which is consistent with the FACS data (**Figure 4**).

Cell Transformation

The Cell Biolabs CytoSelect™ 96-well cell transformation assay did not involve subjective manual counting of colonies, but it used a fluorescent measurement of Cyquant® GR Dye that bound to nucleic acids to quantify the number of cells based on nucleic assay content. We compared the cell proliferation at T0 and T8 between the MCF-7 breast cancer cell line and ALDH^{high} cells. The MCF-7 cell line grew from 130 to 1,233 RFU, and the ALDH^{high} cells rose mildly from 158 to 177 RFU (**Figure 5**).

MTS Cell Proliferation Assay

Cell proliferation was evaluated in ALDH^{high} cells from seeding (day 0) until the end of the culture (day 21). An MTS-based assay revealed that our culture protocol did not affect the proliferation of ALDH^{high} cells. There was a trend of growth that extended until day 21, without any notable drop in cell growth over time (**Figure 6**).

General Characteristics of Tumor Spheres

ALDH^{high} cells of the ADENO and SQUAMO patients were maintained in low attachment cultures in the absence of serum for up to 3 weeks (**Figure 7.1**).

TABLE 2 | SOX2 and NANOG in ALDH$^{high/low}$ cells.

	ΔCt Difference (95% CI)	FC (95% CI)	*p*-value
SOX2			
Adenocarcinoma	−4.37 (−9.31; 0.57)	20.72 (0.68; 635.58)	0.0755
Squamous cell carcinoma	2.79 (−0.17; 5.76)	0.14 (0.02; 1.13)	0.1022
Difference			0.0101
NANOG			
Adenocarcinoma	−4.67 (−8.15; −1.20)	25.49 (2.29; 283.44)	0.0147
Squamous cell carcinoma	3.80 (1.71; 5.88)	0.07 (0.02; 0.31)	0.0073
Difference			0.0005

95% CI, 95% confidence interval; FC, fold change. Relative gene expression of SOX2 and NANOG in ALDH^{high} cells compared to ALDH^{low} cells in adenocarcinoma and squamous cell carcinoma; results from linear mixed model analysis.

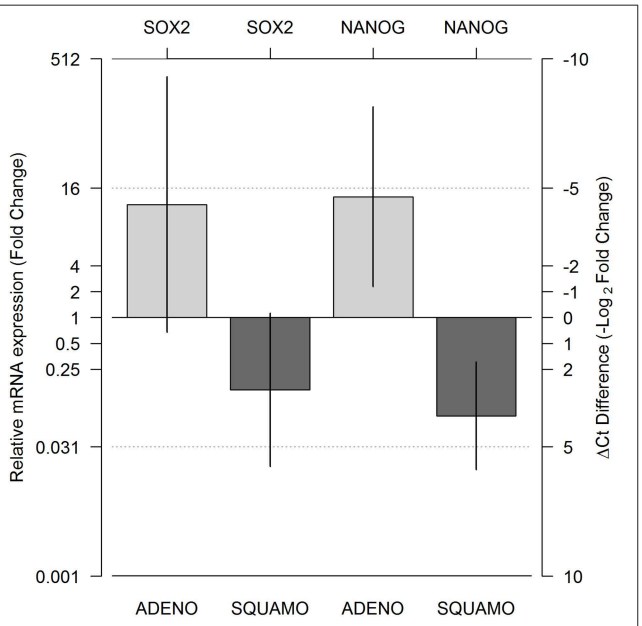


FIGURE 3 | Relative messeger RNA (mRNA) expression of SOX2 and NANOG in ALDH^{high} compared to ALDH^{low} cell populations, using GAPDH as the reference gene. Expression was measured for the ALDH^{low} and ALDH^{high} cell populations in adenocarcinoma and squamous cell carcinoma histotypes using real-time PCR (RT-PCR), and relative expression comparing ALDH^{high} and ALDH^{low} was calculated by means of a linear mixed model. The two light gray bars represent ADENO, and the two dark gray bars represent SQUAMO. The error bars represent the 95% confidence intervals. ADENO, adenocarcinoma; SQUAMO, squamous cell carcinoma.

The tumor spheres that formed in each well were counted and measured for area in ADENO and SQUAMO patients at three different time points (1, 2, and 3 weeks), as shown in **Figure 7.2**. There was a tendency for ADENO to produce a higher number of spheres than SQUAMO, and the spheres produced by ADENO also exhibited greater area than SQUAMO. We observed that the spheres in ADENO had a tendency to grow in area and form larger spheres at 2 and 3 weeks, but there was no evidence of this in SQUAMO spheres, whose distribution did not significantly change over time. In contrast, ALDH^{low} cells of both histotypes died within 3 days, as shown in **Figure 8**.

Genomics

Genomics data were recorded for four ADENO patients. Two of them had a KRAS mutation, one had an EGFR mutation and one patient was wild type for the analyzed mutations.

DISCUSSION

The CSC theory elucidates the origin of tumors, tumor development, metastasis, relapse, and drug resistance (60, 61). Therefore, the establishment of a reliable and efficient method for the isolation, manipulation, and characterization of CSCs is controversial, presumably due to the difficulty of identifying a specific marker. Thirty years ago, Carney and colleagues described a rare population of cells (<1.5%) in small and NSCLC samples that formed colonies in soft agar (62). When inoculated into athymic nude mice, these cells recapitulated the original lung cancer, which suggested that they had progenitor cell features (63). Over the last decade, several investigators isolated tumorigenic cell lines from lung cancers using different phenotypic cancer cell characteristics (48).

In the past, different methods have been used to identify CSCs, such as side population analysis, selection in culture, and cell sorting for a specific marker (64–67). Of all the markers explored, the CD133 marker has received the most attention (48). However, in our samples, the CD133 marker was not useful because of a lack of detection, as described previously (48). Consequently, we analyzed our population for CD44, which is a transmembrane receptor for hyaluronic acid that is a CSC marker of several stem cell-like properties (68). In addition, we used the ALDEFLUOR assay to isolate cancer stem-like cells, as previously described by Sullivan et al. (52). Interestingly, our results confirmed comparable ALDH^{high} and CD44 positive expression.

However, Sullivan et al. identified CSCs in a panel of 11 NSCLC tumor samples, 45 NSCLC lines, and 7 SCLC lines (52) that are used to study ALDH activity and sorted a subpopulation of NSCLC stem-like cells dependent on Notch signaling. Our study used the same method (52), but we focused on analyzing the differences and similarities between adenocarcinoma and squamous cell carcinoma cancer stem-like cells. Our hypothesis was supported by the fact that these populations are the most frequent histotypes in lung cancer patients and account for 50% of adenocarcinoma patients and 30% of squamous cell carcinoma.

We investigated the ALDH$^{high/low}$ populations in both histotypes for the mRNA expression of SOX2 and NANOG, which are stemness-related genes in normal and cancer cells (60, 61).

The RT-PCR data from our patients revealed more SOX2 and NANOG expression in ALDH^{high} cells than in ALDH^{low} cells in adenocarcinoma. However, the opposite result was obtained for squamous cell carcinoma, in which lower SOX2 and NANOG

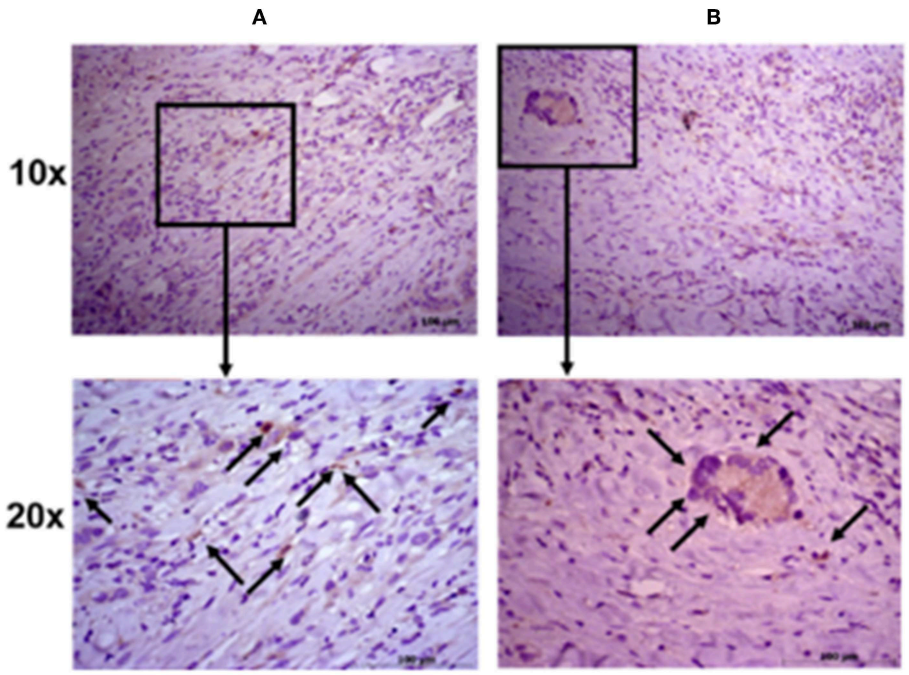


FIGURE 4 | Immunohistochemical staining intensity of aldehyde dehydrogenase. Representative images of two patient samples in squamous cell carcinoma **(A)** and adenocarcinoma **(B)**. Images were taken at 10× (upper panels) and 20× (lower panels) magnification; black arrows indicate positive cells.

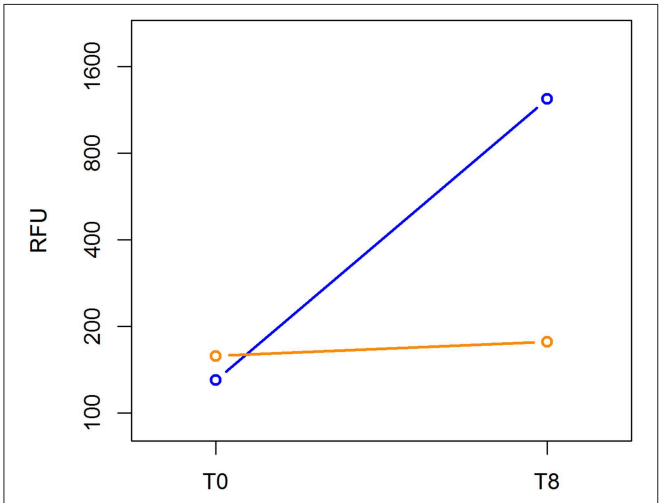


FIGURE 5 | Cell transformation assay. The MCF-7 cell line and ALDH^{high} cells were compared for growth ability in a semisolid agar substrate. Two different time points were evaluated: 0 and 8 days. In the vertical axis, DNA content of each sample was measured using a fluorescent signal released by Cyquant GR Dye. Time points are represented in the horizontal axis.

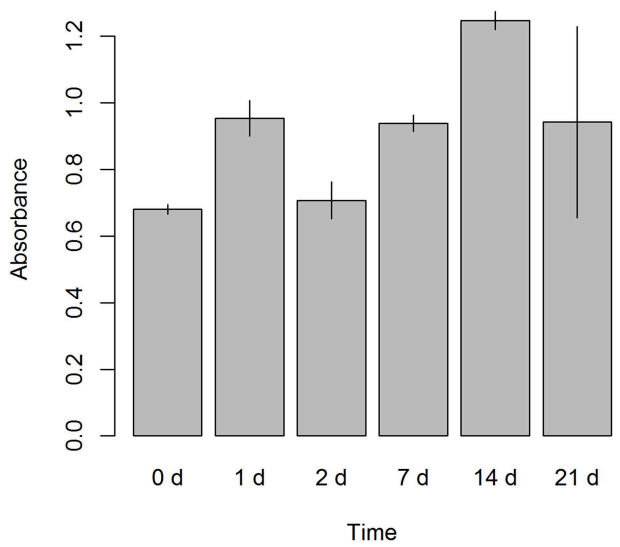


FIGURE 6 | Cell proliferation assay. A colorimetric assay, 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTS) based, was performed on ALDH^{high} cells after seeding. Six different time points were evaluated (0, 1, 2, 7, 14, and 21 days). Absorbance was measured at 490 nm.

expression was found in ALDH^{high} cells than in ALDH^{low} cells. Therefore, there was a concordant trend for SOX2 and NANOG relative mRNA expression, even though only the relative expression of NANOG reached statistical significance. There was a statistically significant difference between the relative mRNA expression of SOX2 and NANOG in adenocarcinoma compared to squamous cell carcinoma. The expression of these genes was discussed in previous attempts to find a connection between these stemness genes and the clinicopathological features of the tumor

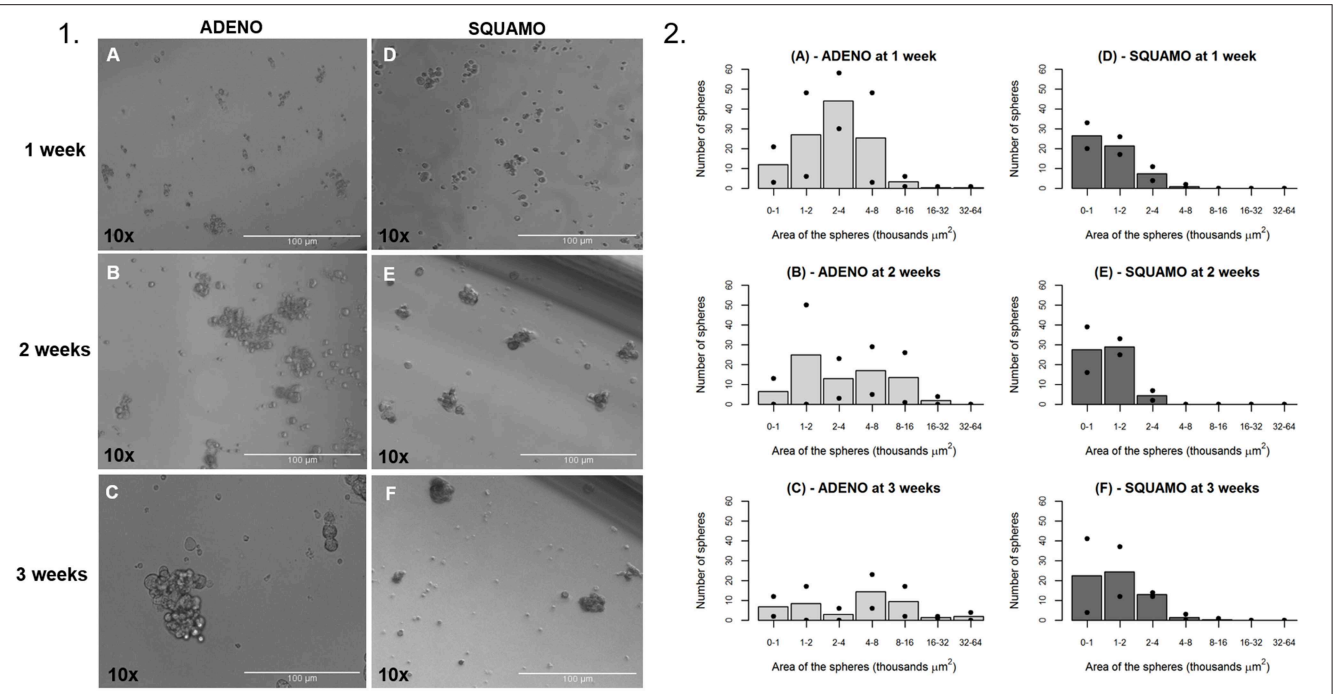


FIGURE 7 | General characteristics over time of tumor spheres in adenocarcinoma and squamous cell carcinoma. **(1A–C)** ALDH^{high} spheres from ADENO at 1, 2, and 3 weeks. **(1D–F)** ALDH^{high} spheres from SQUAMO at 1, 2, and 3 weeks. **(2A–F)** The number of spheres and their areas were calculated for two adenocarcinoma patients and two squamous cell carcinoma patients 1, 2, and 3 weeks after seeding. Points represent values for individual patients, and bars represent average values. ADENO, adenocarcinoma; SQUAMO, squamous cell carcinoma.

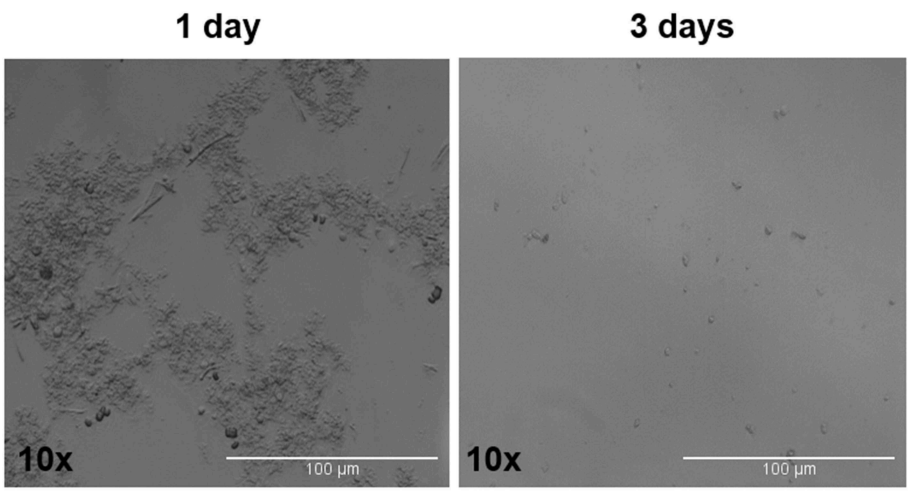


FIGURE 8 | ALDH^{low} cells phase-contrast microscopy. ALDH^{low} cells in low-attachment serum-free culture showed difficult and slow growth that completely stopped at 3 days.

(68–70). Therefore, our study adds a new aspect by considering the existence of different cancer-stem-like cell populations for these two histotypes. However, our data suggest an enrichment of cells with stemness characteristics.

To further confirm CSC-like phenotypes, we analyzed the ability of cells to form tumor spheres in serum-free low-attachment cultures. Tumor sphere formation assays revealed a different pattern in sphere formation, dimension, and growth between adenocarcinoma and squamous cell carcinoma. The former had a tendency to produce a greater number of spheres and larger spheres than the latter. Moreover, we observed growth of adenocarcinoma spheres until the third week, but spheres from the squamous cell carcinoma did not increase in number or size. This result may be related to the distinct aggressiveness and

clinicopathological characteristics of adenocarcinoma compared to squamous cell carcinoma.

Furthermore, genomics data highlighted the presence of genetic mutations in ADENO; however, the relationship of these mutations with the cancer stem-like cells in our study is unclear.

LIMITATIONS OF THE STUDY

The main limitation of the study is the low number of patients included in the analyses (four ADENO and four SQUAMO). The selection of these few patients could have affected our results as well as the statistical power. Nevertheless, we have included consecutive patients—which prevents from selection bias—and we have observed statistically significant differences in gene expression between ADENO and SQUAMO, which partially counterbalances the issue regarding the low statistical power. Moreover, in this study, test samples were obtained by primary cell cultures derived from patients, which is more difficult to obtain than tumor cell lines. On the basis of these limitations, the results obtained in our pilot study should be confirmed by more extensive studies.

CONCLUSION

Even with limited evidence due to the low number of patient samples, our study showed differences between adenocarcinoma and squamous cell carcinoma related to the analyzed stemness genes. $ALDH^{high}$ cancer stem-like cells in adenocarcinoma showed stemness characteristics in gene expression and spheroid culture studies, but squamous cell carcinoma stemness characteristics were not completely clear because of the discrepancy between genes and cellular behavior.

To summarize, our results highlight the importance of a better characterization of cancer stem-like cells in ADENO and SQUAMO histotypes. This may suggest new differential approaches for prognostic and therapeutic purposes in patients with NSCLC.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics committee at University Hospital of Modena, MODENA, Italy, on 17 March 2017, Prot. No. 914/C.E. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

The idea for the manuscript was conceived in September 2016 by BA and MD and was further developed by VM, GG, FB, RD'A, AM, and AS. AM and PS were involved in histopathological diagnosis. BA, VM, and FB wrote the first draft of the manuscript. BA and UM have been involved in surgery and tissue collection. VM and GG performed laboratory experiments, whereas FB and RD'A performed the statistical analysis. BA, VM, FB, MD, RD'A, AM, and UM reviewed the manuscript and were involved in its critical revision before submission. All authors read and approved the final manuscript.

FUNDING

The Project had been supported in parts by funds from the Division of Thoracic Surgery of the University Hospital of Modena and from the Laboratory of Cellular Therapy of the University of Modena and Reggio Emilia, from unrestricted grant from Myriad Inc. (US) and from the Italian Ministry of Education, University and Research: Departments of Excellence 2017.

REFERENCES

1. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer*. (2015) 136:E359–86. doi: 10.1002/ijc.29210
2. Didkowska J, Wojciechowska U, Manczuk M, Łobaszewski J. Lung cancer epidemiology: contemporary and future challenges worldwide. *Ann Transl Med*. (2016) 4:150. doi: 10.21037/atm.2016.03.11
3. Vansteenkiste J, Crino L, Dooms C, Douillard JY, Faivre-Finn C, Lim E, et al. ESMO Consensus guidelines: early stage non-small cell lung cancer consensus on diagnosis, treatment and follow-up. *Ann Oncol*. (2014) 25:1462–74. doi: 10.1093/annonc/mdu089
4. Ettinger DS, Wood DE, Aisner DL, Akerley W, Bauman J, Chirieac LR, et al. Non-small cell lung cancer, version 5.2017, NCCN clinical practice guidelines in oncology. *J Natl Compr Canc Netw*. (2017) 15:504–35. doi: 10.6004/jnccn.2017.0050
5. Zappa C, Mousa SA. Non-small cell lung cancer: current treatment and future advances. *Transl Lung Cancer Res*. (2016) 5:288–300. doi: 10.21037/tlcr.2016.06.07
6. Snyder V, Reed-Newman TC, Arnold L, Thomas SM, Anant S. Cancer stem cell metabolism and potential therapeutic targets. *Front Oncol*. (2018) 8:203. doi: 10.3389/fonc.2018.00203
7. Dong J, Li B, Lin D, Zhou Q, Huang D. Advances in targeted therapy and immunotherapy for non-small cell lung cancer based on accurate molecular typing. *Front Pharmacol*. (2019) 10:230. doi: 10.3389/fphar.2019.00230
8. Gainor JF, Dardaei L, Yoda S, Friboulet L, Leshchiner I, Katayama R, et al. Molecular mechanisms of resistance to first- and second-generation ALK inhibitors in ALK-rearranged lung cancer. *Cancer Discov*. (2016) 6:1118–33. doi: 10.1158/2159-8290.CD-16-0596
9. Gandara DR, Paul SM, Kowanetz M, Schleifman E, Zou W, Li Y, et al. Blood-based tumor mutational burden as a predictor of clinical benefit in non-small-cell lung cancer patients treated with atezolizumab. *Nat Med*. (2018) 24:1441–8. doi: 10.1038/s41591-018-0134-3

10. Gherardi E, Birchmeier W, Birchmeier C, Vande Woude G. Targeting MET in cancer: rationale and progress. *Nat Rev Cancer*. (2012) 12 89–103. doi: 10.1038/nrc3205
11. Goldberg ME, Montesion M, Young L, Suh J, Greenbowe J, Kennedy M, et al. Multiple configurations of EGFR exon 20 resistance mutations after first- and third-generation EGFR TKI treatment affect treatment options in NSCLC. *PLoS ONE*. (2018) 13:e0208097. doi: 10.1371/journal.pone.0208097
12. Haanen JB, Robert C. Immune checkpoint inhibitors. *Prog Tumor Res*. (2015) 42:55–66. doi: 10.1159/000437178
13. Han B, Tjulandin S, Hagiwara K, Normanno N, Wulandari L, Laktionov K, et al. EGFR mutation prevalence in Asia-Pacific and Russian patients with advanced NSCLC of adenocarcinoma and non-adenocarcinoma histology: the IGNITE study. *Lung Cancer*. (2017) 113 37–44. doi: 10.1016/j.lungcan.2017.08.021
14. Hellmann MD, Ciuleanu TE, Pluzanski A, Lee JS, Otterson GA, Audigier-Valette C, et al. Nivolumab plus ipilimumab in lung cancer with a high tumor mutational burden. *N Engl J Med*. (2018) 378 2093–104. doi: 10.1056/NEJMoa1801946
15. Zakaria N, Satar NA, Abu Halim NH, Ngalim SH, Yusoff NM, Lin J, et al. Targeting lung cancer stem cells: research and clinical impacts. *Front Oncol*. (2017) 7:80. doi: 10.3389/fonc.2017.00080
16. Tiran V, Lindenmann J, Brcic L, Heitzer E, Stanzer S, Tabrizi-Wizsy NG, et al. Primary patient-derived lung adenocarcinoma cell culture challenges the association of cancer stem cells with epithelial-tomesenchymal transition. *Sci Rep*. (2016) 7:10040. doi: 10.1038/s41598-017-09929-0
17. Borghaei H, Paz-Ares L, Horn L, Spigel DR, Steins M, Ready NE, et al. Nivolumab versus docetaxel in advanced nonsquamous non-small-cell lung cancer. *N Engl J Med*. (2015) 373:1627–39. doi: 10.1056/NEJMoa1507643
18. Brahmer J, Reckamp KL, Baas P, Crinò L, Eberhardt WE, Poddubskaya E, et al. Nivolumab versus docetaxel in advanced squamous-cell non-small-cell lung cancer. *N Engl J Med*. (2015) 373:123–35. doi: 10.1056/NEJMoa1504627
19. Fehrenbacher L, Spira A, Ballinger M, Kowanetz M, Vansteenkiste J, Mazieres J, et al. Atezolizumab versus docetaxel for patients with previously treated non-small-cell lung cancer (POPLAR): a multicentre, openlabel, phase 2 randomised controlled trial. *Lancet*. (2016) 387:1837–46. doi: 10.1016/S0140-6736(16)00587-0
20. Herbst RS, Baas P, Kim DW, Felip E, Pérez-Gracia JL, Han JY, et al. Pembrolizumab versus docetaxel for previously treated, PDL1-positive, advanced non-small-cell lung cancer (KEYNOTE-010): a randomised controlled trial. *Lancet*. (2016) 387:1540–50. doi: 10.1016/S0140-6736(15)01281-7
21. Rittmeyer A, Barlesi F, Waterkamp D, Park K, Ciardiello F, von Pawel J, et al. Atezolizumab versus docetaxel in patients with previously treated non-small-cell lung cancer (OAK): a phase 3, open-label, multicenter randomised controlled trial. *Lancet*. (2017) 389:255–65. doi: 10.1016/S0140-6736(16)32517-X
22. Reck M, Rodríguez-Abreu D, Robinson AG, Hui R, Csoszi T, Fülöp A, et al. Pembrolizumab versus chemotherapy for PD-L1-positive nonsmall-cell lung cancer. *N Engl J Med*. (2016) 375:1823–33. doi: 10.1056/NEJMoa1606774
23. Carbone DP, Reck M, Paz-Ares L, Creelan B, Horn L, Steins M, et al. First-line nivolumab in stage IV or recurrent non-smallcell lung cancer. *N Engl J Med*. (2017) 376:2415–26. doi: 10.1056/NEJMoa1613493
24. Gettinger S, Horn L, Jackman D, Spigel D, Antonia S, Hellmann M, et al. Five-year follow-up of nivolumab in previously treated advanced non-small-cell lung cancer: results from the CA209–003 study. *J Clin Oncol*. (2018) 36:1675–84. doi: 10.1200/JCO.2017.77.0412
25. Huang YH, Zhu C, Kondo Y, Anderson AC, Gandhi A, Russell A, et al. CEACAM1 regulates TIM-3-mediated tolerance and exhaustion. *Nature*. (2015) 517:386–90. doi: 10.1038/nature13848
26. Ribas A. Adaptive immune resistance: how cancer protects from immune attack. *Cancer Discov*. (2015) 5:915–9. doi: 10.1158/2159-8290.CD-15-0563
27. Champiat S, Dercle L, Ammari S, Massard C, Hollebecque A, Postel-Vinay S, et al. Hyperprogressive disease is a new pattern of progression in cancer patients treated by anti-PD-1/PD-L1. *Clin Cancer Res*. (2017) 23:1920–8. doi: 10.1158/1078-0432.CCR-16-1741
28. Ledford H. Promising cancer drugs may speed tumours in some patients. *Nature*. (2017) 544:13–4. doi: 10.1038/nature.2017.21755
29. Hu Y, Fu L. Targeting cancer stem cells: a new therapy to cure cancer patients. *Am J Cancer Res*. (2012) 2:340–56.
30. Morrison R, Schleicher SM, Sun Y, Niermann KJ, Kim S, Spratt DE, et al. Targeting the mechanisms of resistance to chemotherapy and radiotherapy with the cancer stem cell hypothesis. *J Oncol*. (2011) 2011:941876. doi: 10.1155/2011/941876
31. Codony-Servat J, Rosell R. Cancer stem cells and immunoresistance: clinical implications and solutions. *Transl Lung Cancer Res*. (2015) 4:689–703. doi: 10.3978/j.issn.2218-6751.2015.12.11
32. Wang P, Gao Q, Suo Z, Munthe E, Solberg S, Ma L, et al. Identification and characterization of cells with cancer stem cell properties in human primary lung cancer cell lines. *PLoS ONE*. (2013) 8:e57020. doi: 10.1371/journal.pone.0057020
33. Rivera C, Rivera S, Loriot Y, Vozenin MC, Deutsch E. Lung cancer stem cell: new insights on experimental models and preclinical data. *J Oncol*. (2011) 2011:549181. doi: 10.1155/2011/549181
34. Sullivan JP, Minna JD, Shay JW. Evidence for self-renewing lung cancer stem cells and their implications in tumor initiation, progression, and targeted therapy. *Cancer Metastasis Rev*. (2010) 29:61–72. doi: 10.1007/s10555-010-9216-5
35. Lapidot T, Sirard C, Vormoor J, Murdoch B, Hoang T, Caceres-Cortes J, et al. A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. *Nature*. (1994) 367:645–8. doi: 10.1038/367645a0
36. Bhatia M, Wang JC, Kapp U, Bonnet D, Dick JE. Purification of primitive human hematopoietic cells capable of repopulating immune-deficient mice. *Proc Natl Acad Sci USA*. (1997) 94:5320–5. doi: 10.1073/pnas.94.10.5320
37. Dalerba P, Dylla SJ, Park IK, Liu R, Wang X, Cho RW, et al. Phenotypic characterization of human colorectal cancer stem cells. *Proc Natl Acad Sci USA*. (2007) 104:10158–63. doi: 10.1073/pnas.0703478104
38. Al-Hajj M, Wicha MS, Benito-Hernandez A, Morrison SJ, Clarke MF. Prospective identification of tumorigenic breast cancer cells. *Proc Natl Acad Sci USA*. (2003) 100:3983–8. doi: 10.1073/pnas.0530291100
39. Hermann PC, Huber SL, Herrler T, Aicher A, Ellwart JW, Guba M, et al. Distinct populations of cancer stem cells determine tumor growth and metastatic activity in human pancreatic cancer. *Cell Stem Cell*. (2007) 1:313–23. doi: 10.1016/j.stem.2007.06.002
40. Singh SK, Hawkins C, Clarke ID, Squire JA, Bayani J, Hide T, et al. Identification of human brain tumor initiating cells. *Nature*. (2004) 432:396–401. doi: 10.1038/nature03128
41. O'Brien CA, Pollett A, Gallinger S, Dick JE. A human colon cancer cell capable of initiating tumor growth in immunodeficient mice. *Nature*. (2007) 445:106–10. doi: 10.1038/nature05372
42. Ricci-Vitiani L, Lombardi DG, Pilozzi E, Biffoni M, Todaro M, Peschle C, et al. Identification and expansion of human colon-cancer-initiating cells. *Nature*. (2007) 445:111–5. doi: 10.1038/nature05384
43. Prince ME, Sivanandan R, Kaczorowski A, Wolf GT, Kaplan MJ, Dalerba P, et al. Identification of a subpopulation of cells with cancer stem cell properties in head and neck squamous cell carcinoma. *Proc Natl Acad Sci USA*. (2007) 104:973–8. doi: 10.1073/pnas.0610117104
44. Li C, Heidt DG, Dalerba P, Burant CF, Zhang L, Adsay V, et al. Identification of pancreatic cancer stem cells. *Cancer Res*. (2007) 67:1030–7. doi: 10.1158/0008-5472.CAN-06-2030
45. Collins AT, Berry PA, Hyde C, Stower MJ, Maitland NJ. Prospective identification of tumorigenic prostate cancer stem cells. *Cancer Res*. (2005) 65:10946–51. doi: 10.1158/0008-5472.CAN-05-2018
46. Rodriguez-Torres M, Allan AL. Aldehyde dehydrogenase as a marker and functional mediator of metastasis in solid tumors. *Clin Exp Metastasis*. (2016) 33:97–113. doi: 10.1007/s10585-015-9755-9
47. Zakaria N, Yusoff NM, Zakaria Z, Widera D, Yahaya BH. Inhibition of NF-κB signaling reduces the stemness characteristics of lung cancer stem cell. *Front Oncol*. (2018) 8:166. doi: 10.3389/fonc.2018.00166
48. Tomita H, Tanaka K, Tanaka T, Hara A. Aldehyde dehydrogenase 1A1 in stem cells and cancer. *Oncotarget*. (2016) 7: 11018–32. doi: 10.18632/oncotarget.6920
49. Jiang F, Qiu Q, Khanna A, Todd NW, Deepak J, Xing L, et al. Aldehyde dehydrogenase 1 is a tumor stem cell-associated marker in lung cancer. *Mol Cancer Res*. (2009) 7:330–8. doi: 10.1158/1541-7786.MCR-08-0393

50. Alamgeer M, Ganju V, Szczepny A, Russell PA, Prodanovic Z, Kumar B, et al. The prognostic significance of aldehyde dehydrogenase 1A1 (ALDH1A1) and CD133 expression in early stage non-small cell lung cancer. *Thorax.* (2013) 68:1095–104. doi: 10.1136/thoraxjnl-2012-203021
51. Cortes-Dericks L, Froment L, Boesch R, Schmid RA, Karoubi G. Cisplatin-resistant cells in malignant pleural mesothelioma cell lines show $ALDH^{high}CD44^{+}$ phenotype and sphere-forming capacity. *BMC Cancer*. (2014) 14:304. doi: 10.1186/1471-2407-14-304
52. Sullivan JP, Spinola M, Dodge M, Raso MG, Behrens C, Gao B, et al. Aldehyde dehydrogenase activity selects for lung adenocarcinoma stem cells dependent on Notch signalling. *Cancer Res.* (2010) 70:9937–48. doi: 10.1158/0008-5472.CAN-10-0881
53. Aponte PM, Caicedo A. Stemness in cancer: stem cells, cancer stem cells, and their microenvironment. *Stem Cells Int.* (2017) 2017:5619472. doi: 10.1155/2017/5619472
54. Sodja E, Rijavec M, Koren A, Sadikov A, Korošec P, Cufer T. The prognostic value of whole blood *SOX2*, *NANOG* and *OCT4* mRNA expression in advanced small-cell lung cancer. *Radiol Oncol*. (2016) 50:188–96. doi: 10.1515/raon-2015-0027
55. Kahlert C, Bergmann F, Beck J, Welsch T, Mogler C, Herpel E, et al. Low expression of aldehyde dehydrogenase 1A1 (ALDH1A1) is a prognostic marker for poor survival in pancreatic cancer. *BMC Cancer*. (2011) 11:275. doi: 10.1186/1471-2407-11-275
56. Zhou X, Wang G, Sun Y. A reliable parameter to standardize the scoring of stem cell spheres. *PLoS ONE*. (2015) 10:e0127348. doi: 10.1371/journal.pone.0127348
57. Zhang DG, Jiang AG, Lu HY, Zhang LX, Gao XY. Isolation, cultivation and identification of human lung adenocarcinoma stem cells. *Oncol Lett*. (2015) 9:47–54. doi: 10.3892/ol.2014.2639
58. Steibel JP, Poletto R, Coussens PM, Rosa GJ. A powerful and flexible linear mixed model framework for the analysis of relative quantification RT-PCR data. *Genomics.* (2009) 94:146–52. doi: 10.1016/j.ygeno.2009.04.008
59. Lin L, Jou D, Wang Y, Ma H, Liu T, Fuchs J, et al. STAT3 as a potential therapeutic target in ALDH+ and CD44+/CD24+ stem cell-like pancreatic cancer cells. *Int J Oncol*. (2016) 49:2265–74. doi: 10.3892/ijo.2016.3728
60. Liu A, Yu X, Liu S. Pluripotency transcription factors and cancer stem cells: small genes make a big difference. *Chin J Cancer*. (2013) 32:483–7. doi: 10.5732/cjc.012.10282
61. Zhao W, Li Y, Zhang X. Stemness-related markers in cancer. *Cancer Transl Med.* (2017) 3:87–95. doi: 10.4103/ctm.ctm_69_16
62. Carney DN, Gazdar AF, Bunn PA, Guccion JG. Demonstration of the stem cell nature of clonogenic tumor cells from lung cancer patients. *Stem Cells*. (1982) 1:149–64.
63. Salvador J. Diaz-Cano. Tumor heterogeneity: mechanisms and bases for a reliable application of molecular marker design. *Int J Mol Sci.* (2012) 13:1951–2011. doi: 10.3390/ijms13021951
64. Dobbin ZC, Landen CN. Isolation and characterization of potential cancer stem cells from solid human tumors – potential applications. *Curr Protoc Pharmacol*. (2013) 63:Unit−14.28. doi: 10.1002/0471141755.ph1428s63
65. Greve B, Kelsch R, Spaniol K, Eich HT, Götte M. Flow cytometry in cancer stem cell analysis and separation. *Cytometry Part A*. (2012) 81A: 284–93. doi: 10.1002/cyto.a.22022
66. Bielecka ZF, Maliszewska-Olejniczak K, Safir IJ, Szczylik C, Czarnecka AM. Three-dimensional cell culture model utilization in cancer stem cell research. *Biol Rev*. (2017) 92:1505–20. doi: 10.1111/brv.12293
67. Broakley KWR, Hunn MK, Farrand KJ, Price KM, Grasso C, Miller RJ, et al. Side population is not necessary or sufficient for a cancer stem cell phenotype in glioblastoma multiforme. *Stem Cells.* (2011) 29:452–61. doi: 10.1002/stem.582
68. Wang L, Zuo X, Xie K, Wei D. The role of CD44 and cancer stem cells. *Methods Mol Biol*. (2018) 1692:31–42 doi: 10.1007/978-1-4939-7401-6_3
69. Park E, Park SY, Sun PL, Jin Y, Kim JE, Jheon S, et al. Prognostic significance of stem cell-related marker expression and its correlation with histologic subtypes in lung adenocarcinoma. *Oncotarget*. (2016) 7:42502–12. doi: 10.18632/oncotarget.9894
70. Karachaliou N, Rosell R, Viteri S. The role of SOX2 in small cell lung cancer, lung adenocarcinoma and squamous cell carcinoma of the lung. *Transl Lung Cancer Res.* (2013) 2:172–9. doi: 10.3978/j.issn.2218-6751.2013.01.01

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

SYSTEMATIC REVIEW
published: 20 December 2019
doi: 10.3389/fonc.2019.01441

The Prognostic Role of MET Protein Expression Among Surgically Resected Non-small Cell Lung Cancer Patients: A Meta-Analysis

Guangzhi Ma[1,2†], *Yunfu Deng*[1†], *Wenjie Chen*[1], *Zhenkun Liu*[1], *Cheng Ai*[1], *Xuebing Li*[3*] and *Qinghua Zhou*[1,3*]

[1] Lung Cancer Center, West China Hospital, Sichuan University, Chengdu, China, [2] Department of Thoracic Surgery, West China Hospital, Sichuan University, Chengdu, China, [3] Tianjin Key Laboratory of Lung Cancer Metastasis and Tumor Microenvironment, Tianjin Lung Cancer Institute, Tianjin Medical University General Hospital, Tianjin, China

OPEN ACCESS

Edited by:
Etienne Giroux Leprieur,
Hôpital Ambroise-Paré, France

Reviewed by:
Jai Narendra Patel,
Levine Cancer Institute, United States
Timothy F. Burns,
University of Pittsburgh, United States

***Correspondence:**
Xuebing Li
xbli@tmu.edu.cn
Qinghua Zhou
zhouqh135@163.com

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Thoracic Oncology, a section of the journal Frontiers in Oncology

Received: 05 October 2019
Accepted: 03 December 2019
Published: 20 December 2019

Citation:
Ma G, Deng Y, Chen W, Liu Z, Ai C, Li X and Zhou Q (2019) The Prognostic Role of MET Protein Expression Among Surgically Resected Non-small Cell Lung Cancer Patients: A Meta-Analysis. Front. Oncol. 9:1441. doi: 10.3389/fonc.2019.01441

Objectives: MET protein expression has been reported to be in relevance with the survival of NSCLC patients in various studies, yet the results were inconsistent. The purpose of our study set out to determine the prognostic role of both c-MET and p-MET expression among NSCLC that underwent surgical resection.

Methods: Data were obtained from retrospective cohort studies by searching on PubMed, Cochrane Library, EMBASE and Web of Science, and a meta-analysis was performed to assess the prognostic role of MET expression among NSCLC.

Results: Totally 18 literatures including 5,572 surgically resected NSCLC cases staged I-IV were included for data synthesis. The positive rate of c-MET and p-MET was 1,753/4,315 and 135/1,257. The pooled hazard ratios (HRs) regarding c-MET and p-MET expression for overall survival (OS) was 1.623 (95% CI: 1.176–2.240, $p = 0.003$) and 1.710 (95% CI: 0.823–3.533, $p = 0.15$), respectively. Subgroup analysis results on Asian (HR = 2.115, $p < 0.001$), adenocarcinoma (HR = 2.220, $p < 0.001$) and rabbit polyclonal antibodies (HR = 2.107, $p < 0.001$) etc. were also indicative.

Conclusion: C-MET over-expression among NSCLC patients that underwent surgical resection is a prognostic factor that indicated adverse survival on OS. Whereas, p-met didn't appear to have an impact on the prognosis of NSCLC. The studies are need and the topic could be re-valued by then.

Keywords: MET protein, prognostic role, resected, non-small cell cancer, meta-analysis

INTRODUCTION

Lung cancer remains the leading cause of cancer incidence and mortality worldwide, accounted for ∼1.8 million deaths in 2018 (1). Among which statistically 85–90% of lung cancer cases were non-small cell lung cancer, or NSCLC based on pathologic classification (2). With the development of target-therapy and immunotherapy, alternatives to deal with NSCLC posterior to *en bloc* resection is comprehensive (3). Yet besides the efforts to improve therapeutic methods and diagnostic accuracy, the outcomes of NSCLC patients remains unsatisfactory (4, 5).

MET protein, also known as hepatocyte growth factor receptor (HGFR), has been characterized as a high affinity transmembrane receptor tyrosine kinase (RTK) which is encoded by its homologous oncogene *MET* (6, 7). Being firstly recognized in osteosarcoma derived cell-lines, MET was subsequently identified to have over-expressed in various malignances including NSCLC (8, 9). When c-MET binds to its homologous ligand HGF, the intracellular tyrosine residues of the RTK became activated via auto-phosphorylation (p-MET) (10). P-MET accordingly triggers its downstream pathways such as PI3k-Akt, Ras-MAPK, and STAT3, which physiologically promotes tissue growth, vascularization, and healing (11, 12). Whereas, the aberrant expression of MET would result in tumorigenesis and development of various malignancies, including NSCLC (13, 14). The mechanisms that led abnormal HGF/c-MET signaling were either *c-MET* amplification, mutation or MET/HGF overexpression, and among which MET over-expression most frequently occurred (15, 16). Prior studies have noted alterations regarding HGF/c-MET signaling played a key role among NSCLC patient that acquired resistance to first generation EGFR-TKIs due to its underlying interactions with EGFR pathways (17, 18). In addition, targeting *MET* as well as MET upregulation via either TKIs or MET-antibodies has already become a novel strategy to challenge NSCLC patients with metastatic disease (19–22). Hence, understanding the impact of c-MET/p-MET expression on NSCLC survival should be highlighted. As primary c-MET/p-MET expression status of NSCLC patients was majorly from resected-specimen tumors via immunohistochemistry (IHC), patients that received surgical therapy was our main concern.

To date literatures has emerged with inconsistent conclusions on the prognostic role of MET among NSCLC. C-MET expression was thought to be a favorable biomarker in various studies (23–25), yet others suggested the opposite (26–28). In addition to some studies, neither c-MET nor p-MET expression was related with NSCLC survival (29, 30). Thus, due to the contradictory results from previous studies, we herein set out to conduct a systematic review as well as meta-analysis by summarizing current existing data to examine the survival implications of MET over-expression among lung cancer patients that underwent surgical resection.

MATERIALS AND METHODS

Literature Search

Two reviewers (GM and YD), respectively, conducted electronic search on PubMed, Cochrane Library, EMBASE, and Web of Science for relevant studies up till July 15th, 2019, with the beginning date unlimited. The search terms were as followed: "MET" or "Mesenchymal Epithelial Transition factor" or "Hepatocyte growth factor receptor" and "Non-small cell lung cancer" or "NSCLC" or "Pulmonary carcinoma" or "lung cancer" and "Prognosis" or "Outcomes" or "Survival."

Inclusion Criteria

Eligible studies was required to be in compliance with the following criteria: (1) NSCLC studies, all included participants should be NSCLC patients that underwent surgical resection; (2) MET expression was examined of each resected specimen, with the correlation between MET expression and NSCLC survival been reported; (3) Hazard Ratio (HR) was clearly displayed and feasible for HR synthesis, according to methods described by Parmar et al. (31), Williamson et al. (32), and Tierney et al. (33); (4) Study designs include: randomized controlled trial (RCT) and cohort study.

Exclusion Criteria

Articles were omitted from further consideration if: (1). Systematic review or review; (2) Preclinical studies, such as laboratorial or *in vitro* studies; (3) Case reports; (4). Studies of which survival data (including survival curves yet without HRs reported) unavailable for further calculations.

Data Extraction

Basic information of each eligible study was extracted as followed: name of first author, publication year, country, demographic characteristics (number of patients, gender, and median age), smoking status, pathology, and tumor stage, antibody applied for MET immunohistochemical (IHC) staining, cut-off value of MET over-expression and reported HRs (representing prognosis) for meta-analysis.

The primary data eligible for calculation and results-pooling was hazard ratios (HRs) reported from either multivariate or univariate Cox hazard regression analysis for overall survival (OS). Literatures of eligibility was filtered by two authors (GM and YD) individually, with any discordance being revised and re-assessed.

Quality Evaluation

The Newcastle–Ottawa Scale (NOS) criterion was adopted for quality assessment of included studies (34). The criteria covered three aspects of each study: (1) selection of subject: 0–4; (2) subject comparability: 0–2; and (3) survival: 0–3. The scope regarding the final score ranged between 0 and 9, literature with six or more were reckoned feasible for data incorporation and any scored no <7 were considered of good quality. Two reviewers independently carried out quality evaluation of each study.

Statistical Analysis

Data calculation and meta-analyses were performed via STATA (version 12.0, STATA Corporation, Texas, USA). LogHRs reported in the literature were prior used for HR pooling, otherwise HRs with 95% confidential intervals (CIs) were considered for data syntheses. Multivariate analyses data were prior adopted if multivariate and univariate survival analyses were both conducted. Adjusted HRs was used when unadjusted/adjusted HRs both existed. Chi-square based Q-test and I^2 statistic test were performed to value heterogeneity regarding the pooled HRs (35). The Mantel-Haenszel method or fixed-effect models (36) were adopted when study heterogeneity wasn't statistically considered significant ($I^2 < 50\%$ or $P > 0.10$) whereas random-effect models were applied for calculation in order to minimize potential influence of heterogeneity on pooled results. Apart from random-effect model, sensitivity analysis

by leave one out procedures was also processed uncovering the potential source regarding heterogeneity of pooled results (37). Publication bias were conducted in accordance with Begg's methods (38). Publication bias was reckoned significant when *P*-value was <0.05.

RESULTS

Study Selection

Our initial literature search retrieved 1,151 studies (after duplicates removal) in total. Abstracts of each identified publication was discreetly read and screened. Studies were removed due to the reasons as followed: Reviews or systematic review ($n = 124$), case reports ($n = 158$), irrelevant topic or fundamental observations ($n = 807$). Totally 62 potential studies of eligibly were obtained and scrutinized. Then 45 of which were omitted owing to the following reasons: 34 studies focused on irrelevant topics such as MET gene expression and alterations, 11 remaining studies whose data were either survival curve or illegible of HR estimation. Two studies conducted by Sun et.al included overlapped patients (28, 39). To limit potential risk of bias, we omitted the publication with lesser participants. Hence, altogether 17 studies eventually met our criteria of inclusion and were capable of data extraction as well as meta-analyses.

Summarized process of literature selection was displayed in the flow chart of **Figure 1**.

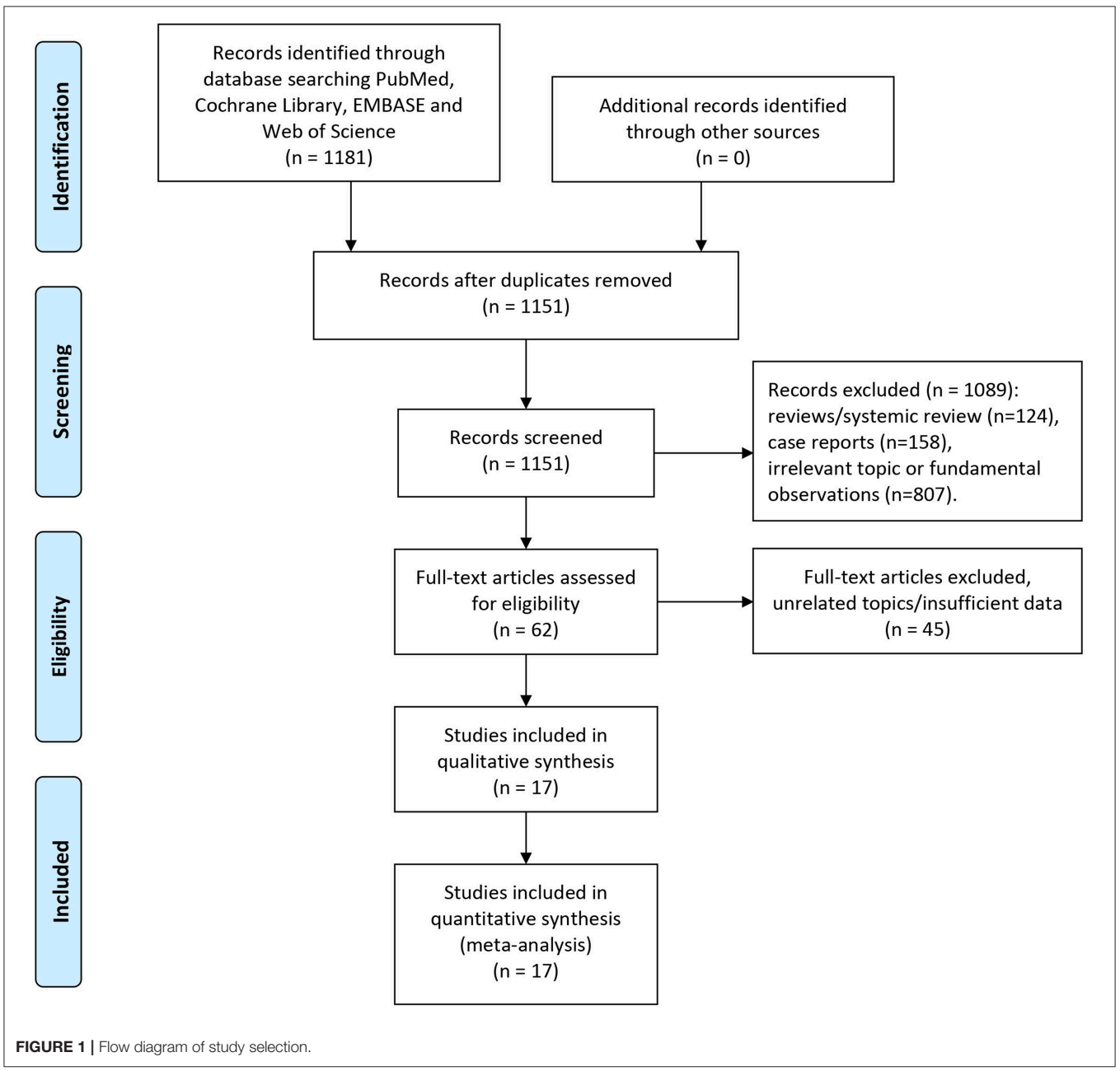


FIGURE 1 | Flow diagram of study selection.

Baseline Characteristics

In all, our topic was demonstrated in 17 studies. Among which Asian studies dominated in quantity, including six from Japan (26, 29, 30, 40–42), five from China (23, 39, 43–45), and two from Korea (27, 46). Caucasian patients that were either from Sweden (25), Netherlands (47), Poland (48), or Australia (24) comprised the rest population of included studies. Totally 4,315 NSCLC patients staged between I-IV that underwent surgical resection were assessed of c-MET expression, while 1,257 participants had p-MET evaluation. Immunohistochemical staining (IHC) was performed on each corresponding NSCLC tissue to value c-MET/p-MET expression, and the rabbit derived antibodies accounted for the majority of antibodies to against MET protein. All studies mentioned HRs that were feasible for data-pooling. MET over-expression were determined in accordance with certain measurements that had varied threshold values such as H-score or H intensity. All studies scored no <6 with reference to the NOS quality criterion, hence applicable for our meta-analysis.

Baseline information was listed on **Table 1**.

Results From Meta-Analyses

The primary end-point of surveillance among included studies was OS. The correlation between MET and outcomes was determined in accordance with combined HRs and related intervals. As a result, the prognostic role of total MET protein or c-MET expression was analyzed in 15 studies of which the combined HR was 1.623 (95% CI: 1.176–2.240, $p = 0.003$), indicating an adverse impact of c-MET expression on NSCLC prognosis. Heterogeneity was significant ($I^2 = 85.9$) thus random-effect model was adopted (**Figure 2A**). With regard to activated c-MET or p-MET, however, apart from potential

TABLE 1 | Baseline characteristics of the included publications.

References	Country	Median age	N(F/M)	Smoking (S/NS)	Histology	Stage	MET type	Antibody	Cut-off value	MET high	MET low	HR estimation
Tsakonas et al. (25)	Sweden	66.5	653 (316/337)	589/64	NSCLC	IA–IIIB	c-MET	PharmDx	H-score ≥ 20	336	117	Multi
Zhang et al. (45)	China	60.4	86 (44/42)	29/57	ADCC	I–IV	c-MET	RM (SP44)	Staining score ≥ 2+ (50%)	54	32	Multi
Kim et al. (46)	Korea	NR	311 (140/171)	109/202	ADCC	IB–IIIA	c-MET	RM (SP44)	Staining score ≥ 2+ (50%)	141	170	Multi
Tran et al. (24)	Australia	67 (–)/ 69 (+)	271 (98/173)	211/9	NSCLC	I–III	c-MET	RM (SP44)	Staining score ≥ 2+ (50%)	248	23	Multi
Tong et al. (44)	China (HK)	66	687 (223/464)	395/223	NSCLC	I–IV	c-MET	RM (SP44)	Staining score ≥ 2+ (50%)	230	457	Uni
Wang et al. (43)	China	57	117 (33/84)	43/74	NSCLC	I–IV	c-MET	R*	H-score ≥ 1.9	36	81	Multi
Huang et al. (23)	China	62	102 (29/73)	47/55	NSCLC	I–IV	c-MET	RM	H-score ≥ 60	52	50	Multi
Sun et al. (39)	China	56.2	183 (42/141)	117/66	ADCC/SCC	I–IV	c-MET	R*	Staining score > 3	123	60	Multi
Tsuta et al. (30)	Japan	61.7	906 (332/574)	416/490	NSCLC	I–IV	c-MET	RM	stained cells ≥ 10%/MA	196	687	Uni
							p-MET	RM	stained cells ≥ 10%/MA	51	829	Uni
Tachibana et al. (29)	Japan	64	106 (55/51)	55/51	ADCC	I–III	c-MET	RP	H intensity ≥ 2+	30	76	Uni
Park et al. (27)	Korea	62	380 (72/308)	279/101	ADCC/SCC	I–IV	c-MET	RP	H-score ≥ 4	52	328	Multi
Dziadziuszko et al. (48)	Poland	63	174 (39/135)	165/9	NSCLC	I–IV	c-MET	RM (SP44)	H-score > 60	83	91	Uni
Onitsuka et al. (26)	Japan	68.5	183 (81/102)	NR	ADCC	I–III	p-MET	M*	IHC Allred score ≥ 3	12	171	Multi
Ruiz et al. (47)	Netherlands	NR	168	NR	NSCLC	I–III	p-MET	NR	H-score > 5	72	96	Multi
Masuya et al. (42)	Japan	NR	88	NR	NSCLC	I–III	c-MET	RP	H intensity > grade 1	36	52	Uni
Tokunou et al. (41)	Japan	59	131 (58/73)	NR	ADCC	I–IV	c-MET	RP	Stained bundles ≥ 1/MA	69	62	Multi
Takanami et al. (40)	Japan	61	120 (51/69)	NR	ADCC	I–IV	c-MET	RP (C-28)	Stained cells ≥ 25%/MA	67	53	Multi

N, Number of patients; F, Female; M, Male; S, Smoker; NS, Non-smoker; NSCLC, Non-small cell lung cancer; ADCC, Adenocarcinoma; SCC, Squamous cell carcinoma; RM, Rabbit monoclonal; RP, Rabbit polyclonal; R, Rabbit; M*, Monoclonal; NR, Not reported; MM, Mouse monoclonal; MA, Microscopic area; Multi, Multivariate analysis; Uni, Univariate analysis.*

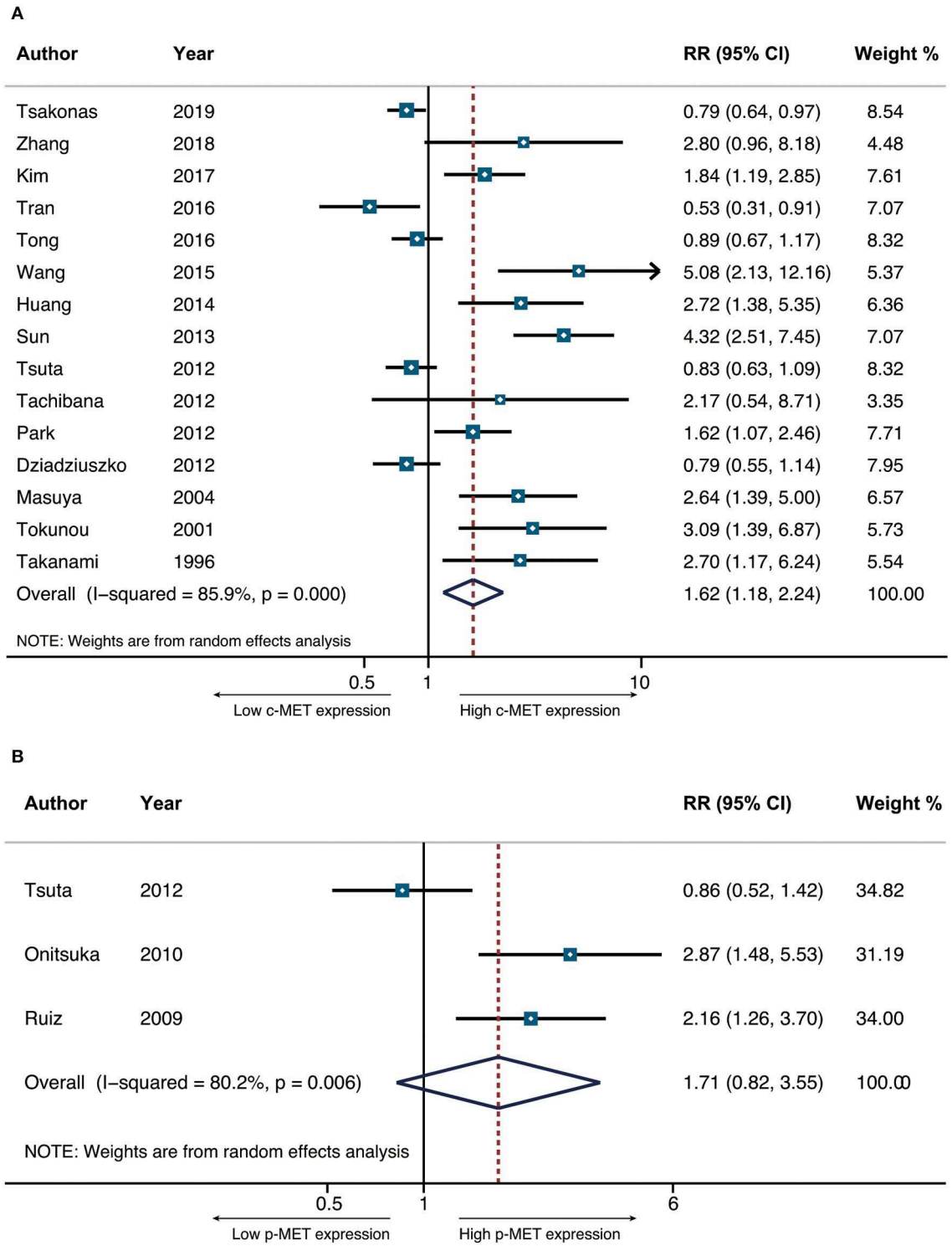


FIGURE 2 | The pooled hazard ratio (HR) for OS in surgically resected NSCLC patients that had positive c-MET expression **(A)** and p-MET expression **(B)**.

heterogeneity ($I^2 = 80.2$, $p = 0.003$) when combining the three related studies, the pooled result for OS (HR = 1.710, 95% CI: 0.823–3.533, $p = 0.15$) was neither indicative (**Figure 2B**).

Subgroup Analyses

Subgroups were performed in terms of demographic distributions and characteristics from all eligible studies.

Subgroups were stratified by (1) Regions (Asian/Non-Asian/Japanese/Chinese); (2) Histology (Adenocarcinoma); (3) Antibodies for IHC (Rabbit monoclonal/Rabbit polyclonal); (4) MET evaluation (H-score); and (5) Derived data (via multivariate analysis or univariate analysis).

Asian/Non-Asian/Japanese/Chinese

Totally 12 studies were conducted in Asia, and the pooled HR for OS was 2.115 (95% CI 1.440–3.108, $P < 0.001$, $I^2 = 83.5\%$). The pooled HRs via random-effect models from five Japanese studies and five Chinese studies was 1.985 (95% CI 0.970–4.058, $P = 0.06$) and 2.627 (95% CI 1.123–6.146, $p = 0.026$, $I^2 = 90.1\%$), respectively. With respect to non-Asian patients, the combined HR for OS from four studies was 0.901 (95% CI 0.586–1.387, $p = 0.637$), with random-effect model due to its significant heterogeneity ($p = 0.002$, $I^2 = 80.0\%$).

Adenocarcinoma

The prospect of our present study was to examine the prognostic role of MET expression on multiple NSCLC types. Yet only adenocarcinoma was applicable. Altogether five studies focused on pulmonary adenocarcinoma, and the synthesized HR of OS was 2.220 (95% CI 1.607–3.066, $P < 0.001$). Fixed-effect model was applied to perform the analysis since heterogeneity was not significant ($P = 0.781$, $I^2 = 0\%$).

IHC Antibodies/H-Score

IHC was performed in all studies, and antibodies for MET staining varied between studies. As to seven studies that applied rabbit monoclonal antibodies, the combined HR for OS was 1.107 (95% CI 0.777–1.579, $P = 0.573$, $I^2 = 78.9\%$). Among which five studies adopted SP44 (Ventana Medical Systems, AZ, USA) antibody, and the pooled HR for survival was 1.031 (95% CI 0.668–1.590, $P = 0.001$, I^2=78.1%). In addition, four studies via SP44 examined MET expression by same cut-off value with reference to methods by Spigel et al. (49), and the pooled HR was 1.031 (95% CI 0.668–1.590, $p = 0.892$). For the survival analysis of five studies that applied rabbit polyclonal antibodies, the pooled HR was 2.107 (95% CI 1.573–2.823, $P < 0.001$). Heterogeneity was not statistically significant ($p = 0.521$, $I^2 = 0\%$) thus fixed-effect model was preferred.

Primary Data

Ten studies addressed the prognostic role of MET over-expression among NSCLC by multivariate analysis. The pooled HR on OS was 2.004 (95% CI 1.229–3.268, $P = 0.005$). The remaining five studies were performed by univariate analysis, of which the pooled HR was 1.051 (95% CI 0.745–1.484, $p = 0.776$). Heterogeneity was significant among either results ($I^2 = 88.4$ and 69.7, respectively). Therefore, random-effect model was adopted for both analyses.

All summarized data was presented on **Table 2** and shown in **Figure 3**.

Sensitivity Analysis and Publication Bias

As shown in **Figure 4A**, the combined results representing the pooled HRs didn't prominently change when each study was sequentially removed, indicating the above synthesized results credible and robust. In addition, publication bias of our systematic review was neither found to exist, in accordance with Begg's plots in **Figure 4B**.

DISCUSSION

Our current study aimed to examine the prognostic role of c-MET/p-MET positivity among NSCLC patients that underwent surgical resection. With incorporated data, a meta-analysis was performed. As a result, although p-MET was not found to be associated with NSCLC survival, c-MET appears to be a prognostic factor that led to shorter OS. In view of Asian population, subgroup results indicated that c-MET

TABLE 2 | Meta-analyses of MET protein over-expression and survival of surgically resected NSCLC.

	N of studies	Model	HR (95% CI)	Log-rank p	Heterogeneity (p, I^2)	Conclusion
C-MET OS	15	Random	1.623 (1.176–2.240)	0.003	<0.001, 85.9%	Positive
P-MET OS	3	Random	1.710 (0.823–3.533)	0.15	0.006, 80.2%	Negative
Asian OS	12	Random	2.115 (1.440–3.108)	<0.001	<0.001, 83.5%	Positive
Non-Asian OS	4	Random	0.901 (0.586–1.387)	0.637	0.002, 80.0%	Negative
Japanese OS	5	Random	1.985 (0.970–4.058)	0.06	<0.001, 82.1%	Negative
Chinese OS	5	Random	2.627 (1.123–6.146)	0.026	<0.001, 90.1%	Positive
ADCC OS	5	Fixed	2.220 (1.607–3.066)	<0.001	0.781, 0%	Positive
RM OS	7	Random	1.107 (0.777–1.579)	0.573	<0.001, 78.9%	Negative
RM (SP44) OS	5	Random	1.031 (0.668–1.590)	0.892	0.001, 78.1%	Negative
H-score	4	Random	1.014 (0.822–1.251)	0.893	0.001, 0.893	Negative
RP OS	5	Fixed	2.107 (1.573–2.823)	<0.001	0.521, 0%	Positive
MVA OS	10	Random	2.004 (1.229–3.268)	0.005	<0.001, 88.4%	Positive
UVA OS	5	Random	1.051 (0.745–1.484)	0.776	0.010, 69.7%	Negative

N, Number; HR, Hazard Ratio; CI, Confidence Interval; OS, Overall Survival; ADCC, Adenocarcinoma; RM, Rabbit Monoclonal; RP, Rabbit Polyclonal; MM, Mouse Monoclonal.

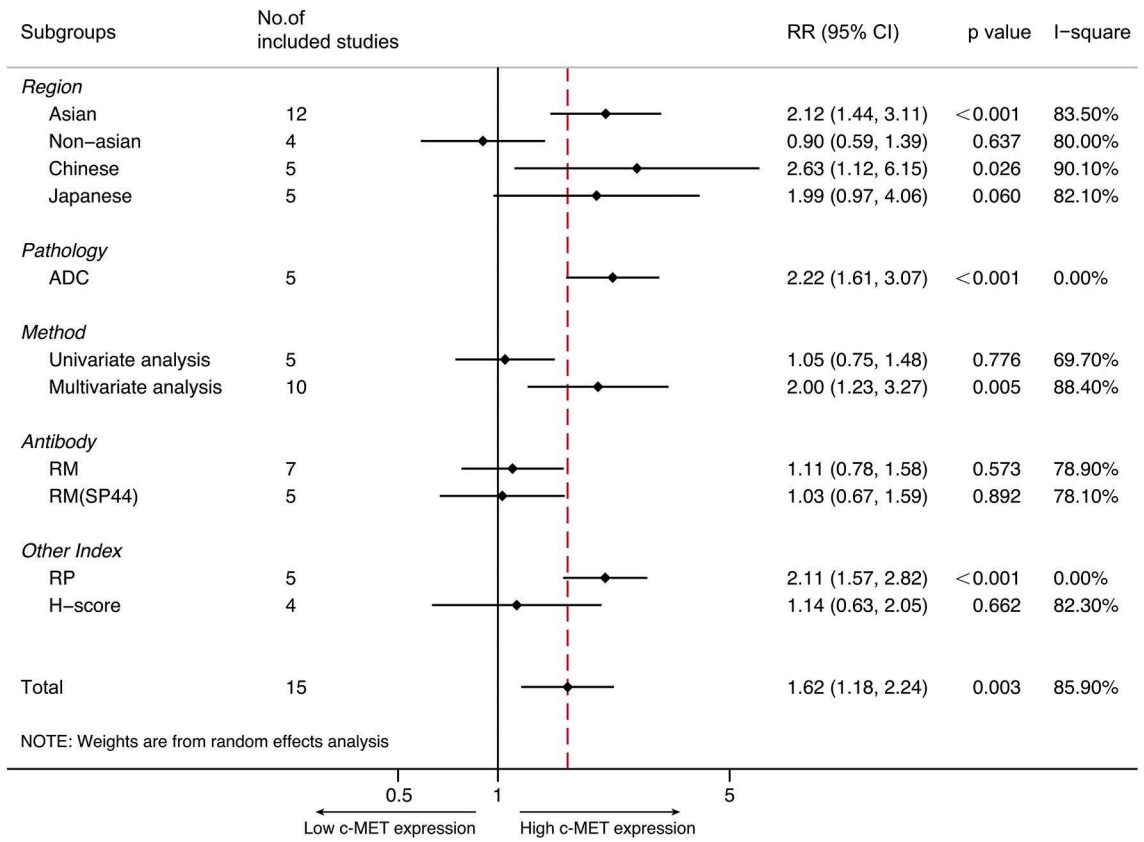


FIGURE 3 | Forest plots representing the pooled results of subgroup analyses.

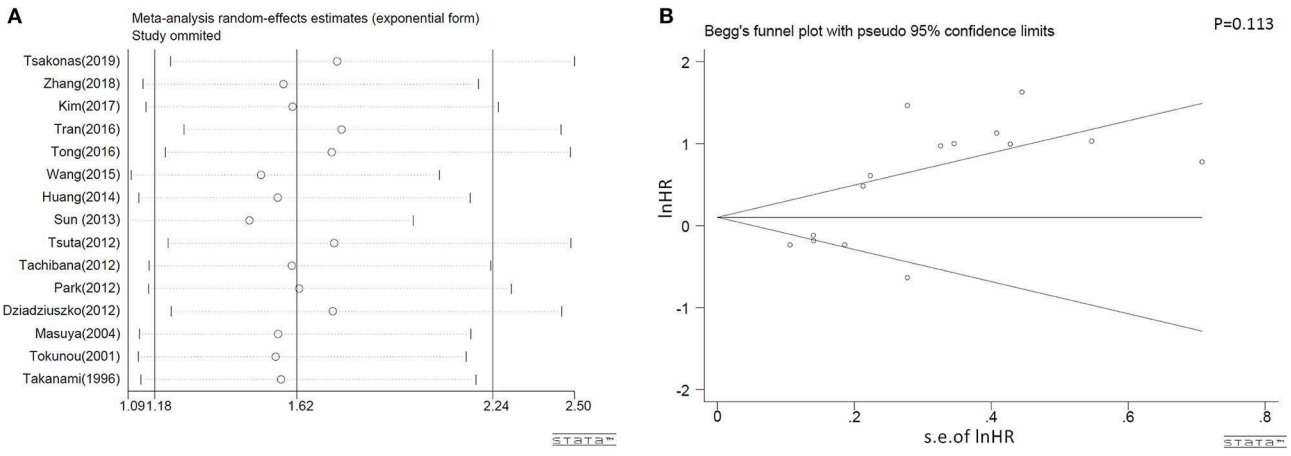


FIGURE 4 | Sensitivity analyses results on omission of each individual study **(A)** and the Begg's publications plots **(B)** of eligible studies that assessed c-MET positivity and NSCLC survival on OS.

was an inferior prognostic marker, and such is the same among Chinese people. Conversely, c-MET wasn't related with outcomes regarding Japanese participants. From a fixed-model, c-MET overexpression was significantly involved with inferior OS of patients with resected pulmonary adenocarcinoma. C-MET was in correlation with poor survival when rabbit polyclonal agents was applied for IHC, whereas neither rabbit monoclonal antibodies nor H-score were indicative when discussing its impact on survival of NSCLC whose c-MET was positive. Pooled result from univariate data suggested

c-MET was not a marker of prognosis. On the contrary, synthesized data via multivariate analyses reflected a decisive conclusion which c-MET was an unfavorable prognostic marker of NSCLC.

From several aspects the adverse role of c-MET expression could be explained. Previous studies have noted that over-expression of c-MET was positively associated with vascular and lymphatic invasion, which led to higher risk of cancer relapse as well as more advanced stage among NSCLC patients (43, 50, 51). From therapy's experience, c-MET positivity was closely related with radio-resistance and chemo-resistance, hence correlated with unfavorable outcomes (52, 53). In terms of pathogenesis, HGF could facilitate tumor metastasis through MET/HGF pathways by inducing epithelial-mesenchymal transition (EMT) process of NSCLC (54, 55). And metastasis is considered as the major cause of lung cancer related death (56). In addition, c-MET over-expression was related with the prognosis of patients that harbored varied EGFR status as MET and EGFR shared signal molecules in downstream pathways (23, 46). Thus, MET over-expression could affect efficacy of patients that received EGFR-TKIs as a result (57). Interestingly, p-MET expression which represents activation level of MET didn't have an impact on survival of NSCLC in our study. As clinical research regarding p-MET is lacking (58), the prognostic role of p-MET remains to be further explored.

It remains controversial to determine how MET over-expressed. Alterations on transcription level of *MET* gene, which includes *MET* amplification and *MET exon 14* skipping mutation could be the potential mechanisms (21). Additionally, high gene copy number (GCN) of *MET* was also found be an adverse survival indicator in several studies (59–61). Nevertheless, MET positivity was notably higher, with a prevalence ranged up-till 70% among NSCLC, compared with *MET* mutations (around 4%) (58). With reference to previous studies, MET overexpression is positively correlated with NSCLCs that harbored *MET exon 14* skipping mutation and amplification both (20, 62). Indicating MET positivity could be adopted to screen NSCLC patients for further genetic profiling, as *MET* alterations has been recognized as a biomarker to receive Crizotinib treatment, and a potential trigger to cause first generation EGFR-TKIs resistance (19, 20, 63). In addition, MET over-expression was reported to be a favorable marker among NSCLC patients that received anti-MET therapy as an alternative. C-MET-positive patients had improved survival when given either anti-MET monoclonal antibodies (Onartuzumab) or MET inhibitors in combination with EGFR-TKIs, with reference to ongoing clinical trials (49, 64). Hence, understanding the nature of MET expression as well as establishing a standardized criteria regarding its evaluation, should be highlighted.

Previously two meta-analysis was published that assessed the impact of MET expression on survival among NSCLC (60, 65). Guo et al. integrated 13 studies and some of the results such as Asian/Non-Asian sub-group analyses were in concordance with ours. Yet a major concern of this systematic review was not making a distinction between c-MET and p-MET, as relevant literatures were combined as a whole. The other study by Pyo et al. also indicated that c-MET was an adverse prognostic factor, which is in agreement with ours, but merely 11 literatures were capable of data pooling. In addition, both systematic reviews adopted survival data via estimation from publications whose HR and CI were not directly provided. To avoid potential risk of bias, those literatures were excluded from our study. As numerous retrospective studies emerged in recent years, our systematic review with 17 publications incorporating 5,572 NSCLC patients has the largest data as well as information summarized in scale. To date it is the first systematic review that highlighted the impact of p-MET on NSCLC survival, as well as the first systematic review that analyzed the correlation between c-MET expression and NSCLC prognosis in many aspects such as IHC cut-off value and antibody selection.

Due to practical constraints, our meta-analysis has several limitations. Firstly, our several results had significant heterogeneity. Efforts such as sensitive analysis and subgroup analyses were performed on the basis of several aspects but a distinct source was still lacking. Hence we speculate that the existing heterogeneity could be attributed to the inconsistency of baseline characteristics from included literatures such as tumor stage, smoking status, post-operative therapies and IHC methodology involving varied cut-off values and antibody adoption. Tumor stage was highlighted in the protocol of the present study for its relationship with MET positivity, yet we failed to analyze the prognostic role of MET in each individual stage due to lack of original data. Moreover, explanations to the positive results derived from Asian population remains obscure. With respect to IHC, although several recent publications performed their IHC analysis with reference to an anti-MET clinical trial (49), a standardized criteria for IHC to determine MET positivity is lacking. Besides, we are unable to interpret the loss of survival when rabbit polyclonal antibodies were applied for MET IHC. Secondly, the amount of eligible literature in our study is relatively small, especially in the analysis of p-MET. Hence the current study could be re-conducted when more evidence have emerged. In addition to above, all data search in our study were carried out in English databases, hence some eligible publications written in other languages could have been neglected. Despite of limitations above, with discreetly pooled statistics and detailed protocols, bias was restrained to the minimum, and the results of the current study is guaranteed reliable.

CONCLUSIONS

In conclusion, c-MET over-expression among resected NSCLC patients is a prognostic factor that indicated adverse survival on OS. Yet p-met didn't appear to have an impact on the prognosis of patients with NSCLC. The existing IHC criteria to define MET positivity is inconsistent, which might be a factor to cause heterogeneity. More studies should be conducted to examine the topic, especially studies that focuses on p-MET expression among NSCLC patients. The prognostic role of c-MET/p-MET both among NSCLC could be re-evaluated when added evidence have emerged by then.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

AUTHOR CONTRIBUTIONS

GM and QZ conceived the study. GM and XL designed the study. GM and YD searched the literature and collected the data and performed the analyses. GM, YD, and WC drafted the manuscript. ZL and GM prepared **Figures 1**–**4**. CA and YD prepared **Tables 1**, **2**. All authors reviewed and revised the manuscript.

FUNDING

This study was partly supported by the grants from the National Natural Science Foundation of China (No. 81572288 to QZ; No. 81302002 to XL), the Key Project of International Cooperation of Science and Technology Innovation between Governments, the National Key Research and Development Plan of China (No. 2016YEE0103400 to QZ), and the Tianjin Natural Science Foundation (Nos. 18JCYBJC92100, 14JCQNJC12300 to XL).

REFERENCES

1. 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. doi: 10.3322/caac.21492
2. Ramalingam SS, Owonikoko TK, Khuri FR. Lung cancer: new biological insights and recent therapeutic advances. *CA Cancer J Clin.* (2011) 61:91–112. doi: 10.3322/caac.20102
3. Ettinger DS, Aisner DL, Wood DE, Akerley W, Bauman J, Chang JY, et al. NCCN guidelines insights: non-small cell lung cancer, version 5.2018. *J Natl Compr Canc Netw*. (2018) 16:807–21. doi: 10.6004/jnccn.2018.0062
4. Aggarwal A, Lewison G, Idir S, Peters M, Aldige C, Boerckel W, et al. The state of lung cancer research: a global analysis. *J Thorac Oncol.* (2016) 11:1040–50. doi: 10.1016/j.jtho.2016.03.010
5. Gettinger S, Horn L, Jackman D, Spigel D, Antonia S, Hellmann M, et al. Five-year follow-up of nivolumab in previously treated advanced non-small-cell lung cancer: results from the CA209-003 study. *J Clin Oncol.* (2018) 36:1675–84. doi: 10.1200/JCO.2017.77.0412.
6. Cooper CS, Park M, Blair DG, Tainsky MA, Huebner K, Croce CM, et al. Molecular cloning of a new transforming gene from a chemically transformed human cell line. *Nature.* (1984) 311:29–33. doi: 10.1038/311029a0.
7. Gherardi E, Birchmeier W, Birchmeier C, Vande Woude G. Targeting MET in cancer: rationale and progress. *Nat Rev Cancer.* (2012) 12:89–103. doi: 10.1038/nrc3205
8. Cipriani NA, Abidoye OO, Vokes E, Salgia R. MET as a target for treatment of chest tumors. *Lung Cancer.* (2009) 63:169–79. doi: 10.1016/j.lungcan.2008.06.011
9. Eder JP, Vande Woude GF, Boerner SA, LoRusso PM. Novel therapeutic inhibitors of the c-Met signaling pathway in cancer. *Clin Cancer Res.* (2009) 15:2207–14. doi: 10.1158/1078-0432.CCR-08-1306
10. Ma PC, Maulik G, Christensen J, Salgia R. c-Met: structure, functions and potential for therapeutic inhibition. *Cancer Metastasis Rev.* (2003) 22:309–25. doi: 10.1023/a:1023768811842
11. Chmielowiec J, Borowiak M, Morkel M, Stradal T, Munz B, Werner S, et al. c-Met is essential for wound healing in the skin. *J Cell Biol.* (2007) 177:151–62. doi: 10.1083/jcb.200701086
12. Trovato M, Torre ML, Ragonese M, Simone A, Scarfi R, Barresi V, et al. HGF/c-met system targeting PI3K/AKT and STAT3/phosphorylated-STAT3 pathways in pituitary adenomas: an immunohistochemical characterization in view of targeted therapies. *Endocrine.* (2013) 44:735–43. doi: 10.1007/s12020-013-9950-x
13. Sierra JR, Tsao MS. c-MET as a potential therapeutic target and biomarker in cancer. *Ther Adv Med Oncol.* (2011) 3:S21–35. doi: 10.1177/1758834011422557
14. Goyal L, Muzumdar MD, Zhu AX. Targeting the HGF/c-MET pathway in hepatocellular carcinoma. *Clin Cancer Res.* (2013) 19:2310–8. doi: 10.1158/1078-0432.CCR-12-2791
15. Peters S, Adjei AA. MET: a promising anticancer therapeutic target. *Nat Rev Clin Oncol.* (2012) 9:314–26. doi: 10.1038/nrclinonc.2012.71
16. Yeung SF, Tong JHM, Law PPW, Chung LY, Lung RWM, Tong CYK, et al. Profiling of oncogenic driver events in lung adenocarcinoma revealed MET mutation as independent prognostic factor. *J Thorac Oncol.* (2015) 10:1292–300. doi: 10.1097/JTO.0000000000000620
17. Engelman JA, Zejnullahu K, Mitsudomi T, Song Y, Hyland C, Park JO, et al. MET amplification leads to gefitinib resistance in lung cancer by activating ERBB3 signaling. *Science.* (2007) 316:1039–43. doi: 10.1126/science.1141478
18. Yano S, Wang W, Li Q, Matsumoto K, Sakurama H, Nakamura T, et al. Hepatocyte growth factor induces gefitinib resistance of lung adenocarcinoma with epidermal growth factor receptor-activating mutations. *Cancer Res.* (2008) 68:9479–87. doi: 10.1158/0008-5472.CAN-08-1643
19. Ou SH, Kwak EL, Siwak-Tapp C, Dy J, Bergethon K, Clark JW, et al. Activity of crizotinib (PF02341066), a dual mesenchymal-epithelial transition (MET) and anaplastic lymphoma kinase (ALK) inhibitor, in a non-small cell lung cancer patient with de novo MET amplification. *J Thorac Oncol.* (2011) 6:942–6. doi: 10.1097/JTO.0b013e31821528d3
20. Paik PK, Drilon A, Fan PD, Yu H, Rekhtman N, Ginsberg MS, et al. Response to MET inhibitors in patients with stage IV lung adenocarcinomas harboring MET mutations causing exon 14 skipping. *Cancer Discov.* (2015) 5:842–9. doi: 10.1158/2159-8290.CD-14-1467
21. Drilon A, Cappuzzo F, Ou SI, Camidge DR. Targeting MET in lung cancer: will expectations finally be MET? *J Thorac Oncol.* (2017) 12:15–26. doi: 10.1016/j.jtho.2016.10.014
22. Kishi K, Sakai H, Seto T, Kozuki T, Nishio M, Imamura F, et al. First-line onartuzumab plus erlotinib treatment for patients with MET-positive and EGFR mutation-positive non-small-cell lung cancer. *Cancer Treat Res Commun.* (2019) 18:100113. doi: 10.1016/j.ctarc.2018.10.004
23. Huang L, An SJ, Chen ZH, Su J, Yan HH, Wu YL. MET expression plays differing roles in non-small-cell lung cancer patients with or without EGFR mutation. *J Thorac Oncol.* (2014) 9:725–8. doi: 10.1097/jto.0000000000000105
24. Tran TN, Selinger CI, Kohonen-Corish MR, McCaughan B, Kennedy C, O'Toole SA, et al. Alterations of MET gene copy number and protein expression in primary non-small-cell lung cancer and corresponding nodal metastases. *Clin Lung Cancer.* (2016) 17:30–8.e31. doi: 10.1016/j.cllc.2015.08.002
25. Tsakonas G, Botling J, Micke P, Rivard C, LaFleur L, Mattsson J, et al. c-MET as a biomarker in patients with surgically resected non-small cell lung cancer. *Lung Cancer.* (2019) 133:69–74. doi: 10.1016/j.lungcan.2019.04.028
26. Onitsuka T, Uramoto H, Ono K, Takenoyama M, Hanagiri T, Oyama T, et al. Comprehensive molecular analyses of lung adenocarcinoma with regard to the epidermal growth factor receptor, K-ras, MET,

and hepatocyte growth factor status. *J Thorac Oncol.* (2010) 5:591–6. doi: 10.1097/JTO.0b013e3181d0a4db
27. Park S, Choi YL, Sung CO, An J, Seo J, Ahn MJ, et al. High MET copy number and MET overexpression: poor outcome in non-small cell lung cancer patients. *Histol Histopathol.* (2012) 27:197–207. doi: 10.14670/HH-27.197
28. Sun W, Song L, Ai T, Zhang Y, Gao Y, Cui J. Prognostic value of MET, cyclin D1 and MET gene copy number in non-small cell lung cancer. *J Biomed Res.* (2013) 27:220–30. doi: 10.7555/JBR.27.20130004
29. Tachibana K, Minami Y, Shiba-Ishii A, Kano J, Nakazato Y, Sato Y, et al. Abnormality of the hepatocyte growth factor/MET pathway in pulmonary adenocarcinogenesis. *Lung Cancer.* (2012) 75:181–8. doi: 10.1016/j.lungcan.2011.07.008
30. Tsuta K, Kozu Y, Mimae T, Yoshida A, Kohno T, Sekine I, et al. c-MET/phospho-MET protein expression and MET gene copy number in non-small cell lung carcinomas. *J Thorac Oncol.* (2012) 7:331–9. doi: 10.1097/JTO.0b013e318241655f
31. Parmar MK, Torri V, Stewart L. Extracting summary statistics to perform meta-analyses of the published literature for survival endpoints. *Stat Med.* (1998) 17:2815–34.
32. Williamson PR, Smith CT, Hutton JL, Marson AG. Aggregate data meta-analysis with time-to-event outcomes. *Stat Med.* (2002) 21:3337–51. doi: 10.1002/sim.1303
33. Tierney JF, Stewart LA, Ghersi D, Burdett S, Sydes MR. Practical methods for incorporating summary time-to-event data into meta-analysis. *Trials.* (2007) 8:16. doi: 10.1186/1745-6215-8-16
34. Stang A. Critical evaluation of the Newcastle-Ottawa scale for the assessment of the quality of nonrandomized studies in meta-analyses. *Eur J Epidemiol.* (2010) 25:603–5. doi: 10.1007/s10654-010-9491-z
35. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ.* (2003) 327:557–60. doi: 10.1136/bmj.327.7414.557
36. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst.* (1959) 22:719–48.
37. Tobias A. Assessing the influence of a single study in the meta-analysis estimate. *Stata Tech Bull.* (1999) 8:7526–9.
38. Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. *Biometrics.* (1994) 50:1088–101.
39. Sun W, Ai T, Gao Y, Zhang Y, Cui J, Song L. Expression and prognostic relevance of MET and phospho-BAD in non-small cell lung cancer. *Onco Targets Ther.* (2013) 6:1315–23. doi: 10.2147/OTT.S50428
40. Takanami I, Tanana F, Hashizume T, Kikuchi K, Yamamoto Y, Yamamoto T, et al. Hepatocyte growth factor and c-Met/hepatocyte growth factor receptor in pulmonary adenocarcinomas: an evaluation of their expression as prognostic markers. *Oncology.* (1996) 53:392–7. doi: 10.1159/000227594
41. Tokunou M, Niki T, Eguchi K, Iba S, Tsuda H, Yamada T, et al. c-MET expression in myofibroblasts: role in autocrine activation and prognostic significance in lung adenocarcinoma. *Am J Pathol.* (2001) 158:1451–63. doi: 10.1016/S0002-9440(10)64096-5
42. Masuya D, Huang C, Liu D, Nakashima T, Kameyama K, Haba R, et al. The tumour-stromal interaction between intratumoral c-Met and stromal hepatocyte growth factor associated with tumour growth and prognosis in non-small-cell lung cancer patients. *Br J Cancer.* (2004) 90:1555–62. doi: 10.1038/sj.bjc.6601718
43. Wang X, Song N, Zhang Y, Cai Y, Liu Y, Qu X, et al. Coexpression of c-Met and Notch-1 correlates with poor prognosis in resected non-small-cell lung cancer. *Tumour Biol.* (2015) 36:7053–9. doi: 10.1007/s13277-015-3404-4
44. Tong JH, Yeung SF, Chan AW, Chung LY, Chau SL, Lung RW, et al. MET Amplification and Exon 14 splice site mutation define unique molecular subgroups of non-small cell lung carcinoma with poor prognosis. *Clin Cancer Res.* (2016) 22:3048–56. doi: 10.1158/1078-0432.CCR-15-2061
45. Zhang J, Sun J, Zhang Z, Liang X, Luo Y, Wu S, et al. Protein overexpression and gene amplification of cellular mesenchymal-epithelial transition factor is associated with poor prognosis in micropapillary-predominant subtype pulmonary adenocarcinoma. *Hum Pathol.* (2018) 72:59–65. doi: 10.1016/j.humpath.2017.10.031
46. Kim IH, Lee IH, Lee JE, Hong SH, Kim TJ, Lee KY, et al. Clinical significance of C-MET overexpression and epidermal growth factor receptor mutation in platinum-based adjuvant chemotherapy outcome in surgically resected lung adenocarcinoma. *Ann Surg Oncol.* (2017) 24:770–7. doi: 10.1245/s10434-016-5599-z
47. Galleges Ruiz MI, Floor K, Steinberg SM, Grunberg K, Thunnissen FB, Belien JA, et al. Combined assessment of EGFR pathway-related molecular markers and prognosis of NSCLC patients. *Br J Cancer.* (2009) 100:145–52. doi: 10.1038/sj.bjc.6604781
48. Dziadziuszko R, Wynes MW, Singh S, Asuncion BR, Ranger-Moore J, Konopa K, et al. Correlation between MET gene copy number by silver *in situ* hybridization and protein expression by immunohistochemistry in non-small cell lung cancer. *J Thorac Oncol.* (2012) 7:340–7. doi: 10.1097/JTO.0b013e318240ca0d
49. Spigel DR, Ervin TJ, Ramlau RA, Daniel DB, Goldschmidt JHJr, Blumenschein GRJr, et al. Randomized phase II trial of Onartuzumab in combination with erlotinib in patients with advanced non-small-cell lung cancer. *J Clin Oncol.* (2013) 31:4105–14. doi: 10.1200/JCO.2012.47.4189
50. Olivero M, Rizzo M, Madeddu R, Casadio C, Pennacchietti S, Nicotra MR, et al. Overexpression and activation of hepatocyte growth factor/scatter factor in human non-small-cell lung carcinomas. *Br J Cancer.* (1996) 74:1862–8. doi: 10.1038/bjc.1996.646
51. Breindel JL, Haskins JW, Cowell EP, Zhao M, Nguyen DX, Stern DF. EGF receptor activates MET through MAPK to enhance non-small cell lung carcinoma invasion and brain metastasis. *Cancer Res.* (2013) 73:5053–65. doi: 10.1158/0008-5472.CAN-12-3775
52. De Bacco F, Luraghi P, Medico E, Reato G, Girolami F, Perera T, et al. Induction of MET by ionizing radiation and its role in radioresistance and invasive growth of cancer. *J Natl Cancer Inst.* (2011) 103:645–61. doi: 10.1093/jnci/djr093
53. Ozasa H, Oguri T, Maeno K, Takakuwa O, Kunii E, Yagi Y, et al. Significance of c-MET overexpression in cytotoxic anticancer drug-resistant small-cell lung cancer cells. *Cancer Sci.* (2014) 105:1032–9. doi: 10.1111/cas.12447
54. Liu F, Song S, Yi Z, Zhang M, Li J, Yang F, et al. HGF induces EMT in non-small-cell lung cancer through the hBVR pathway. *Eur J Pharmacol.* (2017) 811:180–90. doi: 10.1016/j.ejphar.2017.05.040
55. Begemann D, Anastos H, Kyprianou N. Cell death under epithelial-mesenchymal transition control in prostate cancer therapeutic response. *Int J Urol.* (2018) 25:318–26. doi: 10.1111/iju.13505
56. DeSantis CE, Lin CC, Mariotto AB, Siegel RL, Stein KD, Kramer JL, et al. Cancer treatment and survivorship statistics, 2014. *CA Cancer J Clin.* (2014) 64:252–71. doi: 10.3322/caac.21235
57. Shi P, Oh YT, Zhang G, Yao W, Yue P, Li Y, et al. Met gene amplification and protein hyperactivation is a mechanism of resistance to both first and third generation EGFR inhibitors in lung cancer treatment. *Cancer Lett.* (2016) 380:494–504. doi: 10.1016/j.canlet.2016.07.021
58. Wang Q, Yang S, Wang K, Sun SY. MET inhibitors for targeted therapy of EGFR TKI-resistant lung cancer. *J Hematol Oncol.* (2019) 12:63. doi: 10.1186/s13045-019-0759-9
59. Dimou A, Non L, Chae YK, Tester WJ, Syrigos KN. MET gene copy number predicts worse overall survival in patients with non-small cell lung cancer (NSCLC); a systematic review and meta-analysis. *PLoS ONE.* (2014) 9:e107677. doi: 10.1371/journal.pone.0107677
60. Guo B, Cen H, Tan X, Liu W, Ke Q. Prognostic value of MET gene copy number and protein expression in patients with surgically resected non-small cell lung cancer: a meta-analysis of published literatures. *PLoS ONE.* (2014) 9:e99399. doi: 10.1371/journal.pone.0099399.
61. Kim JH, Kim HS, Kim BJ. Prognostic value of MET copy number gain in non-small-cell lung cancer: an updated meta-analysis. *J Cancer.* (2018) 9:1836–45. doi: 10.7150/jca.24980.
62. Awad MM, Oxnard GR, Jackman DM, Savukoski DO, Hall D, Shivdasani P, et al. MET exon 14 mutations in non-small-cell lung cancer are associated

with advanced age and stage-dependent MET genomic amplification and c-Met overexpression. *J Clin Oncol.* (2016) 34:721–30. doi: 10.1200/JCO.2015.63.4600

63. Camidge DR, Pao W, Sequist LV. Acquired resistance to TKIs in solid tumours: learning from lung cancer. *Nat Rev Clin Oncol.* (2014) 11:473–81. doi: 10.1038/nrclinonc.2014.104
64. Scagliotti G, von Pawel J, Novello S, Ramlau R, Favaretto A, Barlesi F, et al. Phase III multinational, randomized, double-blind, placebo-controlled study of tivantinib (ARQ 197) plus erlotinib versus erlotinib alone in previously treated patients with locally advanced or metastatic nonsquamous non-small-cell lung cancer. *J Clin Oncol.* (2015) 33:2667–74. doi: 10.1200/JCO.2014.60.7317
65. Pyo JS, Kang G, Cho WJ, Choi SB. Clinicopathological significance and concordance analysis of c-MET immunohistochemistry in non-small cell lung cancers: a meta-analysis. *Pathol Res Pract.* (2016) 212:710–6. doi: 10.1016/j.prp.2016.05.006

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

ORIGINAL RESEARCH
published: 09 January 2020
doi: 10.3389/fonc.2019.01485

The Potential of Radiomics Nomogram in Non-invasively Prediction of Epidermal Growth Factor Receptor Mutation Status and Subtypes in Lung Adenocarcinoma

Wei Zhao[1,2], Yuzhi Wu[1], Ya'nan Xu[3], Yingli Sun[2], Pan Gao[2], Mingyu Tan[2], Weiling Ma[2], Cheng Li[2], Liang Jin[2], Yanqing Hua[2], Jun Liu[1] and Ming Li[2,4,5]**

[1] Department of Radiology, Second Xiangya Hospital, Central South University, Changsha, China, [2] Department of Radiology, Huadong Hospital Affiliated to Fudan University, Shanghai, China, [3] School of Biomedical Engineering, Capital Medical University, Beijing, China, [4] Diagnosis and Treatment Center of Small Lung Nodules of Huadong Hospital, Shanghai, China, [5] Institute of Functional and Molecular Medical Imaging, Fudan University, Shanghai, China

OPEN ACCESS

Edited by:
Umberto Malapelle,
University of Naples Federico II, Italy

Reviewed by:
Francesco Pepe,
Department of Public Health,
University of Naples Federico II, Italy
Marzia Del Re,
University of Pisa, Italy

***Correspondence:**
Ming Li
minli77@163.com
Jun Liu
junliu123@csu.edu.cn

Specialty section:
This article was submitted to Thoracic Oncology, a section of the journal Frontiers in Oncology

Received: 08 October 2019
Accepted: 10 December 2019
Published: 09 January 2020

Citation:
Zhao W, Wu Y, Xu Y, Sun Y, Gao P, Tan M, Ma W, Li C, Jin L, Hua Y, Liu J and Li M (2020) The Potential of Radiomics Nomogram in Non-invasively Prediction of Epidermal Growth Factor Receptor Mutation Status and Subtypes in Lung Adenocarcinoma. Front. Oncol. 9:1485. doi: 10.3389/fonc.2019.01485

Purpose: Up to 50% of Asian patients with NSCLC have *EGFR* gene mutations, indicating that selecting eligible patients for *EGFR*-TKIs treatments is clinically important. The aim of the study is to develop and validate radiomics-based nomograms, integrating radiomics, CT features and clinical characteristics, to non-invasively predict *EGFR* mutation status and subtypes.

Materials and Methods: We included 637 patients with lung adenocarcinomas, who performed the *EGFR* mutations analysis in the current study. The whole dataset was randomly split into a training dataset ($n = 322$) and validation dataset ($n = 315$). A sub-dataset of *EGFR*-mutant lesions (*EGFR* mutation in exon 19 and in exon 21) was used to explore the capability of radiomic features for predicting *EGFR* mutation subtypes. Four hundred seventy-five radiomic features were extracted and a radiomics sore (R-score) was constructed by using the least absolute shrinkage and selection operator (LASSO) regression in the training dataset. A radiomics-based nomogram, incorporating clinical characteristics, CT features and R-score was developed in the training dataset and evaluated in the validation dataset.

Results: The constructed R-scores achieved promising performance on predicting *EGFR* mutation status and subtypes, with AUCs of 0.694 and 0.708 in two validation datasets, respectively. Moreover, the constructed radiomics-based nomograms excelled the R-scores, clinical, CT features alone in terms of predicting *EGFR* mutation status and subtypes, with AUCs of 0.734 and 0.757 in two validation datasets, respectively.

Conclusions: Radiomics-based nomogram, incorporating clinical characteristics, CT features and radiomic features, can non-invasively and efficiently predict the *EGFR* mutation status and thus potentially fulfill the ultimate purpose of precision medicine. The methodology is a possible promising strategy to predict *EGFR* mutation subtypes, providing the support of clinical treatment scenario.

Keywords: *EGFR*, radiomics, nomogram, lung adenocarcinomas, CT

KEY POINTS

1. We developed and validated two Radiomics-based nomograms, incorporating clinical characteristics, CT features and radiomic features, to non-invasively predict the *EGFR* mutation status and subtypes with the aim to potentially fulfill the ultimate purpose of precision medicine.
2. The presented results indicate that radiomics-based nomogram may potentially facilitate scalable precision medicine on identifying eligible patients of lung adenocarcinoma for EGFR-targeted therapy.

INTRODUCTION

Lung cancer is the leading cause cancer-related death both in male and female (1). Non-small cell lung cancer (NSCLC) accounts for more than 80% of lung cancers, of which lung adenocarcinoma is the most common histological subtype (2). With the advances of genomics, molecular-targeted therapy like using tyrosine kinase inhibitors (TKIs), which targets the epidermal growth factor receptor (*EGFR*) mutations, is recommended as first-line system therapy before first-line therapy by National Comprehensive Cancer Network (NCCN) for patients with advanced *EGFR*-mutant NSCLC (2) and proved to substantially improve the progression-free survival (PFS) compared with conventional chemotherapy (3, 4). Up to 50% of Asian patients with NSCLC have *EGFR* gene mutations (5), indicating that selecting eligible patients for EGFR-TKIs treatments is clinically important. In patients with NSCLC, the most commonly found *EGFR* mutations are deletions in exon 19 (45%) and in exon 21 (L858R in 40%) in patients with *EGFR* mutations (2). Both mutations are associated with sensitivity to the small molecule TKIs as well as erlotinib, gefitinib, afatinib, and osimertinib (2), however, with different survival outcomes in response to both *EGFR*-TKIs and chemotherapy (6). Therefore, identifying *EGFR* mutation subtypes, especially those responsive to TKI treatment, seems to be more critically and scientifically important than just predicting *EGFR* mutation status.

In this context, though more and more research has emerged on the non-invasive prediction of EGFR mutation status in recent years (7–9), no predictors are recommended for selecting patients in clinical decision-making. Moreover, substantial discrepancies are presented to date with regarding to some features, especially semantic features derived from medical images (10). Buoyed by the availability of big data and state-of-art data analysis strategy, such as radiomics and deep learning, decoding tumor phenotype to precisely predict genotype has becoming the point of attention (11). Several studies have investigated the potential ability of radiomics to non-invasively predicting *EGFR* mutation status and show promising results (12–15). Few results are finally applied in clinical practice yet due to the complicated procedure (e.g., time consuming, poor reproducibility, remaining the operator-dependency that is not biases-free, and so on) of radiomic researches (16). In view of this, models that giving an individual numerical probability of a clinical event (e.g., nomogram) rather than a predicting accuracy, may be more suitable and convenient for clinical application.

In the current study, we aim to build radiomics-based nomograms, integrating radiomics, CT imaging features and clinical characteristics, to non-invasively predict *EGFR* mutation status and subtypes (exon 19 and 21mutation).

Abbreviations: *EGFR*, epidermal growth factor receptor; NSCLC, non-small cell lung carcinoma; PFS, progression-free survival; *ALK*, antigen anaplastic lymphoma kinase; *ROS1*, c-ros oncogene 1; TKIs, tyrosine kinase inhibitors.

MATERIALS AND METHODS

Patient Selection and Dataset Preparation

This retrospective study was approved by the institutional review board (No. 20170103), which waived the requirement for patients' informed consent referring to the CIOMS guideline. The flowchart of our study was showed in **Figure 1**.

A search of Picture Achieving and Communication System (PACS) and pathological system from January 2013 to June 2018 was performed by one author with the following inclusion criteria: (1) Available thin-slice chest CT (<1.5 mm) images before biopsies or surgical treatment; (2) Available detailed histological reports of adenocarcinoma; (3) Available detailed *EGFR* mutations testing reports; (4) No any prior treatment before *EGFR* mutations analysis.

Finally, 637 patients were included. Of the 637 lesions, 342 lesions tested positively for an *EGFR* mutation (*EGFR* Mut), where 295 lesions were classified as wild-type lung adenocarcinomas (*EGFR* WT). Note that only one malignant nodule was studied for each patient due to the availability of *EGFR* testing report. Among the 342 patients with *EGFR* Mut, 130 patients were detected an *EGFR* Mut in exon 19, whereas 190 patients were detected an *EGFR* Mut in exon 21. The clinical and histopathologic variables, including age, sex, smoking status, tumor size, tumor location, histological subtypes etc. were presented in **Table 1**.

Two tasks were investigated in the current study: task (a), differentiating *EGFR* Mut from *EGFR* WT; task (b), differentiating *EGFR* Mut in exon 19 from in exon 21. Each dataset in two tasks ($n = 637$ and $n = 320$) was split into 10 groups (1–10), each subset was randomly selected by choosing 10% of each of the 2 categories. In task (a), groups 1–5 were defined as training dataset, the rest groups were defined as validation dataset. In task (b), considering the insufficient training data, groups 1–6 were defined as training dataset, the rest groups were defined as validation dataset. Note that we only included *EGFR* mutations in exon 19 and 21 in task b to avoid the sparse training data and the disbalance of data distribution. Mover, constructing a model to predict these two exons in patients with EGFR mutation is clinically reasonable (see Introduction).

CT Acquisition and Imaging Interpretation

All included patients were performed with the following six scanners: GE Discovery CT750 HD, 64-slice LightSpeed VCT, Revolution CT (GE Medical Systems); Somatom Definition flash, Somatom Sensation-16, Somatom Force (Siemens Medical Solutions). The acquisition parameters were as follows: 120

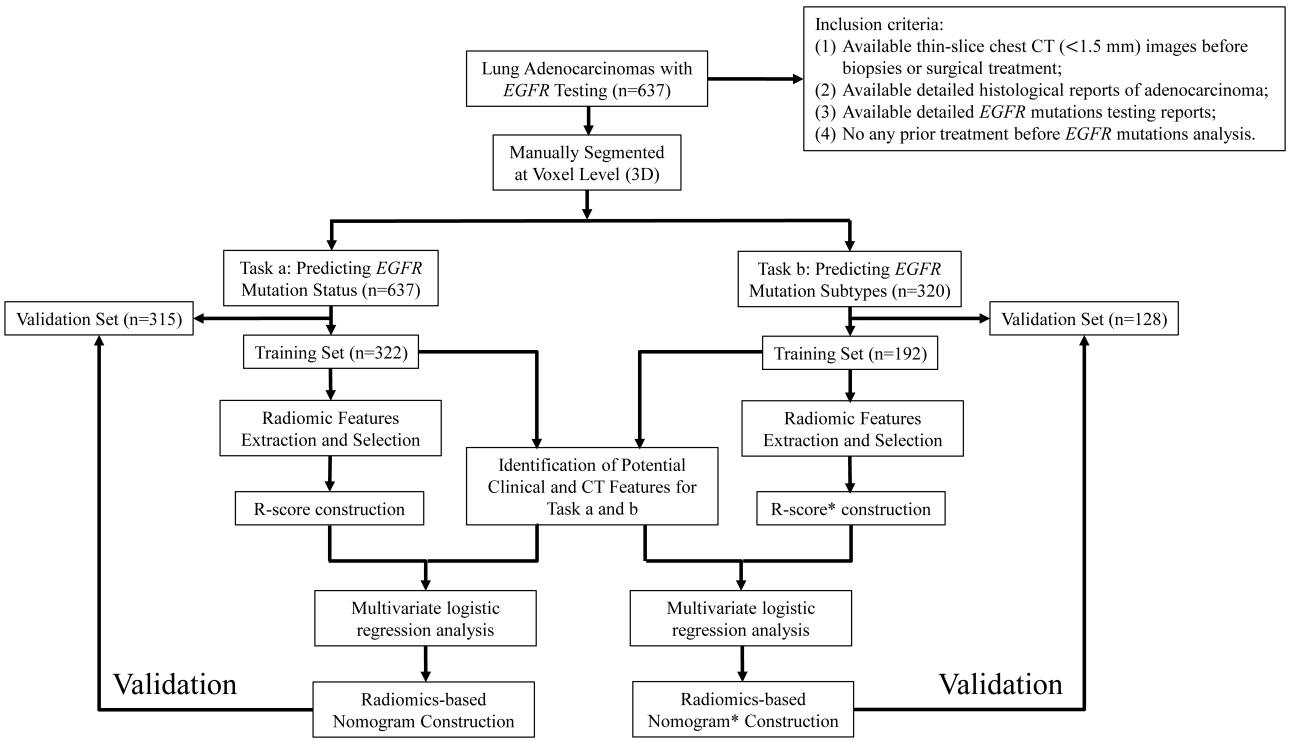


FIGURE 1 | The flowchart of the study.

kVp; 100– 200 mAs; pitch, 0.75–1.5; and collimation, 1–1.5 mm, respectively. All imaging data were reconstructed by using a medium sharp reconstruction algorithm with a thickness of 0.75–1.5 mm. CT images were acquired in the supine position at full inspiration for all patients. Only plain CT images were used in the current study. Two radiologists (with 12 and 3 years of experience in chest CT interpretation) independently interpreted the CT images, blinded to clinical and histologic findings. Thirteen CT features (**Table 1**) were evaluated. The definitions of these features were described in previous study (17, 18). A re-evaluation for achieving a consensus was performed to solve the disagreement between two radiologists.

Segmentation and Radiomic Features Extraction

All nodules were manually delineated slice by slice using a medical image processing and navigation software 3D Slicer (version 4.8.0, Brigham and Women's Hospital) by one author (with 5 years of experience in chest CT interpretation), then the volume of interests (VOIs) were confirmed by another radiologist (with 12 years of experience in chest CT interpretation). Fifty randomly selected nodules were independently segmented by the two authors for feature reproductive analysis. Images and VOIs with NII format were exported for further analysis. Radiomic features from three categories, including 50 gray-level histogram features, 325 gray-level co-occurrence matrix (GLCM) features, and 100 gray-level run lengths matrix (GLRLM) features, were extracted using Matlab 2016b (MathWorks, Natick, USA). The details of extracted radiomic features were presented in **Supplementary Data**. Radiomic feature extraction methodology was described in our previous study (18).

Features Selection and Radiomic Score Construction

The least absolute shrinkage and selection operator (LASSO) method, which is an accepted algorithm for feature selection in high-dimensional variables (19), was applied to select the features that were most distinguishable and build a logistic regression model in training dataset. Then a radiomic score (R-score) was calculated for each lesion using features selected by LASSO and weighted by the respective coefficients.

Nomogram Construction and Validation

Univariate analysis was firstly to identify the potential predictors among clinical characteristics, CT features and R-score. Factors that associated with *EGFR* mutation status and subtypes were then included to multivariate analysis to identify the independent predictors. Furthermore, the identified independent factors were selected to construct the final nomogram in the training dataset.

Histologic Evaluation and EGFR Mutation Analysis

The included lung adenocarcinomas were categorized according to the 2011 IASLC/ATS/ERS classification system (20) (drug target-associated). Molecular analysis of mutation

TABLE 1 | Clinical and histological characteristics of included patients.

Characteristics	Number	Percentage
Gender		
Male	269	42.2
Female	368	57.8
Mean age (range) (year)		
Male	62.0 ± 11.8 (27–85)	–
Female	58.3 ± 11.9 (22–85)	–
Total	59.9 ± 12.0 (22–85)	–
Mean size (range) (cm)	1.85 ± 1.29 (0.4–8.6)	–
Smoke		
Never smoker	588	92.3
Current or former smoker	49	7.7
Location		
Right lobe	378	59.3
Left lobe	259	40.7
Pathology		
Adenocarcinoma *in situ*	32	5.0
Minimally invasive adenocarcinoma	174	27.3
Invasive adenocarcinoma	431	67.7
TMN classification (eighth edition)		
0	32	5.0
I	388	60.9
II	10	1.6
III	12	1.9
IV	195	30.6
EGFR Mut	342	53.4
EGFR Mut in exon 19	130	38.0
EGFR Mut in exon 21	190	55.6

status of EGFR exons 18–21 was examined using a PCR-based amplification-refractory mutation system (ARMS) by AmoyDx company.

Statistical Analysis

Statistical analysis was performed using R software (version 3.4.2; http://www.Rproject.org). The mean values and standard deviations were expressed for continuous variables (age, lesions size), and frequency or percentage for categorical variables. The Wilcoxon rank sum test and the x^2-test were used to compare medians and proportions between two groups, respectively. Predictive performance was evaluated by the area under the curve (AUC) of the receiver operator characteristic (ROC). DeLong test was used to evaluate the difference of the ROCs between various models (21). Hosmer-Lemeshow test was used to evaluate the goodness-of-fit of the constructed nomogram. ICC analysis was performed using "irr" package, features with an ICC> 075 were consider as robust features. Lasso logistic regression was done using the "glmnet" package. Multivariate logistic regression, nomograms and calibration plots were done with the "rms" package. A $P<0.05$ indicated a significant difference.

RESULTS

Associations Between Clinical, CT Features and EGFR Mutation Status and Subtypes

Associations between clinical, CT features and *EGFR* mutation status and subtypes were presented in **Tables 2**, **3**. The incidence of harboring *EGFR* mutation was significantly higher in female than male in two datasets ($P = 0.002$, $P = 0.013$, respectively). Patients with *EGFR* Mut had a higher age in the current study. Smoking was not a significant factor to differentiate *EGFR* Mut lesions from *EGFR* WT lesions. In terms of radiographic features, 9 features, including size, margin, shape, pleural retraction, bronchiole change, lobulation, speculation, peripheral emphysema, peripheral fibrosis were significantly associated with *EGFR* mutation status. Patients with *EGFR* mutations in 21 exon had a higher age than those with *EGFR* mutations in 19 exon in two datasets ($P = 0.013$, $P = 0.003$, respectively). No other clinical and CT features were identified as potential factors to predict the *EGFR* mutation subtypes.

Construction of R-Score and the Association Between R-Score (R-Score*) and EGFR Mutation Status and Subtypes

After performing ICC analysis, 425 of 475 radiomics were identified as robust features. The LASSO logistic regression model was performed to select the most distinguishable features in training dataset, resulting in 11 features left (**Figure 2**). Subsequently, the 11 potential predictors were consequently conducted into a R-score by using the following formula: R-score = −1.072477 + 0.007008 * mean_10_0 + 0.038891 * Homogeneity 0_90_0 + 1.86E-05 * Contrast 45_45_0 + 8.54E-05 * Contrast 90_135_0 + 6.29E-05 * Contrast 90_135_1 – 0.039584 * skewness_1.5 – 0.254939 * skewness_2 + 1.15E-06 * RLN_90_2.5 + 7.46E-05 * Contrast 90_90_2.5 – 9.69E-05 * Contrast 0_0_2.5 – 6.342383311 * Homogeneity 0_0_2.5. The formula caption was presented in **Supplementary Data** (Referring to the formula for calculating R-score*).

The R-score was calculated for each lesion in two datasets of task a. *EGFR*-Mut lung adenocarcinomas had a lower R-score than *EGFR*-WT ones in training dataset (−0.40 ± 0.50 vs. 0.05 ± 0.68, $P = 0.000$), which was confirmed in validation dataset (−0.37 ± 0.51 vs. 0.01 ± 0.0.58, $P = 0.000$) (**Table 2**). The proposed R-score showed a good performance in differentiating *EGFR* mutation status with AUCs of 0.708, 0.694 in training and validation datasets (**Figures 3A,B**). The Hosmer-Lemeshow test for R-score yielded a non-significant statistic in the training and validation datasets ($P = 0.644$, $P = 0.657$, respectively), indicating that there was no departure from a perfect fit.

In terms of the task of predicting the *EGFR* mutation subtypes (task b), 32 features were finally left after performing the LASSO analysis (**Supplementary Data**). A R-score* was also calculated for each lesion by using the formula presented in the **Supplementary Data**. Lung adenocarcinomas with *EGFR* Mut in exon 19 had a lower R-score* than ones with *EGFR* Mut in exon 21 in training dataset (−0.39 ± 1.35 vs. 0.27 ± 1.12, *P*

TABLE 2 | Basal characteristics of patients in training and validation set (task a).

Characteristics	Training set ($n = 322$)		*P*	Validation set ($n = 315$)		*P*
	***EGFR* Mut ($n = 172$)**	***EGFR* WT ($n = 150$)**		***EGFR* Mut ($n = 170$)**	***EGFR* WT ($n = 145$)**	
Clinical Characteristics						
Gender			0.002			0.013
Male	50 (29.1)	69 (46)		70 (41.2)	80 (55.2)	
Female	122 (70.9)	81 (54)		100 (58.8)	65 (44.8)	
Age (year)	61.92 ± 57.64	57.46 ± 12.59	0.001	61.85 ± 11.63	57.63 ± 11.73	0.000
Smoke			0.154			0.404
Never smoker	163 (94.8)	136 (90.7)		158 (92.9)	131 (90.3)	
Current or former smoker	9 (5.2)	14 (9.3)		12 (7.1)	14 (9.7)	
Radiographic Characteristics						
Size (cm)	2.09 ± 1.24	1.53 ± 1.26	0.000	2.06 ± 1.31	1.65 ± 1.25	0.000
Type			0.091			0.472
Pure or part solid GGN	117 (68.0)	103 (68.7)		111 (65.3)	89 (61.4)	
Solid	55 (32.0)	47 (31.3)		59 (34.7)	56 (38.6)	
Margin			0.002			0.039
Easily differentiated	52 (30.2)	71 (47.3)		58 (34.1)	66 (45.5)	
Uneasily differentiated	120 (69.8)	79 (52.7)		112 (65.9)	79 (54.5)	
Shape			0.000			0.007
Round or oval	41 (23.8)	67 (44.7)		46 (27.1)	60 (41.4)	
Irregular	131 (76.2)	83 (55.3)		124 (72.9)	85 (58.6)	
Pleural retraction			0.000			0.015
Present	87 (50.6)	46 (30.7)		83 (48.8)	51 (35.2)	
Absent	85 (49.3)	104 (69.3)		87 (51.2)	94 (64.8)	
Bubble lucency			0.317			0.978
Present	51 (29.7)	37 (24.7)		49 (28.8)	42 (29.0)	
Absent	121 (70.3)	113 (75.3)		121 (71.2)	103 (71.0)	
Vascular change			0.050			0.575
Present	120 (69.8)	89 (59.3)		113 (66.5)	92 (63.4)	
Absent	52 (30.2)	61 (40.7)		57 (33.5)	53 (36.6)	
Bronchiole change			0.000			0.003
Present	90 (52.3)	44 (29.3)		86 (50.6)	49 (33.8)	
Absent	82 (47.7)	106 (70.7)		84 (49.4)	96 (66.2)	
Lobulation			0.031			0.034
Present	79 (45.9)	87 (58.0)		70 (41.2)	77 (53.1)	
Absent	93 (54.7)	63 (42.0)		100 (58.8)	68 (46.9)	
Spiculation			0.003			0.041
Present	87 (50.6)	51 (34.0)		80 (47.1)	85 (58.6)	
Absent	85 (49.4)	99 (66.0)		90 (52.9)	60 (41.4)	
Peripheral Emphysema			0.014			0.002
Present	3 (1.7)	11 (7.3)		6 (3.5)	19 (13.1)	
Absent	169 (98.3)	139 (92.7)		164 (96.5)	126 (86.9)	
Peripheral fibrosis			0.022			0.009
Present	55 (32.0)	31 (20.7)		63 (37.1)	34 (23.4)	
Absent	117 (68.0)	119 (79.3)		107 (62.9)	111 (76.6)	
Pleural effusion			0.177			0.506
Present	6 (3.5)	1 (0.7)		1 (0.6)	3 (2.1)	
Absent	166 (96.5)	149 (99.3)		169 (99.4)	142 (97.9)	
R-score	−0.40 ± 0.50	0.05 ± 0.68	0.000	−0.37 ± 0.51	0.01 ± 0.0.58	0.000

TABLE 3 | Basal characteristics of patients in training and validation set (task b).

Characteristics	Training set ($n = 192$)		*P*	Validation set ($n = 128$)		*P*
	***EGFR* Mut in exon 19 ($n = 78$)**	***EGFR* Mut in exon 21 ($n = 114$)**		***EGFR* Mut in exon 19 ($n = 52$)**	***EGFR* Mut in exon 21 ($n = 76$)**	
Clinical Characteristics						
Gender			0.848			0.858
Male	27 (34.6)	41 (36.0)		17 (32.7)	26 (34.2)	
Female	51 (65.4)	73 (64.0)		35 (67.3)	50 (65.8)	
Age (year)	59.82 ± 11.66	64.13 ± 10.47	0.013	57.77 ± 12.90	64.70 ± 11.13	0.003
Smoke			0.595			0.853
Never smoker	74 (94.9)	106 (93.0)		49 (94.2)	71 (93.4)	
Current or former smoker	4 (5.1)	8 (7.0)		3 (5.8)	5 (6.6)	
Radiographic Characteristics						
Size (cm)	2.07 ± 1.30	2.19 ± 1.27	0.432	1.77 ± 0.98	2.26 ± 1.39	0.051
Type			0.264			0.458
Pure or part solid GGN	48 (61.5)	79 (69.3)		33 (63.5)	53 (69.7)	
Solid	30 (38.5)	35 (30.7)		19 (36.5)	23 (30.3)	
Margin			0.412			0.477
Easily differentiated	21 (26.9)	37 (32.5)		16 (30.8)	28 (36.8)	
Uneasily differentiated	57 (73.1)	77 (67.5)		36 (69.2)	48 (63.2)	
Shape			0.869			0.064
Round or oval	17 (21.8)	26 (22.8)		17 (32.7)	14 (18.4)	
Irregular	61 (78.2)	88 (77.2)		35 (67.3)	62 (81.6)	
Pleural retraction			0.344			0.620
Present	41 (52.6)	52 (45.6)		29 (55.8)	39 (51.3)	
Absent	37 (47.4)	62 (54.4)		23 (44.2)	37 (48.7)	
Bubble lucency			0.567			0.895
Present	19 (24.4)	32 (28.1)		17 (32.7)	24 (31.6)	
Absent	59 (75.6)	82 (71.9)		35 (67.3)	52 (68.4)	
Vascular change			0.663			0.497
Present	55 (70.5)	77 (67.5)		34 (65.4)	54 (71.1)	
Absent	23 (29.5)	37 (32.5)		18 (34.6)	22 (28.9)	
Bronchiole Change			0.739			0.284
Present	45 (57.7)	63 (55.3)		21 (40.4)	38 (50.0)	
Absent	33 (42.3)	51 (44.7)		31 (59.6)	38 (50.0)	
Lobulation			0.113			0.831
Present	52 (66.7)	63 (55.3)		27 (51.9)	38 (50.0)	
Absent	26 (33.3)	51 (44.7)		25 (48.1)	38 (50.0)	
Spiculation			0.179			0.566
Present	46 (59.0)	56 (49.1)		28 (53.8)	37 (48.7)	
Absent	32 (41.0)	58 (50.9)		24 (46.2)	39 (51.3)	
Peripheral emphysema			0.394			0.698
Present	0 (0.0)	3 (2.6)		2 (3.8)	2 (2.6)	
Absent	78 (100.0)	111 (97.4)		50 (96.2)	74 (97.4)	
Peripheral fibrosis			0.673			0.216
Present	29 (37.2)	39 (34.2)		15 (28.8)	30 (39.5)	
Absent	49 (62.8)	75 (65.8)		37 (71.2)	46 (60.5)	
Pleural effusion			0.898			0.795
Present	1 (1.3)	3 (2.6)		1 (1.9)	2 (2.6)	
Absent	77 (98.7)	111 (97.4)		51 (98.1)	74 (97.4)	
R-score*	−0.39 ± 1.35	0.27 ± 1.12	0.000	−0.70 ± 1.25	0.23 ± 1.12	0.000

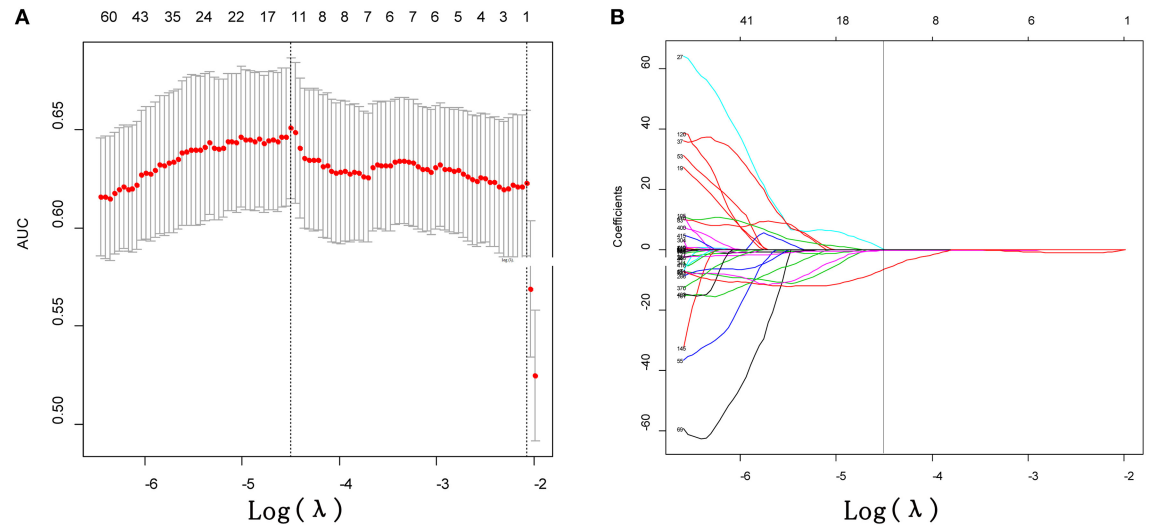


FIGURE 2 | Feature selection using the least absolute shrinkage and selection operator (LASSO) binary logistic regression model. **(A)** Tuning parameter (λ) selection in the LASSO model used 10-fold cross-validation via minimum criteria. The area under the receiver operating characteristic (AUC) curve was plotted vs. log (λ). Dotted vertical lines were drawn at the optimal values by using the minimum criteria. **(B)** LASSO coefficient profiles of the 425 texture features. A coefficient profile plot was produced against the log (λ) sequence. Vertical line was drawn at the value selected using 10-fold cross-validation, where optimal λ (−4.497) resulted in 11 non-zero coefficients.

= 0.000) and in validation dataset (−0.70 ± 1.25 vs. 0.23 ± 1.12, $P = 0.000$) (**Table 3**). The proposed R-score* demonstrated a good performance in differentiating *EGFR* mutation subtypes with AUCs of 0.684, 0.708 in two datasets (**Figures 4A,B**). The Hosmer-Lemeshow test for R-score* yielded a non-significant statistic in the training and validation set ($P = 0.295$, $P = 0.242$, respectively), which suggested that there was no departure from a perfect fit.

Development and Validation of the Radiomics Nomograms for Predicting the EGFR Mutation Status and Subtypes

There was no multicollinearity between the significant factors identified by univariate analysis and R-score. After performing the multivariate analysis, sex, peripheral emphysema and R-score were identified as independent prognostic factors of harboring *EGFR* mutation (**Table 4**) and subsequently incorporated to develop the radiomics-based nomogram (**Figure 3C**). The constructed nomogram obtained a significantly incremental performance for predicting *EGFR* mutation status compared with that of sex and peripheral emphysema (**Table 5**). The Hosmer-Lemeshow test for the nomogram yielded a non-significant statistic in the training and validation set ($P = 0.313$, $P = 0.816$, respectively), which suggested that there was no departure from a perfect fit. Though the nomogram significantly outperformed R-score in training dataset, the statistically difference was not found in validation dataset (**Table 5**).

In terms of task b, age and R-score* were identified as independent prognostic factors of predicting *EGFR* mutation subtypes (**Table 4**) and subsequently incorporated to develop the radiomics-based nomogram* (**Figure 4C**). The Hosmer-Lemeshow test for the nomogram* yielded a non-significant statistic in the training and validation set ($P = 0.760$, $P = 0.413$, respectively), indicating that there was no departure from a perfect fit. Not surprisingly, the constructed nomogram* significantly outperformed age and potentially outperformed R-score* in predicting *EGFR* mutation subtypes (**Table 5**).

DISCUSSION

Non-invasively and preoperatively predicting the *EGFR* mutation status, a field attracted continuous efforts of researchers, can overcome the disadvantages of molecular assays (e.g., high cost, intratumoral heterogeneity, poor sample quality) well and furtherly help clinicians to select the eligible patients for targeted therapy. Moreover, attempting to predict the *EGFR* mutation subtypes, especially those are sensitivity to small TKIs, may provide important information for making finer treatment scenario. In the current study, we developed and validated two radiomics based nomograms, incorporating clinical characteristics, CT features and radiomics, to predict the *EGFR* mutation status and subtypes with promising performance (AUC = 0.734, AUC = 0.757, respectively).

It is well-documented that the *EGFR* Mut rate are significantly higher in female than male (22), which is also confirmed in our study ($P = 0.002$ and $P = 0.013$ in two datasets, respectively). Note that smoking status, another potential clinical factor verified by previous studies (8, 14, 17), failed to show significant association with *EGFR* mutation status in the current study. Note that smoking status is still a contentious risk factor (23) and not recommended to use as the criteria for selecting eligible patients (2). Another reason why the difference between *EGFR* Mut and WT in terms of smoking status is diminished may be that the incidence of lung adenocarcinomas in female

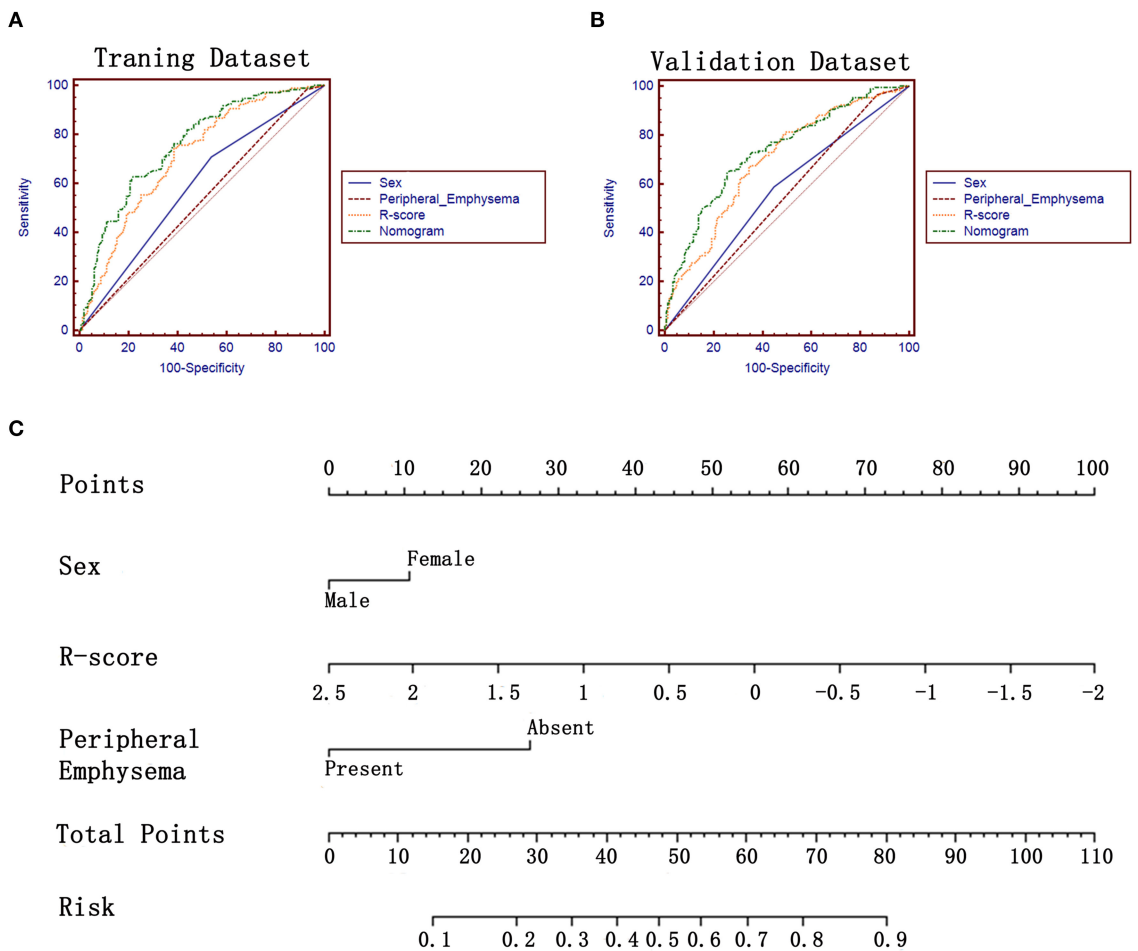


FIGURE 3 | ROC analysis and the constructed nomogram in task a. **(A,B)** ROC analysis of sex, R-score, peripheral emphysema and the constructed nomogram in training and validation datasets, respectively. **(C)** The constructed nomogram for predicting *EGFR* mutation status.

was higher than male (20, 24). Moreover, female may be less likely to be current/former smoker than male. Wu et al. (25) had reported that younger patients (<50 years old) with lung adenocarcinoma had lower EGFR mutation rate, which was not verified in multivariate logistic analysis in the current study. The discrepancy may be contributed to the data bias (ours vs. Wu's: 20.9 vs. 15.5%).

Many prior studies have investigated the role of morphological features in predicting *EGFR* mutation status. The results remain controversial. On one hand, most radiographic features are non-quantitative subjective features and susceptible to the discrepancy of evaluation caused by different knowledge structure of observers. On the other hand, obvious radiographical features are more frequently presented in advanced tumors instead of early-stage tumors. For example, stage I lung cancers account for 60–70% of detected lung cancers in screening programs (26). In our study, 60.9% of included patients were stage I lung cancers. In this context, the differentiating performance of these semantic features may be compromised. Peripheral emphysema was the only independent risk factors for predicting *EGFR* mutation status in the current study.

Another disadvantage of semantic features is that they only reflect few tumor information in biological level. By contrast, radiomics method can encode a more comprehensive level of feature abstraction and thus potentially provide better prediction performance. Previous studies have revealed the potential associations between these engineered features and *EGFR* mutation status (12, 23, 27) and proved that the performance of models can benefit from the integration of radiomics and clinical features (14, 23). These results were also confirmed in our study. Despite the promising results, the complicated process of feature extraction and the inconvenience of formula-based model limit its application in clinical context. Hence, an easy-to-use way for radiomics method is urgently needed. Incorporating multiple radiomic features into a radiomics score can tactfully make multi-marker analyses less complicated to use (18, 28), which is similar to the construction of multi-factor panels. In this study, the LASSO regression model was used to select

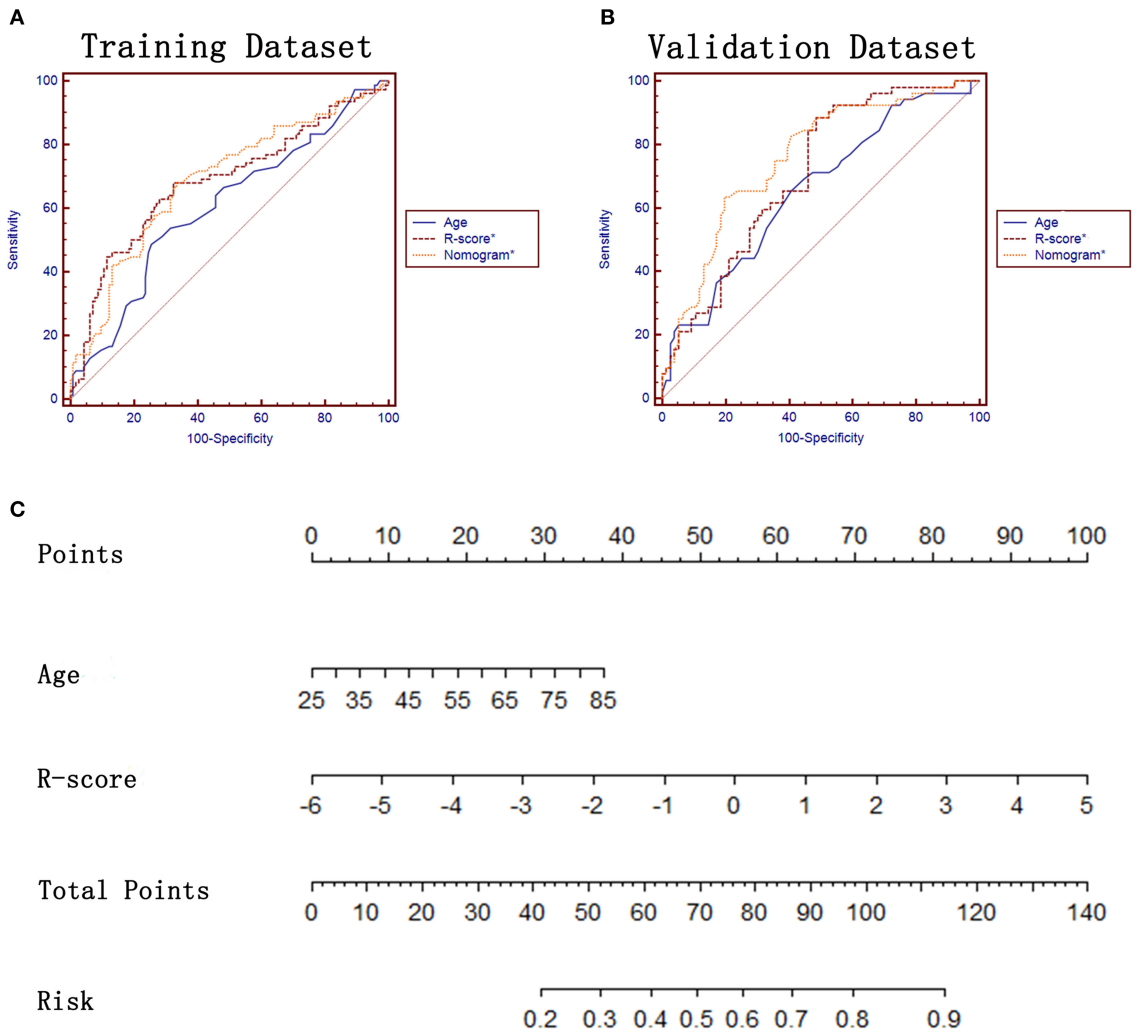


FIGURE 4 | ROC analysis and the constructed nomogram in task b. **(A,B)** ROC analysis of age, R-score* and the constructed nomogram* in training and validation datasets, respectively. **(C)** The constructed nomogram* for predicting *EGFR* mutation subtypes.

and compress the radiomic features and thus construct the R-score. The constructed R-score outperformed the clinical and CT features in predicting *EGFR* mutation status (**Table 5**). Another strategy to make radiomics method friendly use is presenting the results in a more intuitive way, such as the nomogram (29, 30). Nomogram is a statistic model that can provide an individual numerical probability of a clinical event by integrating multiple variables (31, 32). To comprehensively investigate the predictive performance of non-invasively available variables, including clinical factors, imaging semantic features, imaging radiomic features, we adopted the two above-mentioned strategies to construct a radiomics-based nomogram for predicting the *EGFR* mutation status. The constructed nomogram in the current study was conveniently used to individualized predict the probability of harboring *EGFR* mutation by calculation the respective points of three variables (sex, peripheral emphysema, and R-score), with a promising AUC of 0.734. Another promising technique (i.e., liquid biopsy) was consider as an alternative to test *EGFR* mutations. However, the high false-negative rate (30%) need to be further resolved (33).

In patients with NSCLC, the most commonly found *EGFR* mutations are deletions in exon 19 (*Exon19del* in 45%) and in exon 21 (*L858R* in 40%) (2). Both mutations result in activation of the tyrosine kinase domain, and both are associated with sensitivity to the small molecule TKIs. Nevertheless, patients with exon 19 deletion are associated with longer PFS compared to those with L858R mutation after first-line *EGFR*-TKIs (34, 35). In addition, the incidence of T790M mutation, which is associated with acquired resistance to reversible *EGFR*-TKIs (36), might be different between exon 19 deletions and L858R mutations (6), resulting in different treatment scenarios. As a result, furtherly predicting the specific *EGFR* mutation subtypes may be clinically important. Inspired by the satisfied results of predicting *EGFR* mutation status, the potential relations between

TABLE 4 | Multivariate logistic regression of risk characteristics for predicting the *EGFR* mutation status and subtypes in training datasets.

Characteristic	OR (95% CI)	*P*
Task a		
Sex	2.291 (1.329–3.949)	0.003
Peripheral emphysema	5.412 (1.292–22.669)	0.021
R-score	2.262 (1.660–4.415)	0.000
Age	0.988 (0.965–1.012)	0.329
Size	1.58 (0.768–1.458)	0.729
Margin	1.362 (0.769–2.413)	0.290
Shape	1.343 (0.742–2.431)	0.330
Pleural retraction	0.702 (0.392–1.257)	0.234
Bronchiole change	0.661 (0.328–1.329)	0.245
Lobulation	1.278 (0.726–2.250)	0.395
Spiculation	0.953 (0.489–1.860)	0.888
Peripheral fibrosis	1.166 (0.561–2.240)	0.681
Task b		
Age	1.588 (1.041–2.421)	0.032
R-score*	1.921 (1.283–2.875)	0.002

OR, odds ratio; CI, confidence interval; NA, not available.

TABLE 5 | Predictive performance of risk factors in two datasets regarding two tasks.

Characteristic	AUC	
	Training dataset	Validation dataset
Task a		
Sex	0.585#	0.570#
Peripheral Emphysema	0.528#	0.548#
R-score	0.708#	0.694
Nomogram	0.755	0.734
Task b		
Age	0.605#	0.656#
R-score*	0.684	0.708
Nomogram*	0.689	0.757

The constructed R-score and nomogram in task b. #Significantly difference between the performance of nomogram and other factors.

clinical factors, imaging semantic features, imaging radiomic features, and *EGFR* mutation subtypes were also investigated in our study. Age was the only independent factors that can be used to differentiate different *EGFR* mutation subtypes. No CT features were identified as potential predictors, which is consistent with previous studies (37, 38), indicating the difficult of predicting *EGFR* mutation subtypes through semantic features (relatively low-level). This conclusion encouraged us to investigate whether the radiomic features (relatively high-level) might be competent on this task. Surprisingly, the constructed R-score* was identified as an independent predictor and obtained a good performance with an AUC of 0.708. A radiomics-based nomogram*, incorporating age and R-score*, achieved a better performance (AUC = 0.757), indicating the efficiency of radiomic features in different medical tasks.

We may conclude that radiomics outperformed the clinical and semantic features in both tasks (see **Table 5**). The diagnostic benefits may due to the possession of more presentative and discriminative information of radiomics, which could reflect the tumor spatial heterogeneity, tumor microenvironment, as well as tumor gene patterns. Yet, completely interpreting the association between radiomics and the complex biological processes (EGFR mutation status in the current study) remains an intractable challenge. The interpretability warrants further investigation. Please note that one of the implications of our study was to introduce the nomogram, which was easily used in clinical practice and may facilitate the clinical transformation of radiomics researches.

There are several limitations should be noted. First, although the imaging normalization and reproductive analysis can mitigate the influence of radiomic feature variabilities, it cannot make the current study completely immune to the potential confounding variability caused by different CT scanning parameters (39, 40). Paradoxically, studies with homogeneous images must sacrifice the amount of data. How to balance it well is difficult for each radiomics-based research. Second, this was a single-center study and lacked of external validation, indicating the potential data selection bias. Conducting a multi-center study and validating the constructed model in an independent external dataset may not only improve the generalization and robustness of a model efficiently, the models can also substantially benefit from diversified data from different regions, races and countries. Third, the current study narrowly focused on *EGFR* mutation status and subtypes. Constructing a nomogram which could cover all clinical relevant *EGFR* mutation and even other genetic mutations (e.g., *ROS-1*, *ALK*) may be interesting and worth investigating in the future work.

CONCLUSION

Radiomics-based nomogram, incorporating clinical characteristics, CT features and radiomic features, can non-invasively and efficiently predict the *EGFR* mutation status and thus potentially fulfill the ultimate purpose of precision medicine. The methodology is a possible strategy to predict *EGFR* mutation subtypes, providing the support of clinical treatment scenario.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

This retrospective study was approved by the Institutional Review Board of our institution (No. 20170103), which waived the requirement for patients' informed consent referring to the CIOMS guideline.

AUTHOR CONTRIBUTIONS

ML, JL, and WZ: conception and design and provision of study materials or patients. ML, JL, YW, and YH: administrative support. ML, WZ, YS, PG, MT, WM, CL, and LJ: collection and assembly of data. ML, JL, WZ, and YX: data analysis and interpretation. All authors: manuscript writing and final approval of manuscript.

FUNDING

This study was supported by the National Key Research (2017YFC0112905), the Medical Imaging Key Program of Wise Information Technology of 120, Health Commission of Shanghai (Grant No. 2018ZHYL0103), the National Natural Science Foundation of China (Grant Nos. 61976238 and 81671671), the Future Star of famous doctors' training plan of Fudan University, the Shanghai Municipal Commission of Science and Technology (Grant No. 17411963600) and the Key R & D projects in Hunan Province (Grant No. 2019SK2131).

ACKNOWLEDGMENTS

The authors would like to thank Jianhong Cheng for assisting in data analysis.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2019.01485/full#supplementary-material

REFERENCES

1. 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. doi: 10.3322/caac.21492
2. Ettinger DS, Wood DE, Aisner DL, Akerley W, Bauman J, Chirieac LR, et al. Non-small cell lung cancer, Version 5.2017, NCCN clinical practice guidelines in oncology. *J Natl Compr Canc Netw*. (2017) 15:504–35. doi: 10.6004/jnccn.2017.0050
3. Lee SM, Lewanski CR, Counsell N, Ottensmeier C, Bates A, Patel N, et al. Randomized trial of erlotinib plus whole-brain radiotherapy for NSCLC patients with multiple brain metastases. *J Natl Cancer Inst*. (2014) 106:151. doi: 10.1093/jnci/dju151
4. Novello S. Epidermal growth factor receptor tyrosine kinase inhibitors as adjuvant therapy in completely resected non-small-cell lung cancer. *J Clin Oncol*. (2015) 33:3985–6. doi: 10.1200/JCO.2015.63.7587
5. Hirsch FR, Bunn PJ. EGFR testing in lung cancer is ready for prime time. *Lancet Oncol*. (2009) 10:432–3. doi: 10.1016/S1470-2045(09)70110-X
6. Ke EE, Zhou Q, Zhang QY, Su J, Chen ZH, Zhang XC, et al. A higher proportion of the EGFR T790M mutation may contribute to the better survival of patients with exon 19 deletions compared with those with L858R. *J Thorac Oncol*. (2017) 12:1368–75. doi: 10.1016/j.jtho.2017.05.018
7. Sacconi B, Anzidei M, Leonardi A, Boni F, Saba L, Scipione R, et al. Analysis of CT features and quantitative texture analysis in patients with lung adenocarcinoma: a correlation with EGFR mutations and survival rates. *Clin Radiol*. (2017) 72:443–50. doi: 10.1016/j.crad.2017.01.015
8. Hasegawa M, Sakai F, Ishikawa R, Kimura F, Ishida H, Kobayashi K. CT features of epidermal growth factor receptor-mutated adenocarcinoma of the lung: comparison with nonmutated adenocarcinoma. *J Thorac Oncol*. (2016) 11:819–26. doi: 10.1016/j.jtho.2016.02.010
9. Sabri A, Batool M, Xu Z, Bethune D, Abdolell M, Manos D. Predicting EGFR mutation status in lung cancer: proposal for a scoring model using imaging and demographic characteristics. *Eur Radiol*. (2016) 26:4141–7. doi: 10.1007/s00330-016-4252-3
10. Cheng Z, Shan F, Yang Y, Shi Y, Zhang Z. CT characteristics of non-small cell lung cancer with epidermal growth factor receptor mutation: a systematic review and meta-analysis. *BMC Med Imag*. (2017) 17:5. doi: 10.1186/s12880-016-0175-3
11. Aerts HJ, Grossmann P, Tan Y, Oxnard GG, Rizvi N, Schwartz LH, et al. Defining a radiomic response phenotype: a pilot study using targeted therapy in NSCLC. *Sci Rep*. (2016) 6:33860. doi: 10.1038/srep33860
12. Rios VE, Parmar C, Liu Y, Coroller TP, Cruz G, Stringfeld O, et al. Somatic mutations drive distinct imaging phenotypes in lung cancer. *Cancer Res*. (2017) 77:3922–30. doi: 10.1158/0008-5472.CAN-17-0122
13. Yip SS, Kim J, Coroller TP, Parmar C, Velazquez ER, Huynh E, et al. Associations between somatic mutations and metabolic imaging phenotypes in non-small cell lung cancer. *J Nucl Med*. (2017) 58:569–76. doi: 10.2967/jnumed.116.181826
14. Liu Y, Kim J, Balagurunathan Y, Li Q, Garcia AL, Stringfield O, et al. Radiomic features are associated with egfr mutation status in lung adenocarcinomas. *Clin Lung Cancer*. (2016) 17:441–8.e6. doi: 10.1016/j.cllc.2016.02.001
15. Digumarthy SR, Padole AM, Gullo RL, Sequist LV, Kalra MK. Can CT radiomic analysis in NSCLC predict histology and EGFR mutation status? *Medicine*. (2019) 98:e13963. doi: 10.1097/MD.0000000000013963
16. Forghani R, Savadjiev P, Chatterjee A, Muthukrishnan N, Reinhold C, Forghani B. Radiomics and artificial intelligence for biomarker and prediction model development in oncology. *Comput Struct Biotechnol J*. (2019) 17:995–1008. doi: 10.1016/j.csbj.2019.07.001
17. Liu Y, Kim J, Qu F, Liu S, Wang H, Balagurunathan Y, et al. CT features associated with epidermal growth factor receptor mutation status in patients with lung adenocarcinoma. *Radiology*. (2016) 280:271–80. doi: 10.1148/radiol.2016151455
18. Zhao W, Xu Y, Yang Z, Sun Y, Li C, Jin L, et al. Development and validation of a radiomics nomogram for identifying invasiveness of pulmonary adenocarcinomas appearing as subcentimeter ground-glass opacity nodules. *Eur J Radiol*. (2019) 112:161–8. doi: 10.1016/j.ejrad.2019.01.021
19. Sauerbrei W, Royston P, Binder H. Selection of important variables and determination of functional form for continuous predictors in multivariable model building. *Stat Med*. (2007) 26:5512–28. doi: 10.1002/sim.3148
20. Travis WD, Brambilla E, Noguchi M, Nicholson AG, Geisinger KR, Yatabe Y, et al. International association for the study of lung cancer/american thoracic society/european respiratory society international multidisciplinary classification of lung adenocarcinoma. *J Thorac Oncol*. (2011) 6:244–85. doi: 10.1097/JTO.0b013e318206a221
21. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. (1988) 44:837–45. doi: 10.2307/2531595
22. Singal G, Miller PG, Agarwala V, Li G, Kaushik G, Backenroth D, et al. Association of patient characteristics and tumor genomics with clinical outcomes among patients with non-small cell lung cancer using a clinicogenomic database. *JAMA*. (2019) 321:1391–99. doi: 10.1001/jama.2019.3241
23. Jia TY, Xiong JF, Li XY, Yu W, Xu ZY, Cai XW, et al. Identifying EGFR mutations in lung adenocarcinoma by noninvasive imaging using radiomics features and random forest modeling. *Eur Radiol*. (2019) 29:4742–50. doi: 10.1007/s00330-019-06024-y
24. Jemal A, Miller KD, Ma J, Siegel RL, Fedewa SA, Islami F, et al. Higher lung cancer incidence in young women than young men in the United States. *N Engl J Med*. (2018) 378:1999–2009. doi: 10.1056/NEJMoa1715907

25. Wu SG, Chang YL, Yu CJ, Yang PC, Shih JY. Lung adenocarcinoma patients of young age have lower EGFR mutation rate and poorer efficacy of EGFR tyrosine kinase inhibitors. *ERJ Open Res*. (2017) 3:00092-2016. doi: 10.1183/23120541.00092-2016
26. Rami-Porta R, Bolejack V, Crowley J, Ball D, Kim J, Lyons G, et al. The IASLC lung cancer staging project: proposals for the revisions of the T descriptors in the forthcoming eighth edition of the TNM classification for lung cancer. *J Thorac Oncol*. (2015) 10:990–1003. doi: 10.1097/JTO.0000000000000559
27. Yang X, Dong X, Wang J, Li W, Gu Z, Gao D, et al. Computed tomography-based radiomics signature: a potential indicator of epidermal growth factor receptor mutation in pulmonary adenocarcinoma appearing as a subsolid nodule. *Oncologist*. (2019) 24:e1156–64. doi: 10.1634/theoncologist.2018-0706
28. Huang YQ, Liang CH, He L, Tian J, Liang CS, Chen X, et al. Development and validation of a radiomics nomogram for preoperative prediction of lymph node metastasis in colorectal cancer. *J Clin Oncol*. (2016) 34:2157–64. doi: 10.1200/JCO.2015.65.9128
29. Tan Y, Zhang ST, Wei JW, Dong D, Wang XC, Yang GQ, et al. A radiomics nomogram may improve the prediction of IDH genotype for astrocytoma before surgery. *Eur Radiol*. (2019) 29:3325–37. doi: 10.1007/s00330-019-06056-4
30. Ji GW, Zhu FP, Zhang YD, Liu XS, Wu FY, Wang K, et al. A radiomics approach to predict lymph node metastasis and clinical outcome of intrahepatic cholangiocarcinoma. *Eur Radiol*. (2019) 29:3725–35. doi: 10.1007/s00330-019-06142-7
31. Girard N, Sima CS, Jackman DM, Sequist LV, Chen H, Yang JC, et al. Nomogram to predict the presence of EGFR activating mutation in lung adenocarcinoma. *Eur Respir J*. (2012) 39:366–72. doi: 10.1183/09031936.00010111
32. Balachandran VP, Gonen M, Smith JJ, DeMatteo RP. Nomograms in oncology: more than meets the eye. *Lancet Oncol*. (2015) 16:e173–80. doi: 10.1016/S1470-2045(14)71116-7
33. Lindeman NI, Cagle PT, Aisner DL, Arcila ME, Beasley MB, Bernicker EH, et al. Updated molecular testing guideline for the selection of lung cancer patients for treatment with targeted tyrosine kinase inhibitors: guideline from the college of American Pathologists, the International Association for the Study of Lung Cancer, and the Association for Molecular Pathology. *J Thorac Oncol*. (2018) 13:323–58. doi: 10.1016/j.jtho.2017.12.001
34. Lee CK, Wu YL, Ding PN, Lord SJ, Inoue A, Zhou C, et al. Impact of specific epidermal growth factor receptor (EGFR) mutations and clinical characteristics on outcomes after treatment with EGFR tyrosine kinase inhibitors versus chemotherapy in EGFR-mutant lung cancer: a meta-analysis. *J Clin Oncol*. (2015) 33:1958–65. doi: 10.1200/JCO.2014.58.1736
35. Sheng M, Wang F, Zhao Y, Li S, Wang X, Shou T, et al. Comparison of clinical outcomes of patients with non-small-cell lung cancer harbouring epidermal growth factor receptor exon 19 or exon 21 mutations after tyrosine kinase inhibitors treatment: a meta-analysis. *Eur J Clin Pharmacol*. (2016) 72:1–11. doi: 10.1007/s00228-015-1966-0
36. Kobayashi S, Boggon TJ, Dayaram T, Janne PA, Kocher O, Meyerson M, et al. EGFR mutation and resistance of non-small-cell lung cancer to gefitinib. *N Engl J Med*. (2005) 352:786–92. doi: 10.1056/NEJMoa 044238
37. Zhao FN, Zhao YQ, Han LZ, Xie YS, Liu Y, Ye ZX. Clinicoradiological features associated with epidermal growth factor receptor exon 19 and 21 mutation in lung adenocarcinoma. *Clin Radiol*. (2019) 74:80.e7–80.e17. doi: 10.1016/j.crad.2018.10.001
38. Hong SJ, Kim TJ, Choi YW, Park JS, Chung JH, Lee KW. Radiogenomic correlation in lung adenocarcinoma with epidermal growth factor receptor mutations: Imaging features and histological subtypes. *Eur Radiol*. (2016) 26:3660–8. doi: 10.1007/s00330-015-4196-z
39. Berenguer R, Pastor-Juan M, Canales-Vazquez J, Castro-Garcia M, Villas MV, Mansilla LF, et al. Radiomics of CT features may be nonreproducible and redundant: influence of CT acquisition parameters. *Radiology*. (2018) 288:407–15. doi: 10.1148/radiol.2018172361
40. Zhao W, Zhang W, Sun Y, Ye Y, Yang J, Chen W, et al. Convolution kernel and iterative reconstruction affect the diagnostic performance of radiomics and deep learning in lung adenocarcinoma pathological subtypes. *Thorac Cancer*. (2019) 10:1893–1903. doi: 10.1111/1759-7714. 13161

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

ORIGINAL RESEARCH
published: 12 February 2020
doi: 10.3389/fonc.2020.00142

Tissue-Plasma TMB Comparison and Plasma TMB Monitoring in Patients With Metastatic Non-small Cell Lung Cancer Receiving Immune Checkpoint Inhibitors

Alex Friedlaender[1†], *Thierry Nouspikel*[2†], *Yann Christinat*[3], *Liza Ho*[3], *Thomas McKee*[3] *and Alfredo Addeo*[1*]

[1] *Department of Oncology, University Hospital of Geneva (HUG), Geneva, Switzerland,* [2] *Service of Medical Genetics, Diagnostics Department, University Hospital of Geneva, Geneva, Switzerland,* [3] *Department of Pathology, University Hospital of Geneva, Geneva, Switzerland*

Immuno-oncology is an ever growing field that has seen important progress across the spectrum of cancers. Responses can be deep and durable. However, as only a minority of patients respond to checkpoint inhibition, predictive biomarkers are needed. Cancer is a genetic disease arising from the accumulation of somatic mutations in the DNA of affected cells. Tumor mutational burden (TMB), represents the number of somatic mutations in a tumor that form neoantigens, responsible for the immunogenicity of tumors. Randomized controlled trials have so far failed to show a survival benefit when stratifying patients by tissue TMB. TMB has also been evaluated in plasma (PTMB). PTMB is anticipated to represent the biology of the entire cancer, whereas obtaining tissue of an amenable primary or a metastatic lesion may be prone to sampling bias because of tumor heterogeneity. For this reason, we are evaluating the correlation between TMB and PTMB, and prospectively evaluating the impact of these biomarkers on clinical outcomes. We also discuss the technical difficulties inherent to performing and comparing these analyses. Furthermore, we evaluate the correlation between the evolution of PTMB during an immunotherapy treatment and response at 3 and 6 months, as we believe PTMB may be a dynamic biomarker. In this paper, we present results from the first 4 patients in this project.

Keywords: CtDNA, immune check inhibitor (ICI), biomarker, TMB, NSCLC

OPEN ACCESS

Edited by:
Umberto Malapelle,
University of Naples Federico II, Italy

Reviewed by:
Pasquale Pisapia,
University of Naples Federico II, Italy
Fabio Pagni,
University of Milano Bicocca, Italy

***Correspondence:**
Alfredo Addeo
alfdoc2@libero.it

[†] *These authors share first authorship*

Specialty section:
This article was submitted to Thoracic Oncology, a section of the journal Frontiers in Oncology

Received: *05 December 2019*
Accepted: *27 January 2020*
Published: *12 February 2020*

Citation:
Friedlaender A, Nouspikel T, Christinat Y, Ho L, McKee T and Addeo A (2020) Tissue-Plasma TMB Comparison and Plasma TMB Monitoring in Patients With Metastatic Non-small Cell Lung Cancer Receiving Immune Checkpoint Inhibitors. Front. Oncol. 10:142. doi: 10.3389/fonc.2020.00142

INTRODUCTION

Immuno-oncology is an ever growing field that has seen important progress across the spectrum of cancers. While results have changed the prognosis for certain types of tumors, including melanoma, lung, genito-urinary, and digestive cancers, only a minority of patients responds to treatment and reaps these benefits. Different biomarkers have been developed with the goal of predicting response to treatment (1). Yet, only one has reached the maturity needed for clinical utility, programmed death ligand-1 (PD-L1). PD-L1 expression, assessed on tumor cells and immune-cells, correlates with response rate and survival in non-small-cell lung cancer (NSCLC).

In front line management of advanced NSCLC, the Keynote 024 trial concluded that patients with tumors expressing ≥50% PD-L1 derive greater clinical outcomes with pembrolizumab, a PD-1 immune checkpoint inhibitor (ICPI), than with platinum-based chemotherapy, both in terms of response rate (44. vs. 27.8%) and overall survival (OS), with a 3 year OS of 43% (2). The Keynote 042 trial went on to show front-line ICPI activity in all PD-L1 subsets of NSCLC, though higher PD-L1 expression is predictive of response (3).

The 5 year overall survival data of the Keynote 001 trial, presented at the American Society for Clinical Oncology (ASCO) 2019 conference, revealed durable long-term responses to immune-checkpoint inhibition (ICPI), and an unprecedented 5-year OS of 13.4–29.6% (4–6). This is a drastic improvement over pre-ICPI 5 year survival rates, which were below 5% (7).

However, these results also underline that despite being exposed to potential adverse effects of therapy, the majority of patients do not derive a durable benefit from ICPI, even in the selected high PD-L1 population. This highlights the need for further predictive biomarkers for ICPI.

It is known that cancer is a genetic disease and that neoplastic transformation results from the accumulation of somatic mutations in the DNA of affected cells. The genetic modifications in tumors can include non-synonymous mutations comprising mainly missense mutations, as well as synonymous mutations, insertions or deletions, splice site mutations and copy number gains and losses. Tumor mutational burden (TMB), represents the number of somatic mutations in a tumor, but there is no consensus as to which mutations should be included in the calculation: some authors report all mutations (8, 9), others use only non-synonymous mutations (10), and yet others (11–13) consider only mutations that alter the sequence of a protein (i.e., missense and indels within exons). The former vary in prevalence between 0.01 mutations/megabase pair (Mbp) and 400 mutations/Mbp. In this paper, we elected to report miscoding mutations, i.e., mutations that yield the translation of novel peptide epitopes, since it is thought that neoantigens are responsible for the immunogenicity of the tumor by eliciting T cell responses. This is the basis for the hypothesis that higher TMB should allow for greater benefit from ICPI.

This hypothesis has some clinical data in its favor. For instance, tumors known, through DNA sequencing, to harbor multiple acquired mutations, such as microsatellite unstable tumors, melanoma and non-small-cell lung cancer, are those with the best response to ICPI (8). Furthermore, studies have shown improved response rates (RRs) and progression free survival (PFS) in patients with tumors deemed to have high tissue TMB (9, 11). Yet, there appears to be no correlation between OS with single-agent ICPI and TMB in NSCLC, while the predictive value of TMB in combined PD-1 blockade and anti- cytotoxic TIL antigen-4 (CTLA4) inhibition showed promising PFS data (14, 15). Unfortunately, the recently published Checkmate 227 trial has now negated these early results, showing no association between TMB and OS (12).

TMB has also been evaluated in plasma (PTMB). PTMB is anticipated to represent the biology of the entire cancer, whereas obtaining tissue of an amenable primary or a metastatic lesion may be prone to sampling bias because of tumor heterogeneity (16). However, while tissue samples can be microdissected to yield a high percentage of tumor cells, circulating cell free DNA (ccfDNA) can be more challenging to interpret in this regard. Identifying circulating tumor DNA (ctDNA), which can be present as a minute fraction of total ccfDNA, can be limiting for next-generation sequencing (NGS) analysis (17). In a recent publication comparing TMB and PTMB by whole exome sequencing in various cancer types, the sensitivity for mutation detection was 53% in ccfDNA. The lower sensitivity was, as expected, often associated with lower ctDNA levels. Meanwhile, the sensitivity of ctDNA NGS compared to WES was 92%, suggesting that NGS may be a valid surrogate detection method, like in tissue (13).

Important clinical questions remain to be answered: is there a correlation between NGS-based TMB and PTMB in NSCLC? Will one be a stronger predictor of response to immunotherapy?

For this reason, we are evaluating the correlation between TMB and PTMB, and prospectively evaluating the impact of these biomarkers on clinical outcomes. We also discuss the technical difficulties inherent to performing and comparing these analyses. Furthermore, we evaluate the correlation between the evolution of PTMB during an immunotherapy treatment and response at 3 and 6 months, as we believe PTMB may be a dynamic biomarker. In this paper, we present results from the first 4 patients in this project.

METHODS

Patients

We selected 4 consecutive stage IV NSCLC patients before the introduction of pembrolizumab, an anti-PD-1 checkpoint inhibitor. Patients and tumor characteristics were collected from medical records, pathology, and radiology reports.

Samples

Blood was drawn and collected in Streck tubes before treatment and after 3 months of treatment with pembrolizumab. Plasma was prepared by centrifugation 10 min at 1,600 × g, collected, spun again at 16,000 × g for 10 min, and stored at −80°C until used. The tissue sample was collected at diagnosis.

DNA Extraction

Cell-free DNA was prepared from 4 to 5 ml plasma with the MinElute ccfDNA kit (Qiagen) or the Cobas cfDNA kit (Roche) according to the manufacturer instructions. Tumor DNA was prepared from formalin-fixed paraffin-embedded (FFPE) samples using the QIAamp DNA FFPE Tissue kit (Qiagen). Germline DNA was extracted from whole blood with QIAamp DNA blood mini kit (Qiagen). DNA was stored at −20°C until used.

Sequencing

A custom 443-gene, 2 073 Mbases SureSelect-HS library (Agilent) was built from 10 ng ccfDNA or 50 ng genomic DNA. Paired-end sequencing, 2 × 150 nt, was performed on a NextSeq500 sequencer (Illumina).

TABLE 1 | Patient and tumor characteristics.

Patient	Age	Sex	Histology	PD-L1 (%)	TMB (mut/mbp)	Time 0 PTMB (mut/Mbp)	3 month PTMB (mut/Mbp)	ICPI	Line	3 month response	6 month response	TTF
PIT-063	74	Male	ADC	90	6.3	1.4	0.0	Pembro	1	Response	Response	14 (ongoing)
PIT-069	66	Male	ADC	75	8.7	2.9	0.5	Pembro	1	Response	Dissociated	15
PIT-075	66	Male	Squamous	10	3.4	5.8	9.6	Pembro	2	Response	/	14
PIT-077	67	Female	ADC	70	1.4	4.3	12.1	Pembro	1	Progression	Death	4

ADC, adenocarcinoma; PD-L1, programmed death ligand-1; TMB, tumor mutation burden; PTMB, plasma mutation burden; Pembro, pembrolizumab; ICPI, immune checkpoint-inhibitor; TTF, time to treatment failure; Mut/Mbp, mutations per megabase pair.

Analysis

Data from plasma samples were preprocessed with the AGeNT package (Agilent) for molecular tag deduplication, then variants were called with SiNVICT (18) retaining level 5 variants, LoFreq (19) v 2.1.2, single sample mode with parameter $a = 10–9$ for plasma samples, OutLyzer (20) v 2.0, "calling" command, with default parameters or SNVer (21) v 0.5.3 with default parameters. For PTMB calculations, only miscoding variants (exonic variants with the potential of modifying the protein sequence: missense, non-sense, and indels), with a frequency <30% and absent from whole blood DNA were counted, the total was normalized to the total size of the regions sequenced. Tumor data was analyzed with a combination of Strelka (22) v 2.9.6 and MuTect2 (23) v 4.1.0.0, only variants called by both callers, with frequency >2%, frequency in tumor >4-fold higher than in normal tissue and average base quality >20 were retained. For TMB calculations, only miscoding variants were considered, for tumor data in table I only mutations with a frequency >5% were counted. For **Figure 3**, tumor data was re-analyzed with SiNVICT, using the same parameters and same counting criteria as for PTMB, for the sake of comparison.

Outcomes

Response was evaluated radiologically using the immune RECIST criteria (24) and clinically. Progression was defined as radiologic progression or the appearance of new cancer related symptoms or death. Time to treatment failure was calculated from the start of immunotherapy to its interruption due to progression requiring subsequent systemic therapy.

RESULTS

Patients

Of the four included in this preliminary analysis, three had adenocarcinoma, one squamous cell carcinoma. Three patients were male, one female. No patients harbored any druggable driver mutations, analyzed by next-generation sequencing. Three patients had a high PD-L1 expression (above 50%). The age range spanned from 66 to 74 years old (**Table 1**).

Amount of Cell-Free DNA

We measured the amount of cell-free DNA recovered and normalized it to the amount of plasma processed (**Figure 1**). In most cases, cell-free DNA yield matched the values we routinely obtain with healthy controls, around 10 ng/ml plasma. Patient PIT-063 had significantly higher cell-free DNA levels at the time of the first blood draw (23 ng/ml plasma), but this remains within the physiological range, as it is well-known that cell-free DNA levels can vary considerably for a given individual, depending, for instance on physical exercise or even psychosocial stress (25). Patient PIT-77 displayed low amounts of cell-free DNA at the first blood draw, but very high levels at the second, with an almost 100-fold increase over a 3-month period. We assumed that, in this case, the increase was due to a higher amount of ctDNA, which was confirmed by mutation analysis (see below).

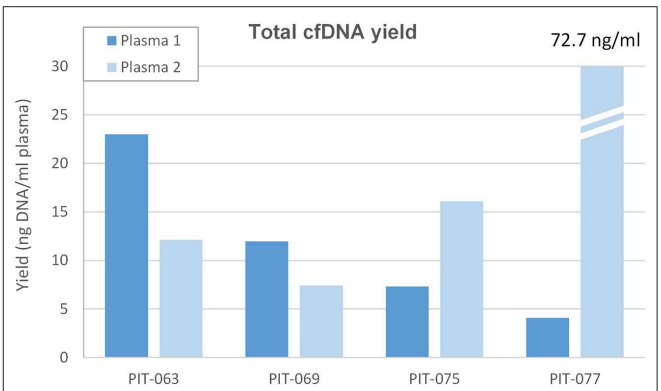


FIGURE 1 | Total cell-free DNA yield. Amount of DNA removed from the first (Plasma 1, concomitant to surgery) and second (Plasma 2, after 3 months therapy) plasma samples, expressed as ng ccfDNA per ml plasma.

Correlation Between Mutations Found in Tumor and Plasma

Bioinformatic identification of somatic mutations in sequencing data is a two-step process. During the initial, straightforward step, individual sequence reads are aligned to the human reference genome, and divergences from the reference (variants) are collated. If molecular barcodes are used, reads can be aggregated based on the original DNA molecule they were amplified from, allowing for some level of error polishing (26). The second and most critical step is to classify variants: some will be germline polymorphisms that are easy to recognize as variant frequencies will approach 50 or 100%, depending on whether the patient is heterozygous or homozygous. The main difficulty

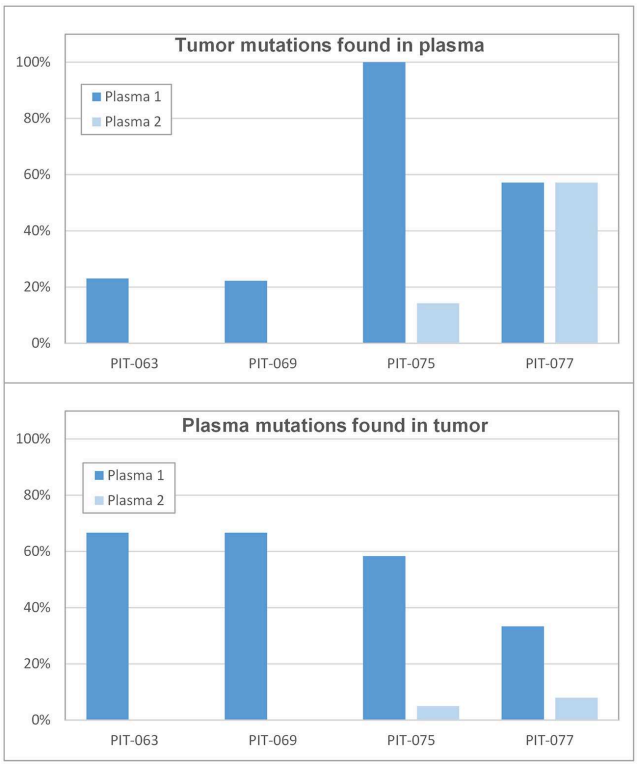


FIGURE 2 | Correlation between tumor and plasma samples (**Top**) percent of tumor mutations that were found in the first (time 0) and second (3 months) plasma samples. (**Bottom**) percent of plasmatic mutations that existed in the original tumor.

lies in the interpretation of low frequency variants, some of which correspond to genuine somatic mutations, while others are background noise, i.e., PCR errors or sequencing mistakes. This filtering process is particularly crucial in the context of mutation burden analysis which essentially tallies passenger mutations, most of which will have appeared in subclones of the tumor. As a result, mutation frequencies are expected to be low in the tumor and even lower in plasma, owing to dilution of ctDNA with ccfDNA from normal cells.

Numerous software packages have been proposed to sift somatic mutations from background noise (27). We extensively tested several, based on different underlying algorithms: SiNVICT (18) (Poisson model with additional heuristic filters), OutLyzer (20) (noise level estimation by recursive Thompson tau tests), SNVer (21) (allele frequency analysis by binomial-binomial model) and LoFreq (19) (allele frequency analysis by Poisson-binomial model). To evaluate the performance of these programs, we used the mutations identified in the tumor as a "truth set" and asked how many of these could be detected in plasma and identified as somatic mutation by a given variant caller.

Most mutations found in the tumor were also present in the corresponding plasma samples, albeit at low frequencies (<3%). As a result, none of the software packages we tested succeeded in calling a major fraction of these mutations. In our hands, the best programs were SiNVICT and LoFreq, the latter slightly more performant. For this reason, the data presented hereafter were produced with LoFreq, although analysis with SiNVICT led to identical conclusions.

The number of miscoding tumor mutations for patients PIT-063, PIT-069, PIT-075, and PIT-077 was 13, 18, 7, and 7, respectively. Of these, the fraction that LoFreq identified in plasma samples varied from zero to 100% (**Figure 2**, top panel). In particular, all tumor mutations were detected in the first plasma sample of PIT-075 and a large fraction in both samples of PIT-077, whereas for PIT-063 and PIT-069 few mutations were detected in the first plasma sample, and none in the second.

We also considered the problem from the opposite angle and asked how many of the variants called by the software did correspond to genuine tumor mutations. All software packages called an implausibly high number of variants (often several hundreds) and LoFreq was no exception. We attempted to identify somatic mutations among these by retaining only variants with a frequency lower than 30%, as variants that were more common than this were likely germline in nature. However, when we checked for the presence of these variants in leukocyte DNA, we found a large fraction of them, indicating that these were not genuine tumor mutations. We thus systematically sequenced leukocyte DNA in each patient and considered only low-frequency variants that were absent from leukocyte DNA as genuine somatic mutations. Among those identified in the first set of plasma samples, only 33–67% corresponded to mutations that had been found in the tumor (**Figure 2**, bottom panel). In the second set of plasma samples, a much lower fraction (0–8%) of presumably somatic mutations actually matched mutations found in the original tumor.

Mutation Burden and Mutation Frequency

We then calculated mutation burden by considering only miscoding mutations among those retained (i.e., called by LoFreq, with a frequency inferior to 30% and absent from leukocyte DNA), and normalizing this number for the size of the target region sequenced (**Figure 3**, top panel and **Table 1**). All 4 patients had TMB inferior to 10 mutations per megabase (Mbp); the corresponding mutation burden in the synchronous plasma samples was generally lower, inferior to 5 mutations per Mbp. The notable exception was patient PIT-075 who displayed a PTMB similar to, even slightly higher than his TMB. At the time of the second blood draw, after 3 months therapy, PTMB was zero for patient PIT-063 and very low for PIT-069, whereas it had notably increased for PIT-075, and reached almost 12 mutations per Mbp for PIT-077.

We also considered the impact of mutation frequencies (i.e., the percentage of mutant reads at a given genomic position) on mutation count, as low-frequency mutations, being harder to distinguish from background, are more likely to be missed. Among our 4 patients, PIT-063 had the lowest average mutation frequency in plasma, and also the lowest plasma mutation burden (**Figure 3**, bottom panel). Average mutation frequencies were slightly higher for PIT-069 and PIT-075, although still inferior to 3%. By contrast, patient PIT-077 showed low mutation frequencies in the first plasma sample but markedly elevated frequencies in the second, with an average of 14.5%. This phenomenon may be partly responsible for the higher

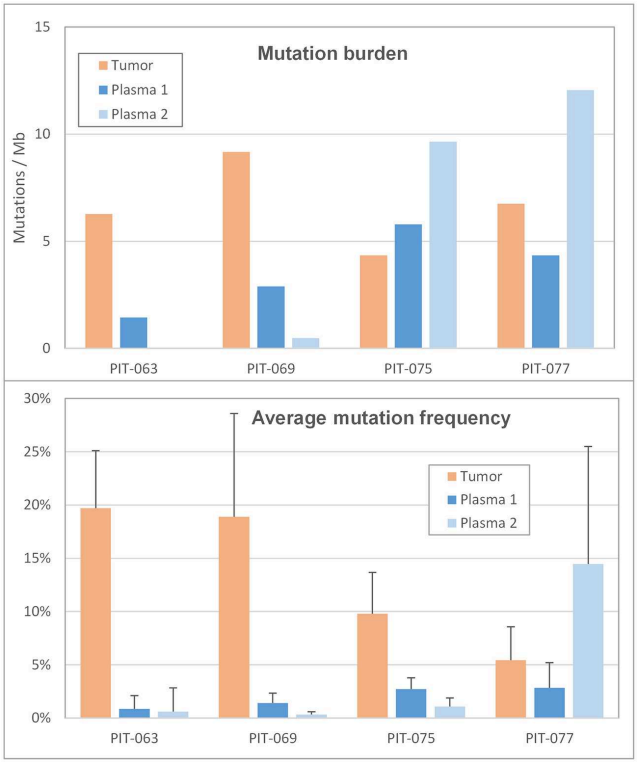


FIGURE 3 | Mutation profile (**Top**) mutation burden in the tumor and in the two plasma samples, expressed as the number of miscoding mutations per megabase. (**Bottom**) Average frequency of the alternate allele in the tumor and in the two plasma samples. Error bars are the standard deviation.

mutation burden we calculated for this sample, by making it easier for the software to distinguish somatic mutations from background errors.

DISCUSSION

Correlations With Clinical History

The 4 patients included in this pilot study illustrate different clinical and pathophysiological scenarios, in their initial situation and subsequent evolution.

Patient PIT-063 had a high TMB, yet the lowest initial PTMB. Most mutations identified in the tumor were present in plasma, but at very low frequencies (<1%), making it almost impossible for the software to distinguish them from background error noise. A compounding factor may be that this patient presented higher than average ccfDNA levels. Given the low mutation frequencies observed, it is likely that most of his circulating DNA did not originate from the tumor, and may just reflect a physiological fluctuation which, unfortunately, contributed to further dilute ctDNA and made mutation detection even more challenging. After 3 months of checkpoint inhibitor therapy, PTMB was down to zero. The clinical evolution was favorable, with a very good partial response at 3 months, maintained at 6 months.

This patient is a perfect illustration of the difficulties inherent to circulating DNA analysis: when a tumor does not release ctDNA into the blood stream, or does so in minute quantities, it is virtually impossible to detect tumor-born mutations, no matter how sophisticated the sequencing technique and how performant the analysis software. From a diagnostic/prognostic point of view, it is important to identify such situations, to distinguish them from real low-TMB cases. In other words, one needs a way to quantify the fraction of cell-free DNA that originates from the tumor. A possible method is to analyze the plasma for driver mutations identified in the tumor, e.g., epidermal growth factor receptor (*EGFR*) exon 21 p.L858R or Kirsten rat sarcoma (*KRAS*) exon 2 p.G12D. Such driver mutations are presumably an early event in tumorigenesis, thus they should be present in all tumor cells, and their frequency in plasma should thus fairly reflect the proportion of ctDNA. In cases when such a mutation was not identified in the tumor, or when no tumor tissue is available, another possibility would be to study cell-free DNA methylation. It has been shown that methylation markers at precise genomic locations can be used to determine the tissue of origin of circulating DNA (28) and in particular to identify ctDNA, as methylation is globally perturbed in cancer cells (29). This approach, however, has not yet been shown to be quantitative and could not be applied to our samples. We thus chose the former approach for PIT-063, using a *TP53* stop-gain mutation that was present at 33% in the tumor, but only at 0.9% in plasma, thereby confirming our hypothesis that very little ctDNA is present in this patient's ccfDNA.

Patient PIT-069 was similar to PIT-063, with low mutation frequencies and a PTMB largely inferior to his TMB. His PTMB had significantly decreased after 3 months of immunotherapy, as well as the average mutation frequency (the amount of ccfDNA was similar at both time points, ruling out a dilution effect). Yet, none of the mutations identified in this second plasma sample were found in the original tumor. This likely represents an example of clonal evolution within the tumor, with novel passenger mutations appearing in numerous subclones, and being detected at low frequencies in the plasma. The clinical evolution entailed a dissociated response, with a notable regression of some metastases at 3 and 6 months, yet the progression of others at 6 months. Treatment was continued, and ultimately, radiotherapy was used on the progressive lesions. The newly identified clones in the plasma may thus correspond to those involved in the progression of the refractory metastases.

Patient PIT-075 was the only one with an initial PTMB matching its TMB. Mutation frequencies in the plasma were the highest of the 4 patients, with average levels of ccfDNA, likely indicating that the tumor was releasing significant amounts of ctDNA in the blood flow. It should be noted that PIT-075 was the only patient enrolled to have squamous histotype. As such, the question arises as to whether the high ctDNA release and PTMB could be related to histology. While current data suggests similar PTMB between ADC and squamous histotypes, we did not find information comparing their ctDNA release (30).

After 3 months of therapy, PTMB had significantly increased, whereas the average mutation frequency had declined, thereby ruling out a detection bias. The clinical evolution mirrors

the decrease in mutation frequency, with a very good partial response. There was no 6-month radiological evaluation, as it was declined by the patient, but he remains asymptomatic at 15 months after treatment initiation.

In this patient, mutation burden analysis in the plasma was successful and matched that in the tumor. Both values were largely inferior to 10 mutations/Mb, the cut-off used in the Checkmate 227 trial (12) as a threshold for high tissue mutation burden, but the latter study counted all mutations in the coding sequence (including synonymous) whereas we only considered miscoding mutations. Counting synonymous mutations for this patient would result in higher values: 4.8 for TMB and 8.7 for PTMB, the latter approaching the 10 mutations/Mb threshold. In patients treated with combination immunotherapy (anti PD-1 and anti-CTLA4 antibodies), with tumors expressing PD-L1 on $< 1\%$ of cells, high TMB was associated with a 1 year progression-free survival rate of 45%, as opposed to 18% with low TMB. It is important to note that ultimately, this trial did not show any survival difference by TMB (11). A preplanned exploratory analysis from the MYSTIC trial showed an OS benefit in patients with PTMB $>$ 16 mutations/Mbp in front-line combined anti-PD-L1 and anti-CTLA4 immunotherapy for advanced NSCLC (31). However, the Neptune trial, which prospectively assessed the role of PTMB in this same setting with the same treatment failed to show any predictive relevance of PTMB in the primary analysis (32). There appears to be a correlation between mutation burden and response in our patient. While this may be a coincidence, the decrease in mutational frequency suggests a dynamic role of PTMB.

Patient PIT-077 had the lowest amount of ccfDNA and one of the highest average mutation frequencies among our 4 patients. It can be assumed that the tumor was releasing a fair amount of ctDNA at the time of the first blood sampling. Her initial PTMB was 4.3 mutations/Mbp, but after 3 months of immunotherapy, the molecular situation had significantly worsened, with a high PTMB (12.1 miscoding mutations/Mbp), high mutations frequencies (14.5% in average) and very high levels of ccfDNA (73 ng/ml plasma). This likely indicates disease progression, with a tumor that accumulated novel mutations (only 8% were present in the initial tumor), and released large amounts of ctDNA into the blood flow. The clinical evolution supports the biological hypothesis, with clear disease progression at 3 months and cancer-related mortality at 6 months.

This patient was illustrative of a situation in which liquid biopsy proved superior to a traditional tumor biopsy, in that it allowed us to non-invasively detect a drastic increase in PTMB and in average mutation frequency after 3 months, which were in line with the poor, eventually fatal evolution of the patient.

Lessons Learned

Although limited in the number of patients, this pilot study allowed us to draw important methodological conclusions pertaining to the analysis of mutation burden in the plasma.

First and foremost, the key to a valid analysis is reliable identification of genuine somatic mutations, originating from the tumor. This is a particularly difficult problem in the case of mutation burden analysis since, by definition, passenger mutations originate from distinct cells and are expected to be present only in small subclones within a tumor. As a result, one can expect mutation frequencies to be quite low in plasma, making it difficult to distinguish genuine mutations from PCR or sequencing errors.

There are several ways to address this issue. One is to reduce background noise via technical improvements to DNA sequencing or library synthesis. In this respect, the advent of molecular barcodes was a major step forward, as it allows the identification of sequencing mistakes and PCR errors occurring after the second PCR cycle. Yet, we found it insufficient in the case of mutation burden analysis. A second key factor is bioinformatic processing, and the development of highly efficient algorithms to discriminate mutations from background noise. Here also, despite the release of a number of specific software packages in the recent years, we did not find one that was fully efficient for the purpose of mutation burden analysis.

A critical point, in our experience, is the need to sequence leukocyte DNA. We observed that a simple mutation frequency filter is not sufficient to reliably identify somatic mutations. For one, because it is conceivable that a somatic mutation may reach a frequency approaching 50%, e.g., in advanced cases when a tumor is releasing large amounts of ctDNA. But more importantly, the bulk of cell-free DNA was shown to originate from hematopoietic cells (33) and it is known that even normal individuals can accumulate clonal mutations in various leucocyte lineages upon aging, a phenomenon known as clonal hematopoiesis (34). Although somatic, these mutations do not originate from the tumor and should not be tallied when calculating PTMB. The only way to exclude them from analysis it to perform two rounds of sequencing, one on plasma and one on leucocytes, obtained either from whole blood or from the buffy coat resulting from plasma preparation.

A second conclusion of this study is that mutation burden measurements in plasma rarely match those in the tumor. There could be several reasons for this: it may happen that a tumor releases very low amounts of ctDNA, as in the case of PIT-063, making it virtually impossible to identify tumor mutations in plasma. Furthermore, as the amount of ccfDNA originating from normal cells fluctuates rapidly, it may dilute ctDNA in various proportions depending when phlebotomy is performed. This is an important consideration to keep in mind if one intends to make use of PTMB for patient follow up over an extended period of time. A possible solution would be to perform several closely spaced blood draws (e.g., 3 or 4 in a day) and sequence them in parallel, or to select the one with the lowest ccfDNA yield for sequencing. This may, however, impose an extra burden on the patient and significantly increases the cost of the analysis.

Conversely, as many tumors (e.g., lung and kidney) are genetically heterogeneous, a surgical biopsy may not account for the whole collection of mutations a tumor comprises. It was shown that ctDNA analysis provides a more complete reflection of the mutational landscape than a surgical biopsy, both in the main tumor (35) and in eventual metastases (36). Thus, provided the above limitations are addressed, PTMB may prove a more reliable prognostic indicator than TMB.

Ultimately, whether mutation burden is predictive or not of response to ICPI treatments, the analytical complexity involved with the biomarker could limit its reproducibility and reliability. While mutation burden has potential, it currently does not deserve a role in therapeutic decision-making. Nonetheless, the potential value of PTMB during therapy, especially given the difficulty in interpreting radiological response to immunotherapy (37), remains to be better elucidated.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Local Ethical Committee in Geneva. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

AA, AF, TM, and TN designed the study. AA and AF provided the samples. LH performed sequencing. YC and TN performed bioinformatic analyses. AF and TN wrote the paper. All authors reviewed and approved the paper.

REFERENCES

1. Friedlaender A, Bauml J, Banna GL, Addeo A. Identifying successful biomarkers for patients with non-small-cell lung cancer. *Lung Cancer Manage*. (2019) 8:LMT17. doi: 10.2217/lmt-2019-0009
2. Reck M, Rodríguez-Abreu D, Robinson AG, Hui R, Csoszi T, Fülöp A, et al. Updated analysis of KEYNOTE-024: pembrolizumab versus platinum-based chemotherapy for advanced non-small-cell lung cancer with PD-L1 tumor proportion score of 50% or greater. *J Clin Oncol.* (2019) 37:537–46. doi: 10.1200/JCO.18.00149
3. Mok TSK, Wu Y-L, Kudaba I, Kowalski DM, Cho BC, Turna HZ, et al. Pembrolizumab versus chemotherapy for previously untreated, PD-L1-expressing, locally advanced or metastatic non-small-cell lung cancer (KEYNOTE-042): a randomised, open-label, controlled, phase 3 trial. *Lancet*. (2019) 393:1819–30. doi: 10.1016/S0140-6736(18)32409-7
4. Novello S, Milella M, Tiseo M, Banna G, Cortinovis D, Di Maio M, et al. Maintenance therapy in NSCLC: why? To whom? Which agent? *J Exp Clin Cancer Res*. (2011) 30:50. doi: 10.1186/1756-9966-30-50
5. Pabani A, Butts CA. Current landscape of immunotherapy for the treatment of metastatic non-small-cell lung cancer. *Curr Oncol*. (2018) 25(Suppl. 1):S94–102. doi: 10.3747/co.25.3750
6. Addeo A, Banna GL, Metro G, Di Maio M. Chemotherapy in combination with immune checkpoint inhibitors for the first-line treatment of patients with advanced non-small cell lung cancer: a systematic review and literature-based meta-analysis. *Front Oncol*. (2019) 9:264. doi: 10.3389/fonc.2019.00264
7. Howlader N, Noone AM, Krapcho M, Miller D, Brest A, Yu M, et al. *SEER Cancer Statistics Review, 1975–2016.* Available online at: https://seer.cancer.gov/csr/1975_2016 (accessed January 07, 2020).
8. Schrock AB, Ouyang C, Sandhu J, Sokol E, Jin D, Ross JS, et al. Tumor mutational burden is predictive of response to immune checkpoint inhibitors in MSI-high metastatic colorectal cancer. *Ann Oncol*. (2019) 30:1096–103. doi: 10.1093/annonc/mdz134
9. Samstein RM, Lee C-H, Shoushtari AN, Hellmann MD, Shen R, Janjigian YY, et al. Tumor mutational load predicts survival after immunotherapy across multiple cancer types. *Nat Genet*. (2019) 51:202–6. doi: 10.1038/s41588-018-0312-8
10. Zehir A, Benayed R, Shah RH, Syed A, Middha S, Kim HR, et al. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat Med.* (2017) 23:703–13. doi: 10.1038/nm.4333
11. Hellmann MD, Ciuleanu T-E, Pluzanski A, Lee JS, Otterson GA, Audigier-Valette C, et al. Nivolumab plus Ipilimumab in lung cancer with a high tumor mutational burden. *N Engl J Med.* (2018) 378:2093–104. doi: 10.1056/NEJMoa1801946
12. Hellmann MD, Paz-Ares L, Bernabe Caro R, Zurawski B, Kim S-W, Carcereny Costa E, et al. Nivolumab plus Ipilimumab in advanced non-small-cell lung cancer. *N Engl J Med.* (2019) 381:2020–31. doi: 10.1056/NEJMoa1910231
13. Koeppel F, Blanchard S, Jovelet C, Genin B, Marcaillou C, Martin E, et al. Whole exome sequencing for determination of tumor mutation load in liquid biopsy from advanced cancer patients. *PLoS ONE*. (2017) 12:e0188174. doi: 10.1371/journal.pone.0188174
14. Addeo A, Banna GL, Weiss GJ. Tumor mutation burden-from hopes to doubts. *JAMA Oncol*. (2019) 5:934–5. doi: 10.1001/jamaoncol.2019.0626
15. Rizvi NA, Hellmann MD, Snyder A, Kvistborg P, Makarov V, Havel JJ, et al. Cancer immunology. mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science*. (2015) 348:124–8. doi: 10.1126/science.aaa1348
16. Zhang Y, Chang L, Yang Y, Fang W, Guan Y, Wu A, et al. The correlations of tumor mutational burden among single-region tissue, multi-region tissues and blood in non-small cell lung cancer. *J Immunother Cancer*. (2019) 7:98. doi: 10.1186/s40425-019-0581-5
17. Fenizia F, Pasquale R, Roma C, Bergantino F, Iannaccone A, Normanno N. Measuring tumor mutation burden in non-small cell lung cancer: tissue versus liquid biopsy. *Transl Lung Cancer Res*. (2018) 7:668–77. doi: 10.21037/tlcr.2018.09.23
18. Kockan C, Hach F, Sarrafi I, Bell RH, McConeghy B, Beja K, et al. SiNVICT: ultra-sensitive detection of single nucleotide variants and indels in circulating tumour DNA. *Bioinformatics*. (2017) 33:26–34. doi: 10.1093/bioinformatics/btw536
19. Wilm A, Aw PPK, Bertrand D, Yeo GHT, Ong SH, Wong CH, et al. LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. *Nucleic Acids Res*. (2012) 40:11189–201. doi: 10.1093/nar/gks918
20. Muller E, Goardon N, Brault B, Rousselin A, Paimparay G, Legros A, et al. OutLyzer: software for extracting low-allele-frequency tumor mutations from sequencing background noise in clinical practice. *Oncotarget*. (2016) 7:79485–93. doi: 10.18632/oncotarget.13103
21. Wei Z, Wang W, Hu P, Lyon GJ, Hakonarson H. SNVer: a statistical tool for variant calling in analysis of pooled or individual next-generation sequencing data. *Nucleic Acids Res*. (2011) 39:e132. doi: 10.1093/nar/gkr599
22. Kim S, Scheffler K, Halpern AL, Bekritsky MA, Noh E, Källberg M, et al. Strelka2: fast and accurate calling of germline and somatic variants. *Nat Methods*. (2018) 15:591–4. doi: 10.1038/s41592-018-0051-x
23. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol*. (2013) 31:213–9. doi: 10.1038/nbt.2514
24. Schwartz LH, Litière S, de Vries E, Ford R, Gwyther S, Mandrekar S, et al. RECIST 1.1-Update and clarification: from the RECIST committee. *Eur J Cancer*. (2016) 62:132–7. doi: 10.1016/j.ejca.2016.03.081
25. Hummel EM, Hessas E, Müller S, Beiter T, Fisch M, Eibl A, et al. Cell-free DNA release under psychosocial and physical stress conditions. *Transl Psychiatr*. (2018) 8:236. doi: 10.1038/s41398-018-0264-x
26. Koessler T, Addeo A, Nouspikel T. Implementing circulating tumor DNA analysis in a clinical laboratory: a user manual. *Adv Clin Chem*. (2019) 89:131–88. doi: 10.1016/bs.acc.2018.12.004
27. Xu C. A review of somatic single nucleotide variant calling algorithms for next-generation sequencing data. *Comput Struct Biotechnol J*. (2018) 16:15–24. doi: 10.1016/j.csbj.2018.01.003

28. Sun K, Jiang P, Chan KCA, Wong J, Cheng YKY, Liang RHS, et al. Plasma DNA tissue mapping by genome-wide methylation sequencing for noninvasive prenatal, cancer, and transplantation assessments. *Proc Natl Acad Sci USA*. (2015) 112:E5503–12. doi: 10.1073/pnas.1508736112
29. Xu R-H, Wei W, Krawczyk M, Wang W, Luo H, Flagg K, et al. Circulating tumour DNA methylation markers for diagnosis and prognosis of hepatocellular carcinoma. *Nat Mater*. (2017) 16:1155–61. doi: 10.1038/nmat4997
30. Gandara DR, Paul SM, Kowanetz M, Schleifman E, Zou W, Li Y, et al. Blood-based tumor mutational burden as a predictor of clinical benefit in non-small-cell lung cancer patients treated with atezolizumab. *Nat Med*. (2018) 24:1441–8. doi: 10.1038/s41591-018-0134-3
31. Peters S, Cho BC, Reinmuth N, Lee KH, Luft A, Ahn M-J, et al. Abstract CT074: tumor mutational burden (TMB) as a biomarker of survival in metastatic non-small cell lung cancer (mNSCLC): blood and tissue TMB analysis from MYSTIC, a Phase III study of first-line durvalumab ± tremelimumab vs chemotherapy. *Cancer Res*. (2019) 79 (Suppl. 13):CT074. doi: 10.1158/1538-7445.AM2019-CT074
32. Update on the Phase III NEPTUNE trial of Imfinzi plus tremelimumab in Stage IV non-small cell lung cancer. Available online at: https://www.astrazeneca.com/media-centre/press-releases/2019/update-on-the-phase-iii-neptune-trial-of-imfinzi-plus-tremelimumab-in-stage-iv-non-small-cell-lung-cancer-21082019.html (cité́d November 20, 2019).
33. Lui YYN, Chik K-W, Chiu RWK, Ho C-Y, Lam CWK, Lo YMD. Predominant hematopoietic origin of cell-free DNA in plasma and serum after sex-mismatched bone marrow transplantation. *Clin Chem*. (2002) 48:421–7. doi: 10.1093/clinchem/48.3.421
34. Steensma DP. Clinical Implications of clonal hematopoiesis. *Mayo Clin Proc*. (2018) 93:1122–30. doi: 10.1016/j.mayocp.2018.04.002
35. Reinert T, Schøler LV, Thomsen R, Tobiasen H, Vang S, Nordentoft I, et al. Analysis of circulating tumour DNA to monitor disease burden following colorectal cancer surgery. *Gut*. (2016) 65:625–34. doi: 10.1136/gutjnl-2014-308859
36. De Mattos-Arruda L, Weigelt B, Cortes J, Won HH, Ng CKY, Nuciforo P, et al. Capturing intra-tumor genetic heterogeneity by *de novo* mutation profiling of circulating cell-free tumor DNA: a proof-of-principle. *Ann Oncol*. (2014) 25:1729–35. doi: 10.1093/annonc/mdu239
37. Solinas C, Porcu M, Hlavata Z, De Silva P, Puzzoni M, Willard-Gallo K, et al. Critical features and challenges associated with imaging in patients undergoing cancer immunotherapy. *Crit Rev Oncol Hematol*. (2017) 120:13–21. doi: 10.1016/j.critrevonc.2017.09.017

Conflict of Interest: AA has received research funding from Boehringer Ingelheim, and has received compensation from Bristol-Myers Squibb, AstraZeneca, Merck Sharpe & Dohme, Takeda, Pfizer, Roche and Boehringer Ingelheim for participating on advisory boards.

AF has received compensation from Roche, Pfizer, Astellas and Bristol-Myers Squibb for service as a consultant.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

ORIGINAL RESEARCH
published: 21 February 2020
doi: 10.3389/fonc.2020.00181

miR-1323 Promotes Cell Migration in Lung Adenocarcinoma by Targeting Cbl-b and Is an Early Prognostic Biomarker

Huan Zhao[1,2†], Chunlei Zheng[3,4†], Yizhe Wang[1], Kezuo Hou[3,4], Xianghong Yang[5], Yang Cheng[1], Xiaofang Che[3,4], Shilin Xie[1], Shuo Wang[3,4], Tieqiong Zhang[1], Jian Kang[6], Yunpeng Liu[3,4], Dianzhu Pan[2], Xiujuan Qu[3,4], Xuejun Hu[1‡] and Yibo Fan[3,4*‡]*

[1] Department of Respiratory and Infectious Disease of Geriatrics, The First Hospital of China Medical University, Shenyang, China, [2] Department of Respiratory, The First Affiliated Hospital of Jinzhou Medical University, Jinzhou, China, [3] Department of Medical Oncology, The First Hospital of China Medical University, Shenyang, China, [4] Key Laboratory of Anticancer Drugs and Biotherapy of Liaoning Province, The First Hospital of China Medical University, Shenyang, China, [5] Department of Pathology, Shengjing Hospital of China Medical University, Shenyang, China, [6] Department of Pulmonary Medicine, The First Hospital of China Medical University, Shenyang, China

OPEN ACCESS

Edited by:
Christian Rolfo,
University of Maryland Medical System, United States

Reviewed by:
Janaki Deepak,
University of Maryland, Baltimore, United States
Guilin Qiao,
University of Illinois at Chicago, United States

***Correspondence:**
Xuejun Hu
huxuejun2015@126.com
Yibo Fan
fany_bo@hotmail.com

[†] These authors share first authorship

[‡] These authors share senior authorship

Specialty section:
This article was submitted to Thoracic Oncology, a section of the journal Frontiers in Oncology

Received: 05 August 2019
Accepted: 03 February 2020
Published: 21 February 2020

Citation:
Zhao H, Zheng C, Wang Y, Hou K, Yang X, Cheng Y, Che X, Xie S, Wang S, Zhang T, Kang J, Liu Y, Pan D, Qu X, Hu X and Fan Y (2020) miR-1323 Promotes Cell Migration in Lung Adenocarcinoma by Targeting Cbl-b and Is an Early Prognostic Biomarker. Front. Oncol. 10:181. doi: 10.3389/fonc.2020.00181

Purpose: MicroRNAs are known to regulate cellular processes in non-small cell lung cancer (NSCLC) cells and predict prognosis. However, identification of specific microRNAs in NSCLC as potential therapeutic targets is controversial. We aim to determine the clinical significance of miR-1323 in the prognosis of patients with lung cancer and the potential mechanism.

Patients and methods: A bioinformatics approach was used to screen the importance microRNA in NSCLC through the online GEO database (GSE42425). The relationship between expression level of miR-1323 and overall survival of lung cancer patients was analyzed. Additionally, an independent corhort including 53 NSCLC cases that underwent resection validated the connection between miR-1323 and LUAD patients' overall survival. Next, the function of miR-1323 was studied *in vitro* by transient transfection. A more in-depth mechanism was studied through luciferase reporter gene experiments.

Results: High miR-1323 expression correlated with poor survival in NSCLC patients ($P = 0.011$), and in lung adenocarcinoma (LUAD) patients ($P = 0.015$) based on GEO database (GSE42425). In the independent cohort based on our hospital, high miR-1323 expression was associated with LUAD patients ($P = 0.025$). Moreover, transfection with mimics of miR-1323 showed an increased migratory capacity in LUAD A549 and HCC827 cells. In addition, E3 ubiquitin-protein ligase Casitas B-lineage Lymphoma-b (Cbl-b) was found to be the target genes of miR-1323 and significantly down regulated after mimics of miR-1323 transfection, and high Cbl-b expression predicted better prognosis in NSCLC and LUAD ($P = 0.00072$ and $P = 0.02$, respectively).

Conclusion: The miR-1323 promoted LUAD migration through inhibiting Cbl-b expression. High miR-1323 expression predicted poor prognosis in LUAD patients.

Keywords: lung adenocarcinoma, miR-1323, CBLB, prognosis, biomarker

INTRODUCTION

Non-small cell lung cancer (NSCLC) is a leading cause of cancer-related death and accounts for 80–85% of all lung cancer cases (1). The prognosis of patients with NSCLC is very poor and hence the majority of patients are diagnosed at an advanced stage owing to the lack of cancer-specific symptoms. Although epidermal growth factor receptor (EGFR) has emerged as a major target for NSCLC therapy, almost all patients on this treatment inevitably acquire drug resistance, but the prognosis of patients remains poor (2). Therefore, identification of new targets in lung cancer is urgently needed.

A previous study has shown that microRNAs are involved in multiple aspects of lung cancer such as cell proliferation, apoptosis, invasion, and EMT (3–6). Additionally, the survival of lung cancer patients has been predicted by a series of microRNAs such as miR-133b, miR-93-5p (7, 8). However, specific microRNAs playing a significant prognostic role in NSCLC are not determined.

Gene Expression Omnibus(GEO) Database (https://www.ncbi.nlm.nih.gov/gds/) is a gene expression database created and maintained by The National Center for Biotechnology Information (9). It contains high-throughput gene expression data submitted by research institutions around the world. Almost all gene expression detection data involved in the studies can be found in this database. Therefore, in this study, we used the online GEO Database to screen and determine the role of miR-1323 in predicting prognosis in lung cancer patients, which makes great significance for early screening and diagnosis. Previously study showed that knocking down of miR-1323 has been shown to restore radiosensitivity in radiation-resistance lung cancer cell (10). Besides, miR-1323 also suppressed the expression of PRKDC and enhanced DNA repair. However, whether miR-1323 is involved in regulating the prognosis of lung cancer patients is not known. We determined the expression of miR-1323 in lung cancer patients' samples from GEO datasets and frozen sections, and found that miR-1323 predicted poor survival in LUAD, whose mechanism might through promote the migration ability of lung cancer cells *in situ*. Furthermore, we found that the way for miR-1323 to regulate migration of LUAD was by targeting Cbl-b and high Cbl-b expression predicted better survival in NSCLC and LUAD patients.

MATERIALS AND METHODS

Patients and Tissue Collection

Primary LUAD tissues ($n = 53$) were obtained from NSCLC patients with permission of Shengjing Hospital of China Medical University (Shenyang, China) between December 2009 and 2010. All the patients underwent the surgery and were histomorphology confirmed. The current study was examined and approved by the Research Ethics Committee of Shengjing Hospital of China Medical University.

Bioinformatics Analysis

GEO Database was used to screen vital prognostic marker miRNA in early-stage lung adenocarcinoma. The search terms include following key words: [("lung neoplasms" OR ("lung neoplasms" OR "lung cancer")) AND ["mirnas" OR ("micrornas" OR "MicroRNA")] AND ("gene expression" OR "expression")] AND "stage." Suitable gene set was selected for subsequent analysis.

GEO2R, which is a online tool provided by official, allows users to compare two or more groups of Samples in a GEO Series and identify genes that are differentially expressed across experimental conditions. We divided all samples into good prognosis groups and poor prognosis groups according to median survival time and use GEO2R to analysis differentially expressed genes between two groups.

The corresponding clinical data were also acquired from the GEO database. The expression value of miRNAs were collected for each case and then divided into miR-1323 high expression and miR-1323 low expression groups. Correlation of miR-1323 expression level and clinicopathologic parameters was evaluated by Spearman assay. Kaplan-Meier analysis were performed to do the survival analysis. The relationship between miR-1323 and clinical stges of LUAD patients was analyzed by OncomiR database (11). MicroRNA target predictions were performed by online databases based on different prediction methods, including miRDB (12), miRWalk (13), Targetscan (14), starBase (15). The annotation of predicted gene symbols was done by DAVID Bioinformatics Resources 6.7 (16), Kaplan- Meier Plotter (17) was used to screen the target mRNA related with prognosis of LUAD.

RNA Isolation and Quantitative Real-Time PCR

Total RNA was purifed from formalin-fixed, paraffin-embedded tissue sections using miRNeasy FFPE Kit (Qiagen, USA) in accordance with manufacturer's protocol. RNA was quantified and purified at absorbance of 260/280 nm using NanoDrop spectrophotometer. cDNA was synthesized using the One Step PrimeScript@ miRNA (Takara, Naha, Japan). miR-1323 quantifcation was done using SYBR@ Premix Ex TaqTM II (Takara) Kit. qRT-PCR experiment was conducted in triplicates, normalized to U6 Small nuclear RNA and performed on the Applied Biosystems 7500 Thermocycler. Relative expression were calculated based on 2-ΔΔCt method.

Cells and Cell Culture

The human lung adenocarcinoma cell lines (A549 and HCC827) were obtained from ATCC. The cells were cultured in an humidified atmosphere with 95%air and 5%CO_2 at 37°C,and supplemented with RPMI-1640 medium (Gibco, ThermoFisher, Shanghai, China) with 10% heat-inactivated fetal bovine serum, streptomycin (100 U/mL) and penicillin (100 U/mL). A549 and HCC827 cells were split every 2–3 days at a concentration of 1.5*105 and cells/ml.

Transient Transfection

Before Wound healing assays, Transwell assays, WB assays, ELISA and Dual luciferase reporter assays, cells were transfected with plasmids, miR-1323 mimics, or Cbl-b siRNAs at indicated doses and times using jetPRIME® (Polyplus Transfection,

New York, USA) following the manufacturer's instructions. qRT-PCR and Western blot was performed to verify the transfection effciency.

Western Blot

Protein expression were assessed through Western blot analysis. Cell lysates were obtained using RIPA lysis [0.1%SDS, 1%Triton-100, 1 mmol EDTA (pH 8.0), 150 mmol/L NaCl, 10 mmol/L Tris-HCl (pH 7.5)] supplemented with protease inhibitors (100 μg/ml PMSF and 2 μg/ml Aprotitin). The protein level was measured by Coomassie brilliant blue method. Total protein extract was separated by SDS-PAGE and electrophoretically transferred onto a PVDF membrane. After blocking with 5% non-fat milk in TTBS buffer, membranes were incubated with antibodies [β-actin (1:1000, sc-47778, Santa Cruz Biotechnology), Cbl-b (1:250, sc-8006, Santa Cruz Biotechnology), Cyclin E (1:1000, sc-377100, Santa Cruz Biotechnology), Occludin (1:1000, 331500, Life Technologies)] overnight at 4°C. Membranes were then incubated for 30 min with secondary antibodies. Protein was detected using an ECL reagent in the Electrophoresis Gel Imaging analysis system.

Transwell

Transwell migration assays were performed to measure cell migratory ability. 1.0×10^4 cells were seeded in the top chamber of 24-well transwell plates, Medium supplemented with serum in the lower chamber was used to attract the cells in the top chamber which were suspended in medium without serum or growth factors. After migration at 37°C for 24 h, the non-migratory cells were removed by cotton wool. The migratory cells on the membrane were stained with 0.1% Giemsa stain solution for 2 h and the cells were quantified by counting the number of cells in randomly chosen 4 fields under microscope at 200-fold magnification.

Wound Healing Assay

Migration of A549 and HCC827 cells was investigated by wound healing assay. Viable cells were plated at 3×10^5 cells per well in six-well culture plates using growth media containing 10% FBS. After the cells were attached, cells were transfected with miR-1323 mimic or Cbl-b siRNAs for 48 h. As the cells reached semi-confluence, *in vitro* scratch wounds were created by scraping the cell monolayers with a 1 ml sterile pipette tip. After washing away suspended cells, the cells were treated with normal growth media containing lower concentration FBS, in order to eliminate the effect of proliferation on results. Photomicrograph was taken immediately (time 0 h) with an inverted microscope equipped with a digital camera, and the wounded cultures were allowed to grow for 24 h at 37°C. At this time, another photomicrograph was taken at the same position. Photos were randomly selected per hole for comparative analysis. Migration was quantified by counting cell numbers at the indicated distances from the wound edge. Data shown are representative of minimum three independent experiments.

ELISA

After treated A549 cells by transient transfection for 24 or 48 h, the culture supernatants were collected and the number of cells was counted. The levels of interleukin-6 (IL-6) were analyzed by human IL-6 ELISA kit (Biolegend, San Diego,USA.) according to the manufacturer's protocol and the optical density (OD) was measured at 450 nm with a microplate reader. Finally, IL-6 level was adjusted by the total number of cells.

Dual Luciferase Reporter Assay

Dual luciferase reporter assay was performed according to our previously study (18). We initially obtained the 3′-UTR sequence of Cbl-b was obtained through gene synthesis (OriGene, Rockville, MD, USA), and cloned into the vector pMirTarget through two restriction enzyme cutting sites (SgfI-MluI), resulting in the generation of SC209114. All reagents and methods are provided by OriGene Technologies (OriGene, Rockville, MD, USA). The sequencing results were compared with the standard template sequences of the BLAST software on the PUBMED and CHROMAS software to identify the gene mutation loci. In order to generate the Cbl-b mutant reporter, the seed region was mutated to remove all complementary nucleotides to miR-1323. A549 cells were co-transfected with firefly luciferase reporter plasmids (0.5 μg), pRL-TK luciferase control vector (0.005 μg) and miR-1323 or NC (50 nmol) in the 24-well plates. Luciferase assays were performed 24 h after transfection, using the dual-luciferase reporter assay system (Promega, Madison, WI, USA) according to the manufacturer's protocol (18).

Statistical Analysis and Graphing

Kaplan-Meier was performed to do survival analysis. X-tile program was used to choose the optimal cutoff value. Sperarman correlation analysis and Student's *t*-test and was used to analyze the comparisons between two miR-1323 expression levels and clinicopathological characteristics of patients. $P < 0.05$ was considered to be statistically significant. Statistical analyses were performed using SPSS 18.0 software (SPSS, Inc., Chicago, IL) and GraphPad Prism 7 (GraphPad, La Jolla, CA) were used for statistical analysis.

RESULTS

miR-1323 Was High Expression in Early NSCLC With Poor Prognosis

Previous studies on NSCLC implied the prognostic predictive value of microRNAs, and microRNA targeting showed great potential as a novel therapeutic strategy for NSCLC. In order to explore the important roles of microRNAs playing in early NSCLC, we initially searched for a group of data from the GEO database (GSE42425), which included 71 NSCLC lung tissues of stage I, and compared their microRNA expressions.

As shown in **Figure 1**, 363 miRNAs were found to be differently expressed in patients with short overall survival than in those with long overall survival according to the criteria of fold change >1.2. Among these miRNAs, 168 showed higher expression in NSCLC patients with poor prognosis, and 79

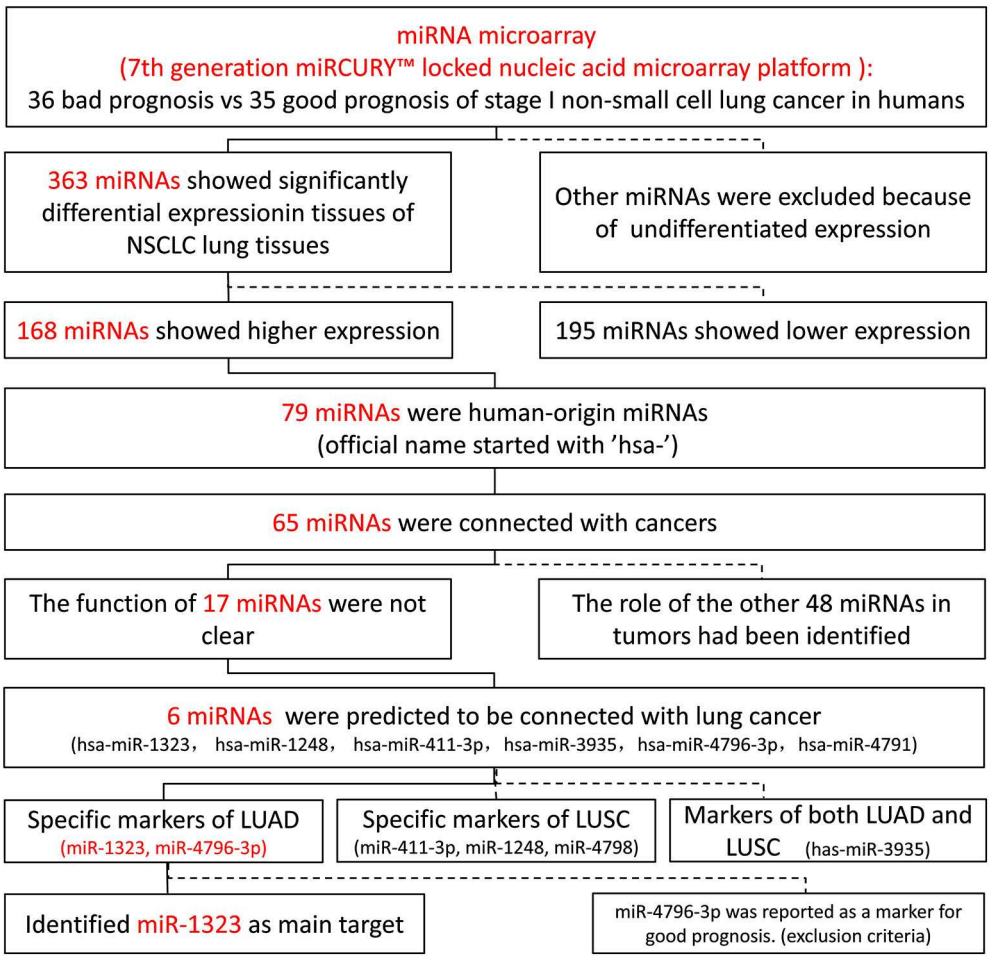


FIGURE 1 | Screening process of vital microRNA of early NSCLC.

miRNAs were chosen as they were human-origin miRNA. Afterwards, we searched the literature and screened out 65 miRNAs that had been identified their potential crucial roles in tumors. Since we hoped to find the key miRNA that could serve as clinical markers or targets for early diagnosis in NSCLC, 48 miRNAs were excluded owing to the weaker relationship with cancer. Among the remaining 17 miRNAs, only six were reported to be associated with lung cancer. During the process of analysis of these six miRNAs, we surprisingly found that they caused different effects to prognosis according to varies of pathological types. Two miRNAs (miR-1323, miR-4796) might be the alternative marker for LUAD (**Figures 2A,B**), while one miRNA (miR-3935, **Figure 2C**) is pathological type non-specific indication andthree miRNAs (miR-411-3p, miR-1248, miR-4791) have a clear indication for prognosis of lung squamous cell carcinoma (LUSC, **Figures 2D–F**). Since miR-4796 was reported that its low expression was associated with primary resistance of EGFR-tyrosine kinase inhibitors (EGFR-TKIs), which is the first-line treatment of LUAD. Furthermore, little literature had concerned about it. Thus, we chose miR-1323 for further experiments. The characteristics of the patients in GSE42425 are shown in **Table 1**. We found that there was no significant association between the expression levels of miR-1323 and other clinicopathological parameters except histological type, which predicts that miR-1323 has important prognostic significance in lung adenocarcinoma.

Validation the Effect of miR-1323 in an Independent Cohort of LUAD Patients

In order to verify the prognosis predictive function of miR-1323 expression in LUAD, we selected 53 LUAD cases that underwent resection at Shengjing Hospital of China Medical University (**Table 2**). The cohort, analyzed for the effect of prognosis, included 53 patients (30 men, 23 women) with a median age at surgery of 59 years (range, 35–75 years). Of these, 37 (69.8%) patients were at clinical stage I or II, and 16 (30.2%) were at stage III. The median survival duration was 57 months. Using surgically resected and paraffin sections from these cases, we performed qRT-PCR, and found that the expression level of miR-1323 in the paraffin sections ranged from 0.000085 to 0.056. Using the X-tile program, we chose an optimal cut-off value to divide the patients into two groups (miR-1323 low expression

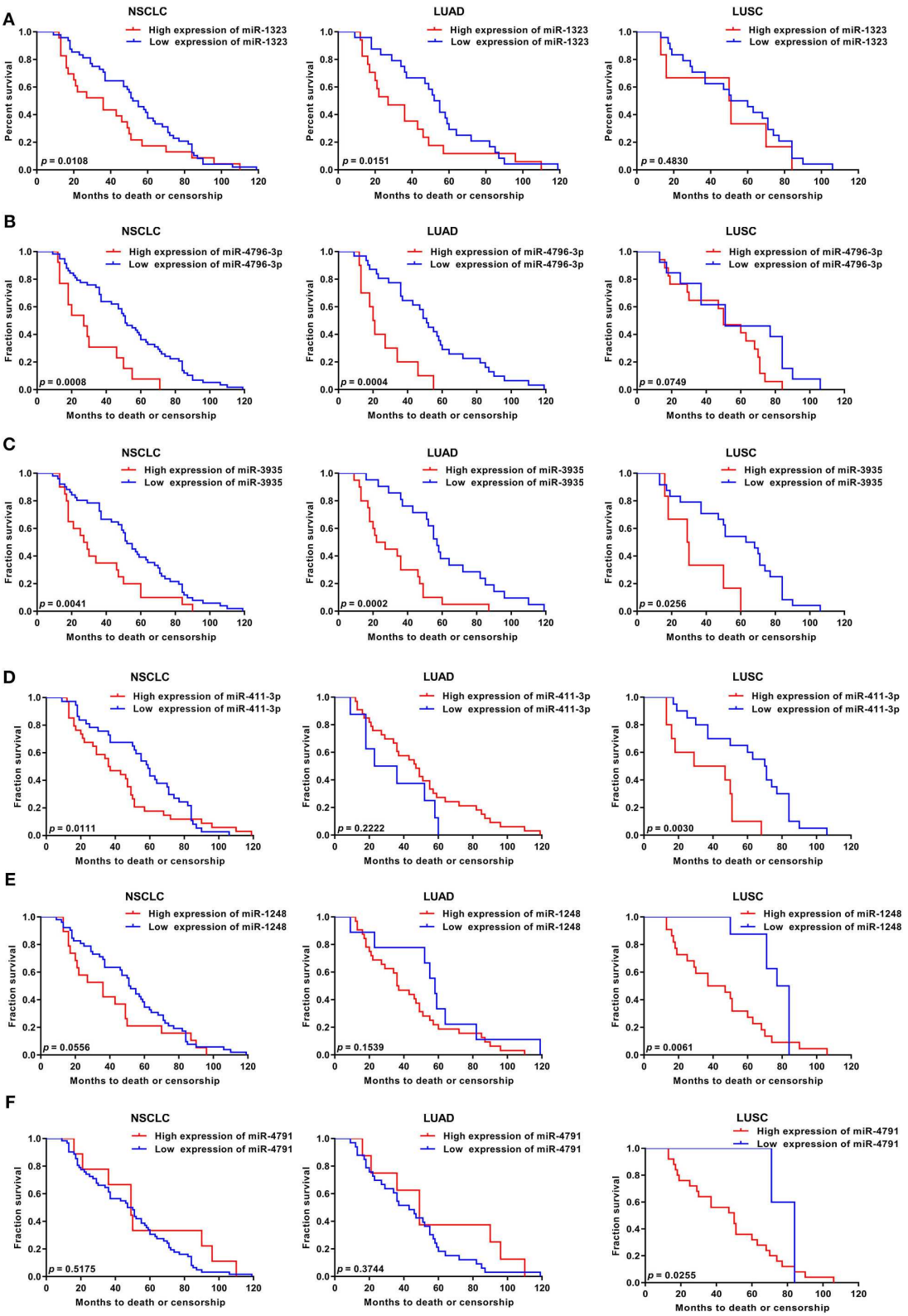


FIGURE 2 | KM survival curve and log-rank test for patients with six miRNAs high xpression in NSCLC, LUAD, and LUSC using GEO database. Using GSE42425 from GEO database, the expression of **(A)** miR-1323, **(B)** miR-4796-3p, **(C)** miR-3935, **(D)** miR-411-3p, **(E)** miR-1248, **(F)** miR-4791, and with OS in NSCLC, LUAD, and LUSC.

TABLE 1 | Characteristics of patients in GSE42425.

Characteristic	Number	miR-1323 expression		*P*-value
		High	Low	
Age, years	68 (46.0–83.0)			0.614
Median (range)				
≤68.0	36 (50.7)	23 (47.9)	13 (56.5)	
>68.0	35 (49.3)	25 (52.1)	10 (43.5)	
Gender, no. (%)				0.607
Male	34 (47.9)	24 (50.0)	10 (43.5)	
Female	37 (52.1)	24 (50.0)	13 (56.5)	
Histological type, no (%)				0.056
LUAD	41 (40.6)	24 (50.0)	17 (73.9)	
LSCC	30 (39.1)	24 (50.0)	6 (26.1)	
Race, no (%)				0.677
White	65 (91.5)	44 (91.7)	21 (91.3)	
Black	4 (6.3)	2 (4.3)	2 (8.7)	
Asian	1 (1.6)	1 (2.0)	0 (0.0)	
American indian	1 (1.6)	1 (2.0)	0 (0.0)	
History of smoking, no (%)				0.929
Never	8 (11.3)	5 (10.4)	3 (13.0)	
Former	43 (60.6)	29 (60.4)	14 (60.8)	
Current	20 (28.2)	14 (29.2)	6 (26.1)	
Stage, no (%)				0.613
IA	38 (53.5)	27 (56.3)	11 (47.8)	
IB	33 (46.5)	21 (43.7)	12 (52.2)	
Status of recurrence of cancer				0.405
No	39 (54.9)	28 (58.3)	11 (47.8)	
Yes	32 (45.1)	20 (41.7)	12 (52.2)	

LUAD, lung adenocarcinoma; LUSC, Lung squamous cell carcinoma; BAC, bronchioloalveolar carcinoma; AS, adenosquamous carcinoma.

TABLE 2 | Correlation of clinical features of NSCLC samples with miR-1323 expression levels of NSCLC cases.

Characteristic	Number	miR-1323 expression		*P*
		High expression	Low expression	
Age, years, no. (%)				0.884
Median (range)	59.0 (35.0–75.0)			
≤59.0	27 (50.9)	13 (52.0)	14 (50.0)	
>59.0	26 (49.1)	12 (48.0)	14 (50.0)	
Gender, no. (%)				0.033*
Male	30 (56.6)	12 (42.9)	18 (72.0)	
Female	23 (43.4)	16 (57.1)	7 (28.0)	
EGFR mutant state, no. (%)				0.674
19-Del	15 (28.3)	7 (25.0)	8 (32.0)	
L858R	18 (34.0)	11 (39.3)	7 (28.0)	
Unknown	20 (37.7)	10 (35.7)	10 (40.0)	
Depth of invasion, no (%)				0.471
T1+T2	47 (88.7)	24 (85.7)	23 (92.0)	
T3+T4	6 (11.3)	4 (14.3)	2 (8.0)	
Lymph node status, no (%)				0.442
N0–1	31 (58.5)	15 (53.6)	16 (64.0)	
N2–3	22 (41.5)	13 (46.4)	9 (36.0)	
TNM Stage, no (%)				0.354
I+II	37 (69.8)	18 (64.3)	19 (76.0)	
III	16 (30.2)	10 (35.7)	6 (24.0)	

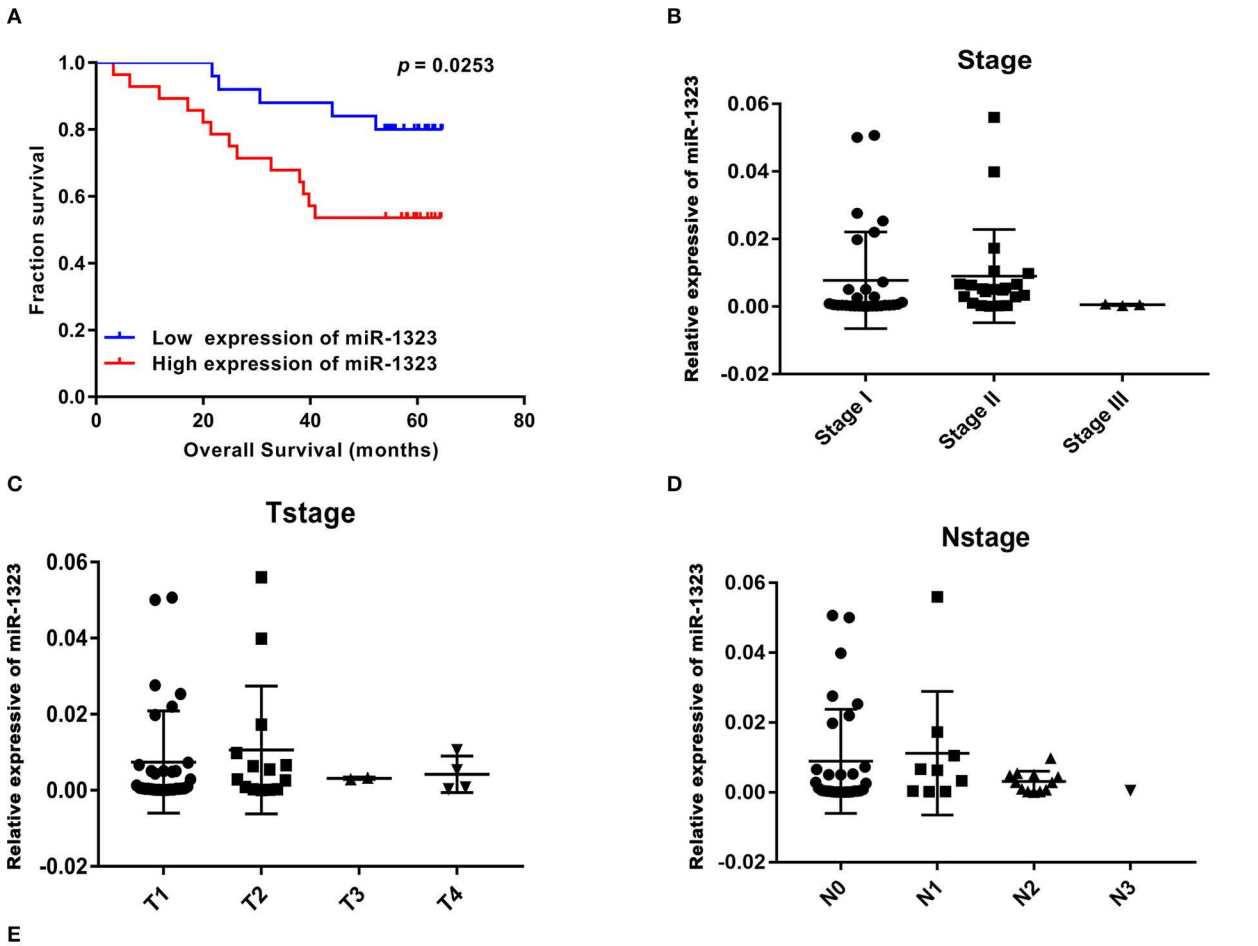
***$p < 0.05$.*
LUAD, lung adenocarcinoma; LUSC, Lung squamous cell carcinoma; BAC, bronchioloalveolar carcinoma; AS, adenosquamous carcinoma.

and miR-1323 high expression). Kaplan Meier survival analysis results showed high miR-1323 expression was found to indicate poor prognosis in LUAD ($P = 0.0253$, **Figure 3A**). In detail, higher expression of miR-1323 was associated with shorter overall survival (OS) in NSCLC and LUAD. These suggest that miR-1323 would like to be used as a specific early prognostic marker for LUAD to guide clinical treatment.

Cox univariate and multivariate regression analysis was performed on the prognostic value of pathological parameters such as age, gender, T stage, N stage, M stage, and miR-1323 expression using the above data. Univariate analysis showed that T-stage, N-stage, and miR-1323 expression levels were prognostic factors for lung adenocarcinoma ($P < 0.1$, **Table 3**). Further applying of Cox regression model, further multivariate analysis of factors with $P < 0.1$ in univariate analysis showed that N-stage and miR-1323 expression in tumor patients were independent prognostic factors for lung adenocarcinoma. These results suggest that miR-1323 is an independent risk factor for the prognosis of lung adenocarcinoma.These data further vertified the important role of mir-1323 in guiding prognosis for LUAD patients.

miR-1323 Promotes Migration of A549 and HCC827 Cells

Considering that mir-1323 has such a conductive effect on the prognosis of LUAD patients, we wonder that how it participates in regulating biological behavior and if it could be a novel target for clinical treatment. OncomiR, an online resource for exploring miRNA dysregulation in cancer, indicated that mir-1323 expression was correlated with pathologic M Status and N status (**Figure 3E**). Considering the Kaplan–Meier survival analysis showed that LUAD patients with high miR-1323 expression had shorter overall survival shown in **Figure 2**, we speculated higher miR-1323 expression might imply higher migration ability for LUAD. In order to investigate the effect of miR-1323 on LUAD, lung cancer cell lines A549 and HCC827 were transfected with an miR-1323 mimics or a negative control (NC) for 24 h (**Figure 4A**). The experimental results of both transwell assays and wound healing assays showed that miR-1323 significantly increased the migration of A549 cells and HCC827 cells post-transfection compared to the negative control groups (**Figures 4B–E**), which indicated the way of miR-1323 leading poor prognosis of LUAD patients was by promoting tumor cell metastasis.



miRNA Name	Cancer Abbreviation	Clinical Parameter	ANOVA P-value	ANOVA FDR	Multivariate Log Rank P-value	Multivariate Log Rank FDR
hsa-miR-1323	LUAD	Pathologic M Status	2.92e-04	1.09e-02	2.55e-01	6.56e-01
hsa-miR-1323	LUAD	Pathologic N Status	7.85e-01	9.99e-01	6.17e-02	4.20e-01
hsa-miR-1323	LUAD	Pathologic Stage	1.62e-03	5.74e-02	4.68e-01	8.57e-01

FIGURE 3 | KM survival curve and pathological stage for miR-1323 expression in LUAD cases. **(A)** KM survival curve of OS for 54 LUAD cases. **(B)** Pathological stage, **(C)** T stage, **(D)** N stage of 54 LUAD cases. **(E)** The relation between clinical parameters in LUAD and hsa-miR-1323 by ONCOMIR.

TABLE 3 | Univariate and multivariate analysis of the influence of clinicopathological features on prognosis in lung adenocarcinoma.

Variables	Univariate analysis			Multivariate analysis		
	HR	95% CI	*P*	HR	95% CI	*P*
Age (years)	0.985	0.941–1.032	0.530			
Gender (male vs. female)	1.548	0.580–4.126	0.383			
T staging	2.560	0.840–7.803	0.098*			
N staging	2.595	1.005–6.701	0.049**	2.541	0.983–6.567	0.054*
EGFR mutant state	1.044	0.732–1.489	0.811			
mir-1323	2.937	1.045–8.255	0.041**	2.881	1.024–8.108	0.045**

*$p < 0.05$, *$p < 0.1$.

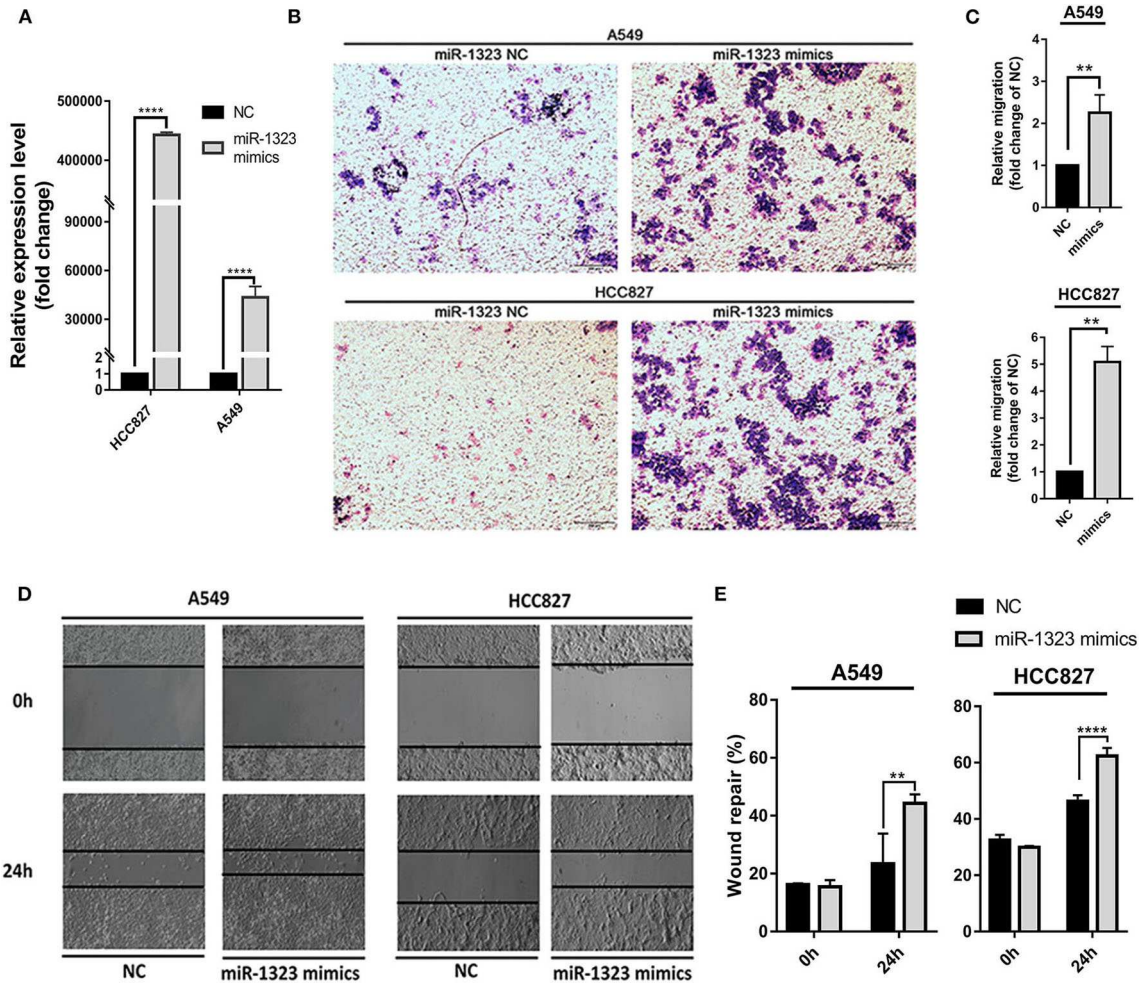


FIGURE 4 | Overexpression of miR-1323 significantly promotes the migration of A549 cells and HCC827 cells. **(A)** A549 cells and HCC827 cells were transfected with miR-1323 mimics or NC. **(B)** qRT-PCR was used to confirm the overexpression efficiency of miR-1323 mimics. **(C,D)** Transwell assays and **(E)** wound healing assays was to detect the migration of A549 cells and HCC827 cells. *t*-test was used to assess statistically significant differences between groups.Mean ± SD, results of three independent experiments, $^{**}p < 0.01$, $^{****}p < 0.0001$.

Cbl-b Is the Target of miR-1323 in NSCLC and Related With Patients' Prognosis

To examine the mechanisms through which miR-1323 exerts its lung cancer-promoting effects, we used bioinformatic algorithms (**Figure 5A**) and identified 251 potential target genes of miR-1323, among which only four genes are common tumor-associated genes related with prognosis of patients with lung adenocarcinoma in the Kaplan- Meier Plotter database ($P < 0.05$). After miR-1323 overexpression in A549 cells, only the protein expression level of Cbl-b was obviously suppressed among these candidates (**Figure 5B**). The level of IL-6 was generally measured by ELISA (**Figure 5C**). Since the expression level of IL-6 was negatively correlated with miR-1323 mimics in a time-dependent manner in A549 cells, which is inconsistent with the fact that IL-6 promotes inflammation and promotes cancer as it was reported, it wouldn't be considered as the main target gene of miR-1323 in the development of lung adenocarcinoma. Additionally, using computational algorithms, we identified that the 3′UTR of Cbl-b mRNA contains a complementary binding site for the miR-1323 seed region (**Figure 5D**). We also performed 3′UTR luciferase reporter assays to validate whether miR-1323 directly targets Cbl-b. Cotransfection of miR-1323 with the wild type 3′UTR of Cbl-b significantly repressed the relative luciferase ($P < 0.05$), and cotransfection of miR-1323 with the mutated 3′UTR of Cbl-b has no effects ($P = 0.100$), suggesting that miR-1323 targets Cbl-b directly (**Figure 5E**). These results indicate that Cbl-b is a target gene of miR-1323.

miR-1323 Promotes the Migration of NSCLC Cells Through Inhibiting Cbl-b

To investigate the detailed relationship between miR-1323 and Cbl-b and the effect of Cbl-b on the migration in NSCLC cells, we further transfect Cbl-b siRNA and a negtive control in A549

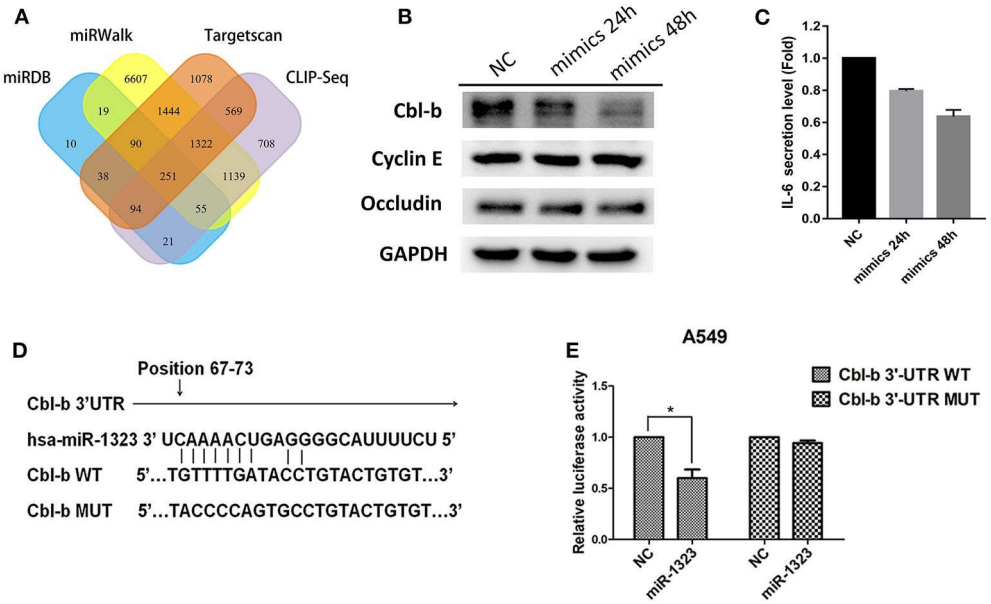


FIGURE 5 | miR-1323 can promotes the migration of A549 cells through inhibiting Cbl-b. **(A)** A schematic view of miR-1323 predicted target genes through four miRNA databases of different mechanisms. The A549 cells were transfected with miR-1323 mimic or NC, **(B)** WB detected the expression of proteins and **(C)** Elisa detected the secretory level of IL-6 in A549 cells. **(D)** The bindng site of miR-1323 to Cbl-b 3′UTR. **(E)** The A549 cells were co-transfected with pMirTarget-Cbl-b WT (or MUT) plasmid, miR-1323 mimic (or NC), and pRL-TK. The dual luciferase reporter assay detected the activity of luciferase. The histogram shows the relative activity of luciferase. *t*-test was used to assess statistically signifificant differences between groups.Mean ± SD, results of three independent experiments, $^{*}p < 0.05$.

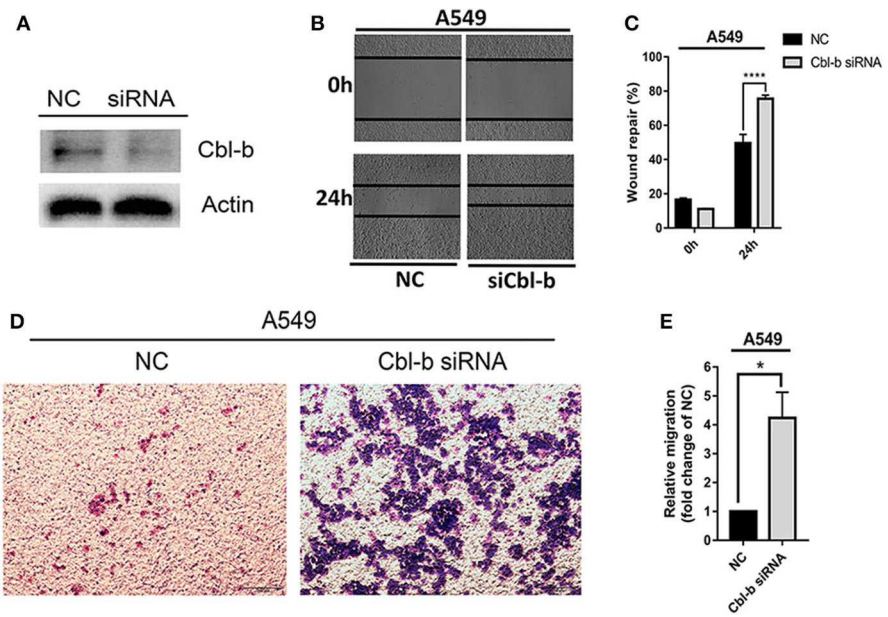


FIGURE 6 | Decreasing the expression of Cbl-b increases the migration ability of A549 cells. A549 cells were transfected with Cbl-b siRNA or NC, **(A)** WB detected the expression of Cbl-b in A549 cells. **(B,C)** Wound healing assays and **(D,E)** transwell assays showed the migration ability of A549 cells. *t*-test was used to assess statistically signifificant differences between groups.Mean ± SD, results of three independent experiments, $^{*}p < 0.05$, $^{****}p < 0.0001$.

and HCC827 cells. The transfection efficiency was ensured by western blot analysis (**Figure 6A**). Treatment with Cbl-b siRNA significantly promoted the cell migration when compared to that with the negtive control in both A549 cells (**Figures 6B–E**). These results emphasize miR-1323 inhibits the expression of Cbl-b and promotes the metastasis of tumor cells, leading to poor prognosis in patients with early LUAD. Besides, we used the Kaplan-Meier plotter and found that Cbl-b overexpression was associated with

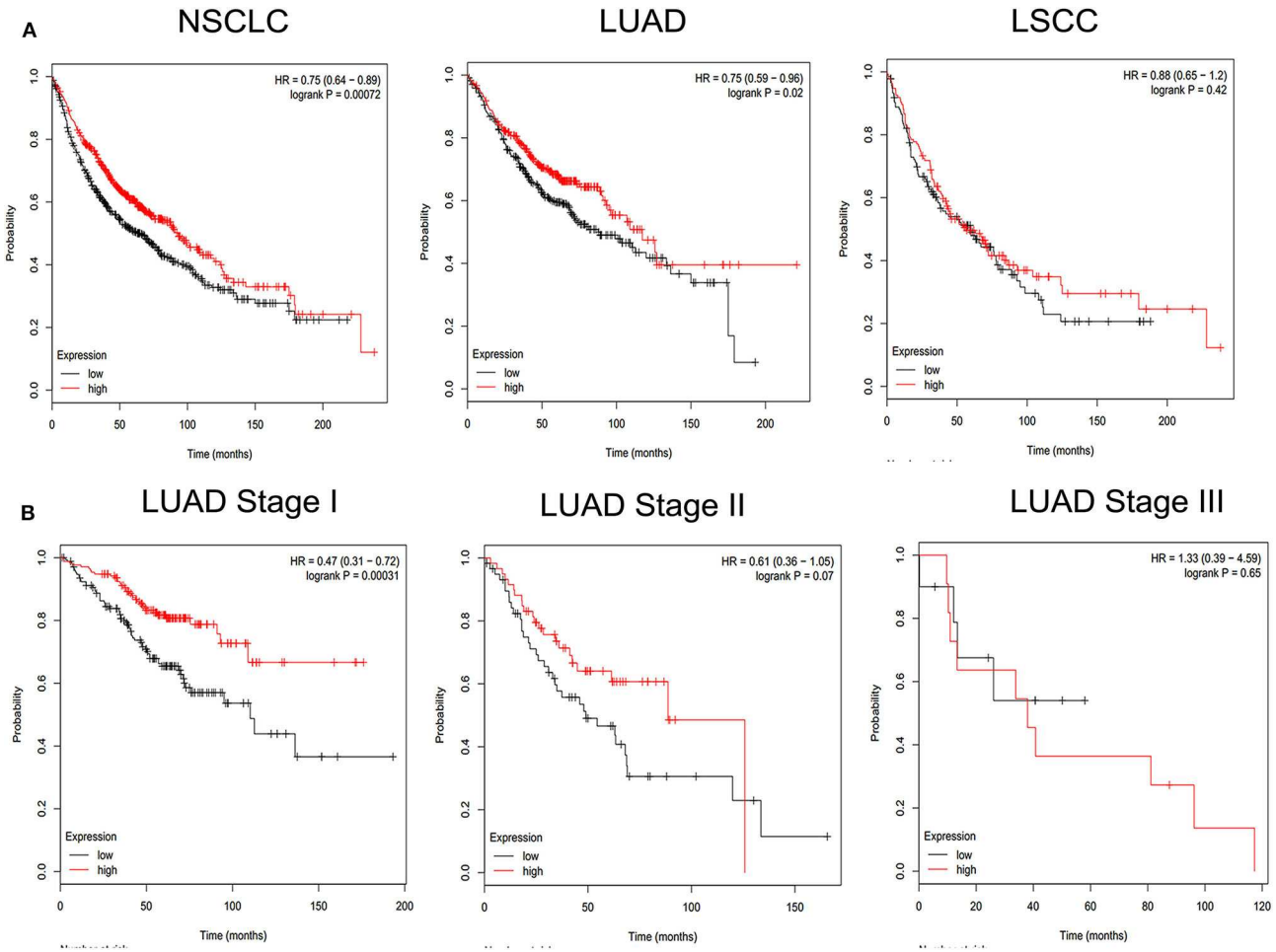


FIGURE 7 | KM survival curve and log-rank test for patients with Cbl-b high or low expression in different pathological types of NSCLC and LUAD at Stage I, Stage II, Stage III. Online Kaplan Meier plotter Database was used to analysis. **(A)** KM survival curve of Cbl-b in non-small cell lung cancer (NSCLC), lung adenocarcinoma (LUAD), lung squamous carcinoma (LUSC). **(B)** KM survival curve of Cbl-b in different pathological stage in LUAD.

a long overall survival time in NSCLC ($P = 0.00072$) and in LUAD ($P = 0.02$,), but there is not significant differences in LUSC (**Figure 7**). Additionally, high Cbl-b expression was significantly associated with better prognosis in stage I ($P = 0.00031$), whereas there was no significant association in stage II and III ($P = 0.07$ and $P = 0.65$, **Figure 7**). To summarize, Cbl-b, whose low expression indicates poor prognosis in stage I, is the target of miR-1323 of NSCLC patients and the correlation between Cbl-b and prognosis depends on histological type and clinical stage.

DISCUSSION

Previous studies have shown that microRNAs participate in lung cancer development and progression, and may be potential prognostic biomarkers for lung cancer (19, 20). In this study, miR-1323 was selected through an online database, and was found to predict the survival of lung cancer patients. We validated miR-1323 in an independent cohort and demonstrated that high miR-1323 expression predicted poor survival in LUAD. Mechanistically, miR-1323 promoted the migration of lung cancer cells by targeting Cbl-b.

It is well-known that microRNAs are involved in multiple biological processes such as cell differentiation, proliferation, apoptosis, EMT, and cell migration, through targeting gene expression (21–23). Previously, a study has shown that miR-1323 promoted radiation-resistant lung cancer cells (10). However, the effect of miR-1323 in LUAD is unknown. In the present study, we examined the expression of miR-1323 in 53 LUAD, and found that its high expression predicted poor survival in LUAD ($P = 0.0253$). It is worth noting that the data from the GEO database (GSE42425) came from different populations, including the Whites, Blacks, Asians, and American Indians, but mainly the Whites. The patients' sample data from our results is mainly the Asians. However, both cohorts show that the expression of miR-1323 affects the prognosis of patients with LUAD. These results validated miR-1323 as a potential prognostic marker in LUAD patients.

The executive effect of miR-1323 in LUAD is still unknown. In the present study, we found that miR-1323 promoted cell migration in LUAD cell lines A549 and HCC827. A previous study has shown that PRKDC was the target gene of miR-1323, which was involved in radiation resistance.In the present study, we used bioinformatics tools to predict the potential target genes of miR-1323 and found Cbl-b as a potential target gene, which was validated by the dual luciferase reporter gene assay. Cbl-b was found to be a direct target gene of miR-1323. Cbl-b is the second member of the c-Cbl RING finger E3 ubiquitin protein ligases. It targets tyrosine-kinase receptors and growth signaling proteins for ubiquitination and down regulation (24–26). Our previous studies demonstrated that Cbl-b inhibited migration of gastric cancer and breast cancer cell (27, 28). However, whether Cbl-b is involved in the migration of lung cancer cells was unknown. In the present study, we first demonstrated Cbl-b was the target gene of miR-1323, and inhibited the migration of A549 cells. Additionally, analysis of an online database showed that Cbl-b is associated with the survival of LUAD patients in stage I.

In conclusion, our present study highlights the important significance of miR-1323 in predicting the poor survival in early LUAD and regulating of lung adenocarcinoma cells metastasis by targeting Cbl-b. We provide a potential early diagnosis biomarker for lung adenocarcinoma, which could help lung cancer patients development effective and appropriate strategies as early as possible when transferred into clinical practice. However, the mechanism of miR-1323 needs to be further explored. Our team will devote ourselves to in-depth research based on present results.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE42425. The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Research Ethics Committee of Shengjing Hospital of China Medical University and all patients signed informed consent forms to allow analyses to be performed on their tissue samples.

AUTHOR CONTRIBUTIONS

HZ, XH, and YF contributed conception and design of the study. DP, CZ, and KH gave the technical support. XY, SX, SW, and TZ performed the experiment and statistical analysis. YW, XC, and YC wrote the manuscript. JK, YL, XH, and XQ provided fund supports. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

This work was supported by the National Science and Technology Major Project of the Ministry of Science and Technology of China [grant number 2017ZX09304025]; the Science and Technology Plan Project of Liaoning Province [grant number 2014226033]; the Science and Technology Plan Project of Liaoning Province [grant number 2016007010]; the Key Research and Development Program of Shenyang [grant number 17-230-9-01]; the National Natural Science Foundation of China [grant number 81472193]; and the general project of liaoning province department of education [grant number L2014296]. Thank you for these supports.

REFERENCES

1. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2013. *CA Cancer J Clin*. (2013) 63:11–30. doi: 10.3322/caac.21166
2. Janne PA, Yang JC, Kim DW, Planchard D, Ohe Y, Ramalingam SS, et al. AZD9291 in EGFR inhibitor-resistant non-small-cell lung cancer. *N Engl J Med*. (2015) 372:1689–99. doi: 10.1056/NEJMoa1411817
3. Wei K, Pan C, Yao G, Liu B, Ma T, Xia Y, et al. miR-106b-5p promotes proliferation and inhibits apoptosis by regulating BTG3 in non-small cell lung cancer. *Cell Physiol Biochem*. (2017) 44:1545–58. doi: 10.1159/000485650
4. Li JH, Sun SS, Li N, Lv P, Xie SY, Wang PY. miR-205 as a promising biomarker in the diagnosis and prognosis of lung cancer. *Oncotarget*. (2017) 8:91938–49. doi: 10.18632/oncotarget.20262
5. Zhu Y, Li T, Chen G, Yan G, Zhang X, Wan Y, et al. Identification of a serum microRNA expression signature for detection of lung cancer, involving miR-23b, miR-221, miR-148b, and miR-423-3p. *Lung Cancer*. (2017) 114:6–11. doi: 10.1016/j.lungcan.2017.10.002
6. Xu X, Cao L, Zhang Y, Lian H, Sun Z, Cui Y. MicroRNA-1246 inhibits cell invasion and epithelial mesenchymal transition process by targeting CXCR4 in lung cancer cells. *Cancer Biomark*. (2018) 21:251–60. doi: 10.3233/CBM-170317
7. Qian L, Ji AH, Zhang WJ, Zhao N. HuR, TTP, and miR-133b expression in NSCLC and their association with prognosis. *Eur Rev Med Pharmacol Sci*. (2018) 22:430–42. doi: 10.26355/eurrev_201801_14192
8. Yang W, Bai J, Liu D, Wang S, Zhao N, Che R, et al. miR-93-5p up-regulation is involved in non-small cell lung cancer cells proliferation and migration and poor prognosis. *Gene*. (2018) 647:13–20. doi: 10.1016/j.gene.2018.06.012
9. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data sets–update. *Nucleic Acids Res*. (2013) 41:D991–5. doi: 10.1093/nar/gks1193
10. Li Y, Han W, Ni TT, Lu L, Huang M, Zhang Y, et al. Knockdown of microRNA-1323 restores sensitivity to radiation by suppression of PRKDC activity in radiation-resistant lung cancer cells. *Oncol Rep*. (2015) 33:2821–8. doi: 10.3892/or.2015.3884
11. Wong NW, Chen Y, Chen S, Wang X. OncomiR: an online resource for exploring pan-cancer microRNA dysregulation. *Bioinformatics*. (2018) 34:713–5. doi: 10.1093/bioinformatics/btx627
12. Wang X. Improving microRNA target prediction by modeling with unambiguously identified microRNA-target pairs from CLIP-ligation studies. *Bioinformatics*. (2016) 32:1316–22. doi: 10.1093/bioinformatics/btw002
13. Dweep H, Gretz N. miRWalk2.0: a comprehensive atlas of microRNA-target interactions. *Nat Methods*. (2015) 12:697. doi: 10.1038/nmeth.3485
14. Agarwal V, Bell GW, Nam JW, Bartel DP. Predicting effective microRNA target sites in mammalian mRNAs. *Elife*. (2015) 4:28. doi: 10.7554/eLife.05005.028
15. Lin L, Asthana S, Chan E, Bandyopadhyay S, Martins MM, Olivas V, et al. Mapping the molecular determinants of BRAF oncogene dependence

in human lung cancer. *Proc Natl Acad Sci USA*. (2014) 111:E748–57. doi: 10.1073/pnas.1320956111
16. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. (2009) 4:44–57. doi: 10.1038/nprot.2008.211
17. Gyorffy 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. doi: 10.1371/journal.pone.0082241
18. Li C, Dong Q, Che X, Xu L, Li Z, Fan Y, et al. MicroRNA-29b-2-5p inhibits cell proliferation by directly targeting Cbl-b in pancreatic ductal adenocarcinoma. *BMC Cancer*. (2018) 18:681. doi: 10.1186/s12885-018-4526-z
19. Yang M, Shen H, Qiu C, Ni Y, Wang L, Dong W, et al. High expression of miR-21 and miR-155 predicts recurrence and unfavourable survival in non-small cell lung cancer. *Eur J Cancer*. (2013) 49:604–15. doi: 10.1016/j.ejca.2012.09.031
20. Bobbili MR, Mader RM, Grillari J, Dellago H. OncomiR-17-5p: alarm signal in cancer? *Oncotarget*. (2017) 8:71206–22. doi: 10.18632/oncotarget.19331
21. Jiang J, Yi B, Qin C, Ding S, Cao W. Upregulation of microRNA27b contributes to the migration and invasion of gastric cancer cells via the inhibition of sprouty2mediated ERK signaling. *Mol Med Rep*. (2016) 13:2267–72. doi: 10.3892/mmr.2016.4779
22. Zhang P, Tang WM, Zhang H, Li YQ, Peng Y, Wang J, et al. miR-646 inhibited cell proliferation and EMT-induced metastasis by targeting FOXK1 in gastric cancer. *Br J Cancer*. (2017) 117:525–34. doi: 10.1038/bjc.2017.181
23. Lei C, Du F, Sun L, Li T, Li T, Min Y, et al. miR-143 and miR-145 inhibit gastric cancer cell migration and metastasis by suppressing MYO6. *Cell Death Dis*. (2017) 8:e3101. doi: 10.1038/cddis.2017.493
24. Nau MM, Lipkowitz S. Comparative genomic organization of the cbl genes. *Gene*. (2003) 308:103–13. doi: 10.1016/S0378-1119(03)00471-2
25. Paolino M, Thien CB, Gruber T, Hinterleitner R, Baier G, Langdon WY, et al. Essential role of E3 ubiquitin ligase activity in Cbl-b-regulated T cell functions. *J Immunol*. (2011) 186:2138–47. doi: 10.4049/jimmunol.1003390
26. Welchman RL, Gordon C, Mayer RJ. Ubiquitin and ubiquitin-like proteins as multifunctional signals. *Nat Rev Mol Cell Biol*. (2005) 6:599–609. doi: 10.1038/nrm1700
27. Vennin C, Spruyt N, Dahmani F, Julien S, Bertucci F, Finetti P, et al. H19 non coding RNA-derived miR-675 enhances tumorigenesis and metastasis of breast cancer cells by downregulating c-Cbl and Cbl-b. *Oncotarget*. (2015) 6:29209–23. doi: 10.18632/oncotarget.4976
28. Dong Q, Li C, Che X, Qu J, Fan Y, Li X, et al. MicroRNA-891b is an independent prognostic factor of pancreatic cancer by targeting Cbl-b to suppress the growth of pancreatic cancer cells. *Oncotarget*. (2016) 7:82338–53. doi: 10.18632/oncotarget.11001

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

ORIGINAL RESEARCH
published: 21 February 2020
doi: 10.3389/fonc.2020.00168

Association of *MSH2* Expression With Tumor Mutational Burden and the Immune Microenvironment in Lung Adenocarcinoma

Mingming Jia[1,2,3], *Linli Yao*[4], *Qin Yang*[4*] and *Tian Chi*[1,5*]

[1] School of Life Science and Technology, ShanghaiTech University, Shanghai, China, [2] CAS Center for Excellence in Molecular Cell Science, Shanghai Institute of Biochemistry and Cell Biology, Chinese Academy of Sciences, Shanghai, China, [3] University of Chinese Academy of Sciences, Beijing, China, [4] State Key Laboratory of Oncogenes and Related Genes, Shanghai Cancer Institute, Ren Ji Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, China, [5] Department Immunobiology, Yale University School of Medicine, New Haven, CT, United States

OPEN ACCESS

Edited by:
Umberto Malapelle, University of Naples Federico II, Italy

Reviewed by:
Pasquale Pisapia, University of Naples Federico II, Italy
Fabio Pagni, University of Milano Bicocca, Italy

***Correspondence:**
Qin Yang
qyang@shsci.org
Tian Chi
chitian@shanghaitech.edu.cn

Specialty section:
This article was submitted to Thoracic Oncology, a section of the journal Frontiers in Oncology

Received: *30 October 2019*
Accepted: *30 January 2020*
Published: *21 February 2020*

Citation:
Jia M, Yao L, Yang Q and Chi T (2020) Association of MSH2 Expression With Tumor Mutational Burden and the Immune Microenvironment in Lung Adenocarcinoma. Front. Oncol. 10:168. doi: 10.3389/fonc.2020.00168

Immune checkpoint blockade (ICB) therapies that target programmed cell death 1 (PD1) and PD1 ligand 1 (PDL1) have demonstrated promising benefits in lung adenocarcinoma (LUAD), and tumor mutational burden (TMB) is the most robust biomarker associated with the efficacy of PD-1-PD-L1 axis blockade in LUAD, but the assessment of TMB by whole-exome sequencing (WES) is rather expensive and time-consuming. Although targeted panel sequencing has been developed and approved by the US Food and Drug Administration (FDA) to estimate TMB, we found that its predictive accuracy for ICB response was significantly lower than WES in LUAD. Given that previous studies were mainly focusing on genomic variations to explore surrogate biomarkers of TMB, we turned to examine the transcriptome-based correlation with TMB in this study. Combining three immunotherapeutic cohorts with two independent The Cancer Genome Atlas (TCGA) datasets, we revealed that the expression of mutS homolog 2 (*MSH2*), one of the most crucial genes involved in DNA mismatch repair (MMR) pathway, was the strongest feature associated with increased TMB in multivariate analysis. Furthermore, *MSH2* expression also displayed a significantly positive correlation with smoking signature while an inverse association with MMR deficiency (MMRd) signature in LUAD. More importantly, high expression of *MSH2* markedly correlated with increased *PD-L1* expression and CD8+ T cell infiltration, both suggesting a prominent immunotherapy-responsive microenvironment in LUAD. Notably, detecting *MSH2* expression is much easier, faster, and cheaper than TMB in clinical practice. Taken together, this study demonstrates the association of *MSH2* expression with TMB and the immune microenvironment in LUAD. *MSH2* expression may be developed as a potential surrogate biomarker of TMB to identify ICB responders in LUAD.

Keywords: lung adenocarcinoma, *MSH2* expression, biomarkers, immunotherapy, tumor mutation burden

INTRODUCTION

Recent clinical trials with immune checkpoint blockade (ICB) therapies have demonstrated durable clinical responses in patients with non-small cell lung cancer (NSCLC), but only a minority of patients respond (1–3). The combination of ICB therapies can improve response rates but also result in more severe adverse effects than single-agent therapy (4). Previous studies have reported that tumor mutational burden (TMB) (1–3, 5–7), programmed death-ligand 1 (PD-L1) expression (5, 8), CD8+ T cell infiltration (1, 9, 10), and DNA mismatch repair deficiency (MMRd) (5, 11) could affect the efficacy of PD-1 blockade immunotherapy (5, 12, 13). However, only TMB and PD-L1 expression are validated as predictive biomarkers for ICB response in phase III clinical trials across multiple cancer types (5, 7).

Currently, TMB performs much better than other biomarkers for predicting ICB response in NSCLC (1–3). High TMB can potentially generate higher immunogenic neoantigens presented on the tumor cell surface and then facilitate immune recognition of tumor cells as foreign (1, 2, 7). However, the assessment of TMB is expensive and time-consuming (6, 14). Alternatively, PD-L1 expression assessed by immunohistochemistry (IHC) is much cheaper and timelier to select candidates for ICB therapies, but many patients whose tumors are PD-L1-positive do not respond (1). Additionally, the localization (on tumor-infiltrating immune cells or tumor cells) and positivity threshold of PD-L1 expression for predicting ICB efficacy are still undetermined, which may affect its clinical application (1, 5, 6, 8, 15). Therefore, we hypothesized that other factors, which highly correlated with increased TMB and were as convenient as PD-L1 expression to be detected, might also be developed as potential biomarkers to predict ICB response in NSCLC.

To test our hypothesis, we recruited three well-studied NSCLC immunotherapeutic cohorts (1–3) and one multidimensional non-immunotherapeutic The Cancer Genome Atlas (TCGA) dataset. As previous studies reported (2, 10, 16–19), TCGA samples without ICB therapies are still informative to explore tumor immune escape and can also derive surrogate biomarkers for ICB therapies. Combining these four cohorts, we revealed that *MSH2* expression was the most robust feature associated with increased TMB and smoking signature in multivariate analysis and might be developed as a potential surrogate biomarker of TMB for identifying ICB responders in lung adenocarcinoma (LUAD), one of the commonest types of NSCLC (20, 21).

MATERIALS AND METHODS

Clinical Immunotherapeutic Patients

Given the intratumoral heterogeneity across different cancer subtypes, it is more reliable to discover the specific determinants for ICB efficacy within the same cancer subtype (5), so we only focused on the LUAD subtype according to its dominating proportion in previous NSCLC immunotherapeutic cohorts (1–3). We collected three LUAD cohorts containing both clinical and genomic characteristics, which were initially reported in *Science* (1), *Journal of Clinical Oncology (JCO)* (3), and *Cancer Cell* (2) journals. For the *Science-LUAD* cohort, it contained 29 LUAD patients treated with PD-1 blockade (pembrolizumab) (1). For the *Cancer Cell-LUAD* cohort, it involved 59 LUAD patients treated with PD-1 plus CTLA-4 blockade (nivolumab plus ipilimumab) (2). For the *JCO-LUAD* cohort, it contained 186 LUAD patients who had received anti-PD-(L)1 monotherapy or in combination with anti-CTLA-4 (3).

TCGA-LUAD Datasets Without Immunotherapy

Non-immunotherapeutic *TCGA-LUAD* datasets were extracted from the UCSC Xena multi-omics database platform (22) (https://tcga.xenahubs.net), including somatic mutation (n = 543) and RNA-seq expression (n = 576) profiles. We first removed adjacent normal samples from RNA-seq expression data and then only analyzed those LUAD samples that had both genomic and transcriptomic profiles (n = 478).

Tumor Mutational Burden (TMB) Estimates

TMB was defined as the number of somatic non-synonymous single nucleotide variants. Raw somatic mutation data in three immunotherapeutic cohorts were extracted from the respective **Supplementary Materials** (1–3). Mutation profiles were assessed by whole-exome sequencing (WES) on the Illumina platform in *Science-LUAD* (1) and *Cancer Cell-LUAD* (2) cohorts while determined by MSK-IMPACT targeted panel sequencing on specific cancer-associated genes in *JCO-LUAD* (3) cohort. The detailed methodology for generating mutation calls has previously been described (1–3). For the *TCGA-LUAD* dataset, somatic mutation data were retrieved from the UCSC Xena multi-omics database platform (https://xenabrowser.net/datapages/?dataset=TCGA.LUAD.sampleMap%2Fmutation_broad&host=https%3A%2F%2Ftcga.xenahubs.net&removeHub=https%3A%2F%2Fxena.treehouse.gi.ucsc.edu%3A443) and preprocessed at the Broad Institute Genome Sequencing Center (22). WES data were generated on the Illumina platform. Mutation calls were calculated using the MuTect method (23), and only calls with variant allele frequency (VAF) >4.0% were included (22). The R package "maftools" (24) was then used to calculate the total number of somatic non-synonymous point mutations within each sample.

RNA-seq and Gene Set Enrichment Analysis (GSEA)

For three immunotherapeutic cohorts, RNA-seq data were not available. For *TCGA-LUAD* datasets, RNA-seq data were assessed using the Illumina RNA sequencing platform. We downloaded the level 4 gene expression data from the UCSC Xena platform (22). The pre-processing and quality control of expression data have previously been described (22). The unit of mRNA expression value is pan-cancer normalized $\log_2$ (norm_count+1).

For pathway enrichment analysis, we used MSigDB (Molecular Signatures Database) of KEGG gene sets (25) to enrich the significant pathways, which were determined by a list of genes that highly correlated with increased TMB (**Table S2**; AUC > 0.65, $P < 0.0001$). For the enriched results, a $P > 0.05$

was considered statistically significant. We also recruited another tool, GSEA software (http://software.broadinstitute.org/gsea/index.jsp) (25), to confirm the pathway we enriched. GSEA integrates the expression data with phenotypes' information to determine whether a gene set significantly correlates with a defined phenotype. The normalized enrichment score (NES) and the nominal *P*-value are two primary statistics to examine the GSEA results. The ranking metric score is used to measure the correlation of a gene with a phenotype, with a positive value indicating a correlation with the first phenotype and a negative value indicating a correlation with the second.

Mutational Signature Analysis

We used the "SignatureAnalyzer" R package (26, 27) to calculate the percentage of mutation signature within each tumor sample. "SignatureAnalyzer" can capture the non-negative matrix factorization algorithm (NMF) to decipher mutation signatures within cancer genomes, and then it automatically calculates the optimal number of mutation signatures (W) and the fraction of mutation signature in an individual sample. Mutation Annotation Format (MAF) files are available in the *TCGA-LUAD* dataset (http://gdac.broadinstitute.org/) and necessary for this analysis. The detailed method of mutation signature analysis has been described (https://software.broadinstitute.org/cancer/cga/msp).

Immune Cellular Infiltration Estimates

The abundance of tumor-infiltrating immune cells (CD8+ T cells, T-regulatory cells, and macrophages) in LUAD samples was assessed using the CIBERSORT algorithm (28). CIBERSORT is an influential deconvolution method that uses support vector regression to quantify the cellular components from bulk tissue gene expression profiles. Based on gene expression data, CIBERSORT can accurately estimate the immune composition within a given tumor sample. We extracted the relative proportion of immune cells of *TCGA-LUAD* samples from the Pan-Cancer Atlas (https://www.cell.com/pb-assets/consortium/PanCancerAtlas/PanCani3/index.html) (18) and then compared them according to the indicated *MSH2* expression status.

Statistical Analyses

Statistical analyses were performed using R software (version 3.5.2) and GraphPad Prism software (version 7.0.0). Student's *t*-test or Mann-Whitney *U* test was used to determine the differences between two groups. Kruskal-Wallis test was used to determine the differences among three or more groups. We used ROC curves with the highest Youden index to determine the optimum cut-off of TMB and *MSH2* expression.

The proportion of gene mutation was compared using Fisher's exact test. Pairwise correlations were calculated using the Spearman correlation formula. Multivariate logistic and linear regression models were conducted to assess the impact of gene expression on TMB, adjusting for other covariates described. All reported *p*-values were two-sided.

RESULTS

Clinical and Genomic Characteristics of Selected Cohorts

We retrieved many previous studies and cancer databases, only collecting four high-quality LUAD datasets that contained both clinical and genomic information: 29 LUAD patients treated with anti-PD-1 therapy (*Science-LUAD*) (1), 59 LUAD patients treated with PD-1 plus CTLA-4 blockade (*Cancer Cell-LUAD*) (2), 186 LUAD patients treated with anti-PD-1/PD-L1 therapies or in combination with anti-CTLA-4 therapy (*JCO-LUAD*) (3), and 478 LUAD patients without immunotherapy (*TCGA-LUAD*) (**Figure 1**; **Table S1**). Pre-therapy tissues from LUAD patients were assessed by whole-exome sequencing (WES) in *Science-LUAD*, *Cancer Cell-LUAD*, and *TCGA-LUAD* cohorts while targeted panel sequencing (MSK-IMPACT panel, covering specific cancer-related genes) in *JCO-LUAD* dataset (**Figure 1**; **Table S1**). Since *TCGA-LUAD* datasets had more samples than the other three cohorts, we randomly divided it into two independent cohorts to further validate our hypothesis (*Discovery-LUAD* and *Validation-LUAD*, respectively) (**Figure 1**; **Table S1**).

TMB was defined as the total number of somatic non-synonymous point mutations. Except for the *JCO-LUAD* cohort [median six and interquartile range (IQR) 3-11] assessed by targeted panel sequencing, the quantity and range of TMB in *TCGA-LUAD* (median 178 and IQR 80-326 in the *Discovery-LUAD* cohort; median 167 and IQR 68-313 in the *Validation-LUAD* cohort) were similar to that in *Science-LUAD* (median 201 and IQR 109-302) and *Cancer Cell-LUAD* (median 143 and IQR 40-296) cohorts (**Figure 2A**), suggesting the homogeneity of these cohorts as previously reported (1, 2).

WES Outperformed Targeted Panel Sequencing for Assessing TMB as an ICB Biomarker

Previous studies reported that TMB assessed by WES or targeted panel sequencing was significantly associated with improved efficacy of ICB therapies in LUAD (1–3). Moreover, the assessment of TMB by targeted panel sequencing also highly correlated with WES ($r = 0.86$, $P < 0.001$) (3). However, using receiver operator characteristic (ROC) curves as previously suggested (1–3, 16), we found that WES-based TMB achieved consistently better performance than targeted panel sequencing for predicting ICB response in LUAD (**Figure 2B**; AUC = 0.82 (*Science-LUAD*), 0.80 (*Cancer Cell-LUAD*), and 0.60 (*JCO-LUAD*), respectively). Additionally, TMB assessed by WES [**Figure 2B**; AUC = 0.82 (*Science-LUAD*); and 0.80 (*Cancer Cell-LUAD*), respectively] also performed better than PD-L1 expression detected by IHC for predicting ICB response in LUAD [**Figure S1A**; AUC = 0.61 (*Cancer Cell-LUAD*); and 0.69 (*JCO-LUAD*), respectively].

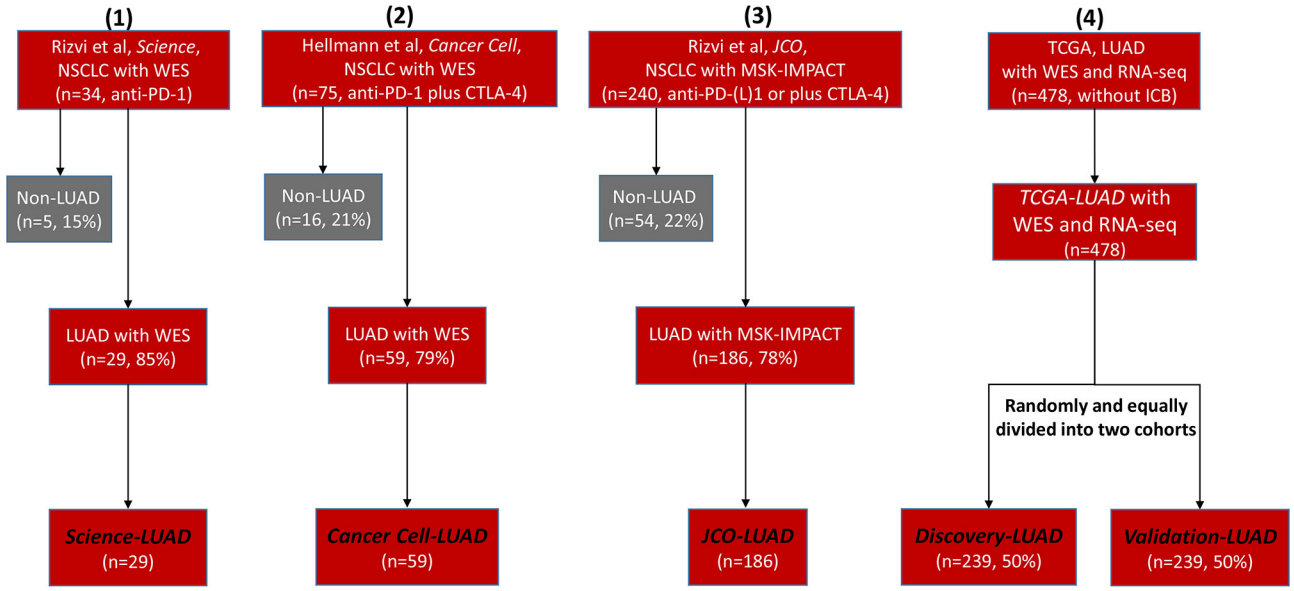


FIGURE 1 | Flowchart of four selected clinical cohorts for statistical analysis. NSCLC, non–small cell lung cancer; WES, Whole-Exome Sequencing; LUAD, lung adenocarcinoma; PD-1, programmed cell death-1; CTLA-4, cytotoxic T lymphocyte antigen-4; MSK-IMPACT, Memorial Sloan Kettering-Integrated Mutation Profiling of Actionable Cancer Targets; TCGA, The Cancer Genome Atlas; ICB, immune checkpoint blockade.

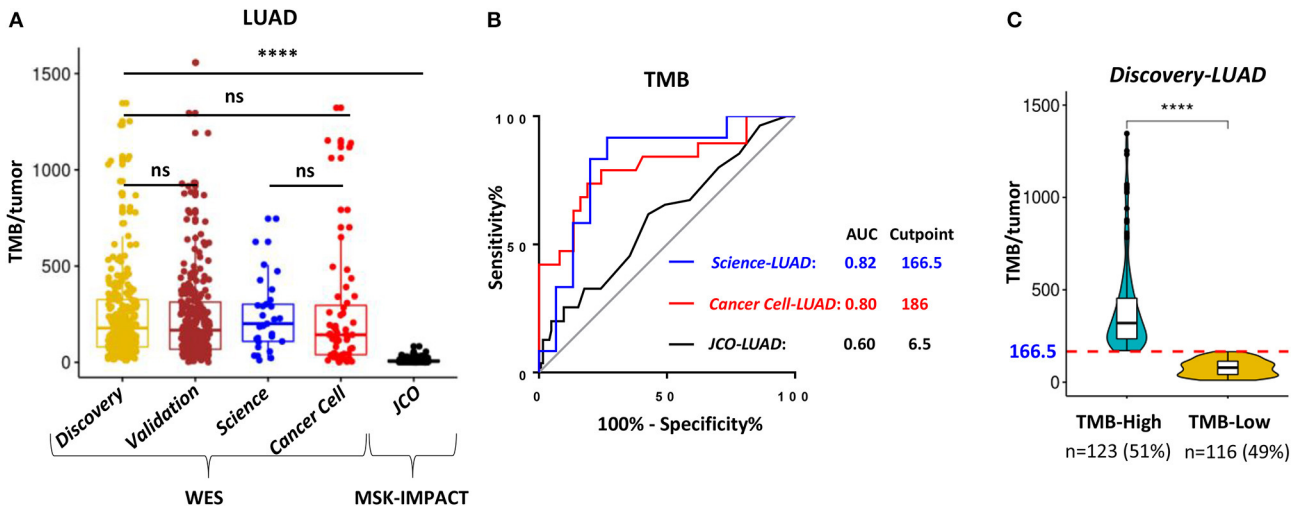


FIGURE 2 | WES outperformed targeting-sequencing for assessing TMB as an ICB biomarker. **(A)** Quantitative analysis of TMB between *TCGA- LUAD* and the other three published LUAD cohorts for immunotherapy. **(B)** Receiver operating characteristic (ROC) curves for the correlation of TMB with clinical response to ICB therapies in the three cohorts. AUC, area under a ROC curve; Cutpoint, the Youden index-associated cutoff value of TMB. **(C)** Samples in the *Discovery-LUAD* cohort were stratified into two groups based on the Youden index-associated cutpoint of TMB from the *Science-LUAD* cohort. TMB-High: ≥166.5; TMB-Low: <166.5. ****P < 0.0001, ns, non significant.

MSH2 Expression Significantly Correlated With Increased TMB and Performed Better Than *PD-L1* on Predicting TMB in LUAD

Given that transcriptomic data in three immunotherapeutic cohorts were not available, we turned to use multidimensional *TCGA-LUAD* datasets, which contained both genomic and transcriptomic features, to further explore the potential determinants associated with increased TMB in LUAD.

To demonstrate the potential clinical usefulness of TMB for predicting ICB response in LUAD, the Youden index was used to choose the optimum cut point of TMB (16, 29). The index-associated cut point of TMB in *Science-LUAD* was very close to that in the *Cancer Cell-LUAD* cohort (**Figure 2B**; TMB = 166.5 and 186, respectively), which was also very approximate to a previous report in NSCLC (TMB = 178) (1). Given that *Science-LUAD* cohort was only treated with PD-1 blockade and

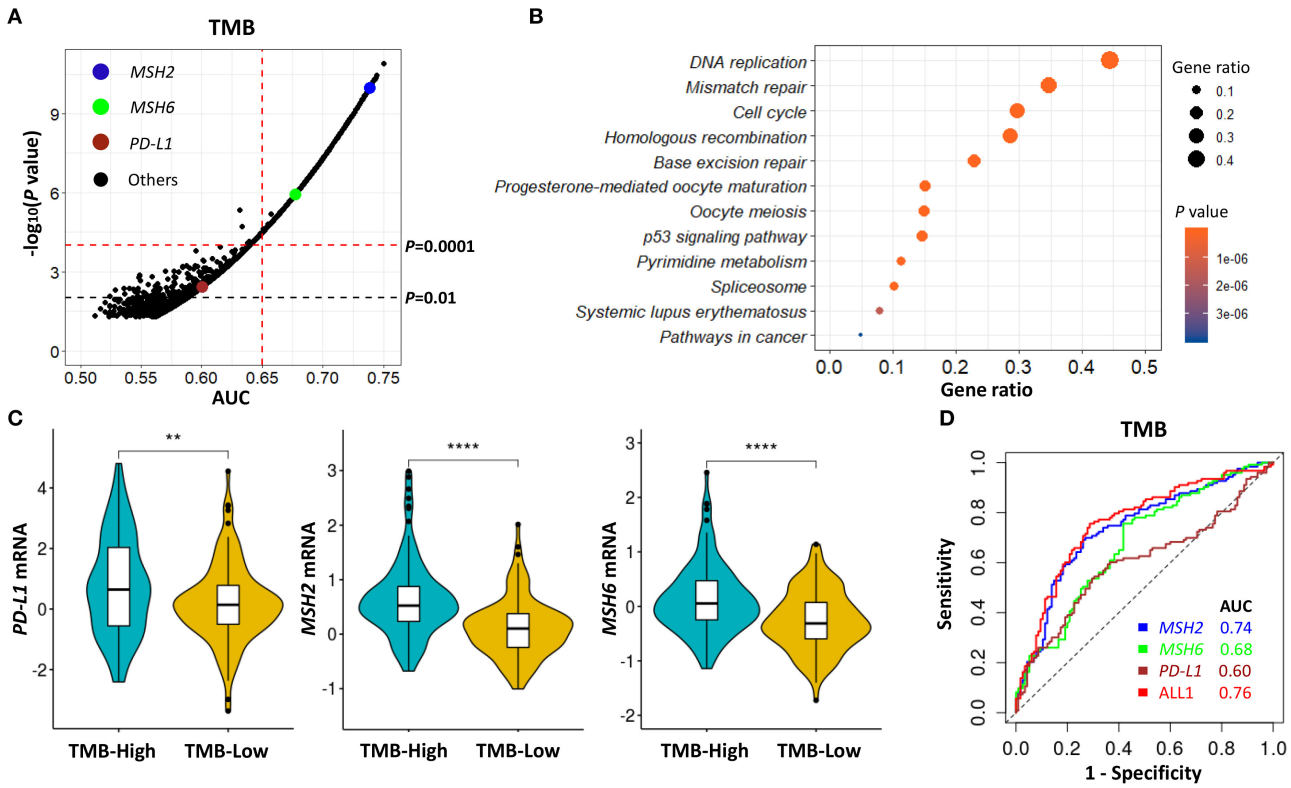


FIGURE 3 | *MSH2* expression highly correlated with increased TMB and performed better than *PD-L1* on predicting TMB in the *Discovery-LUAD* cohort. **(A)** Volcano plot of AUC and rank sum *P*-values demonstrating the positive correlation of transcriptomic features with increased TMB in the *Discovery-LUAD* cohort. AUC, area under a ROC curve. **(B)** Pathway enrichment analysis of genes whose expression significantly and positively associated with increased TMB (AUC > 0.65, $P < 0.0001$), with $P > 0.05$ considered statistically significant. **(C)** Quantitative analysis of *PD-L1*, *MSH2*, and *MSH6* mRNA expression in two groups according to indicated TMB status. **(D)** ROC curves for the association of *PD-L1*, *MSH2*, and *MSH6* mRNA expression with TMB status. ALL1: the combination of *PD-L1*, *MSH2*, and *MSH6* mRNA expression. ****$P < 0.0001$, **$P < 0.01$.

performed better than *Cancer Cell-LUAD* cohort on predicting ICB efficacy (**Figure 2B**; AUC = 0.82 and 0.80, respectively), we stratified the *Discovery-LUAD* cohort into two groups based on the TMB cutoff from *Science-LUAD* cohort (**Figure 2C**; TMB = 166.5).

According to the above TMB-defined groups in the *Discovery-LUAD* cohort, we performed the ROC test to all genes, examining the association of TMB with all transcriptomic features (**Figure 3A**). A list of genes, which highly and positively correlated with increased TMB (AUC > 0.65, $P < 0.0001$), were significantly enriched in the mismatch repair (MMR) pathway (**Figures 3A,B**; **Tables S2, S3**), consistent with the result of gene set enrichment analysis (GSEA) (**Figure S2A**). Notably, *MSH2* and *MSH6* are two key cancer-related MMR genes and were as similar as *PD-L1* expression significantly up-regulated in patients with high TMB in *Discovery-LUAD* cohort (**Figures 3A–C**; **Table S3**). These results could be reasonably speculated that patients with high TMB would potentially accelerate the expression of MMR-related genes to repair the impaired genome.

Additionally, we also examined the impact of *MSH2* and *MSH6* expression on TMB in the context of *PD-L1* expression. There were moderate correlations of TMB with *MSH2* and *MSH6* expression (**Figure S3A**; $r = 0.46$ and 0.39, respectively) while no significant association with *PD-L1* expression (**Figure S3A**; $r = 0.13$). In multivariate analysis incorporating *MSH2*, *MSH6*, and *PD-L1* expression, *MSH2* expression was the most robust gene associated with increased TMB in the *Discovery-LUAD* cohort (**Figure 3D**; **Figures S3B,C**). Of note, the ROC test incorporating *MSH2*, *MSH6*, and *PD-L1* expression did not significantly improve the predictive ability for TMB compared with *MSH2* expression alone [**Figure 3D**; AUC = 0.74 (*MSH2*) and 0.76 (ALL1), respectively].

MSH2 Expression Outperformed Other MMR-Related Genes for Predicting TMB in LUAD

The MMR pathway is crucial for maintaining genomic integrity, and the deficiency of MMR (MMRd) is also highly sensitive to ICB therapies (11, 30). The potential mechanism is that tumors with MMRd can result in microsatellite instability (MSI) and are a specific subset of high TMB tumors (5). However, in LUAD, the positivity rate of MMRd/MSI assessed by genomic variations is

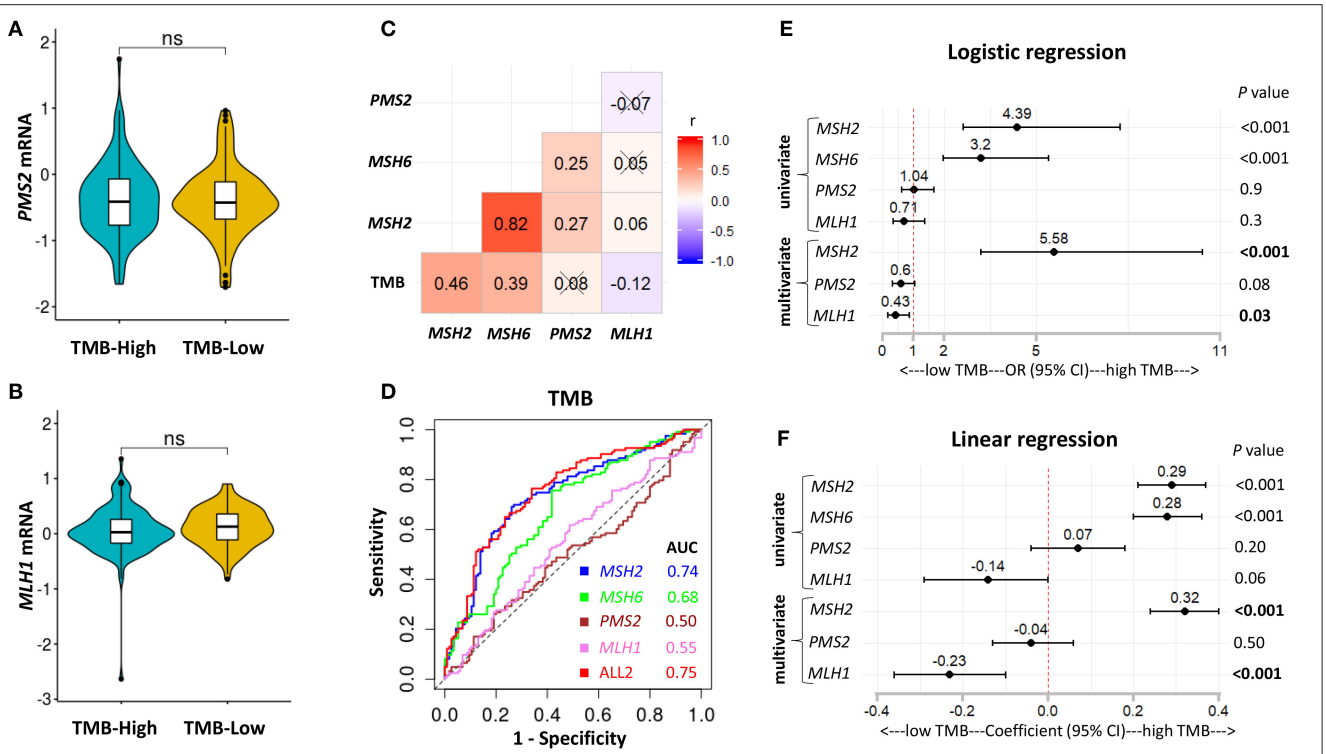


FIGURE 4 | *MSH2* expression outperformed other MMR-related genes for predicting TMB in the *Discovery-LUAD* cohort. **(A,B)** Quantitative analysis of *PMS2* and *MLH1* mRNA expression in two groups based on TMB status. **(C)** Pairwise correlations between TMB and four MMR-related gene expression. Cross indicated no significant correlation ($P > 0.05$); r: Spearman correlation coefficient. **(D)** ROC curves for the association of *MSH2*, *MSH6*, *PMS2*, and *MLH1* mRNA expression with TMB status. ALL2: the combination of *MSH2*, *MSH6*, *PMS2*, and *MLH1* mRNA expression. **(E,F)** Forest plots for univariate and multivariate logistic or linear analysis of TMB with indicated gene expression. Of note, *MSH6* expression was not analyzed in multivariate regression model given its tight correlation with *MSH2* expression. OR, odds ratio. ns, not significant.

<1% and much lower than the objective response rate to PD-1 blockade in unselected patients (13, 19, 30–33). Given that previous studies were mainly focusing on genomic variations to explore the MMRd mechanism in LUAD (13, 33) and our data showed that *MSH2* expression was strongly associated with increased TMB in *Discovery-LUAD* cohort (**Figure 3**; **Figure S3**), we further examined the transcriptome-based MMRd status in LUAD patients with high TMB.

MSH2, *MSH6*, *PMS2*, and *MLH1* are four genes that play a critical role in DNA MMR (13, 30). Four proteins codified by these genes function in heterodimer pairs (MSH2-MSH6 and MLH1-PMS2) to preserve genomic integrity (13, 30). In clinical practice, the inactivation of one of the four genes detected by next-generation sequencing (NGS) or IHC suggests an MMRd mechanism within a tumor (13, 30). However, MSH2 and MLH1 are obligatory partners for forming the two heterodimers, while MSH6 and PMS2 can be replaced by other MMR proteins, such as MSH3, PMS1, and MLH3 (13, 30). We observed that *MSH2* was significantly mutated in patients with high TMB, but it only accounted for 5.7% of high TMB tumors in LUAD (**Figures S4A–D**). In addition, except for the other two MMR genes, *MLH1* and *PMS2* were not significantly up-regulated in patients with high TMB (**Figures 3C**, **4A,B**), suggesting a low abundance of MLH1-PMS2 heterodimers existed in high TMB tumors. Furthermore, *MSH2* whose expression displayed the strongest correlation with increased TMB than the other three MMR genes (**Figure 4C**), and the ROC test incorporating *MSH2*, *MSH6*, *PMS2*, and *MLH1* expression did not significantly improve the predictive ability for TMB compared with *MSH2* expression alone [**Figure 4D**; AUC = 0.74 (*MSH2*); and 0.75 (ALL2), respectively].

Interestingly, using both multivariate logistic and linear regression analysis, we revealed that TMB displayed a significantly positive association with *MSH2* expression while an inverse correlation with *MLH1* expression in *Discovery-LUAD* cohort (**Figures 4E,F**). Furthermore, GSEA incorporating all MMR-related genes also confirmed these findings (**Figures S2B,C**). These results suggested that down-regulated *MLH1* expression in patients with high TMB might result in the dysfunction of the MMR machinery and then potentially facilitate the accumulation of mutations in LUAD.

To further consolidate and extend our findings, we performed two additional analyses. First, we used multivariate regression analysis to demonstrate that *MSH2* expression was independently associated with increased TMB, with adjustment for patients' sex, age, and pack-year (smoking index) in the *Discovery-LUAD* cohort (**Figures S5A,B**). Second, we validated the hypothesis that *MSH2* expression was the strongest determinant associated with increased TMB in another independent *Validation-LUAD* cohort (**Figures 5A–D**).

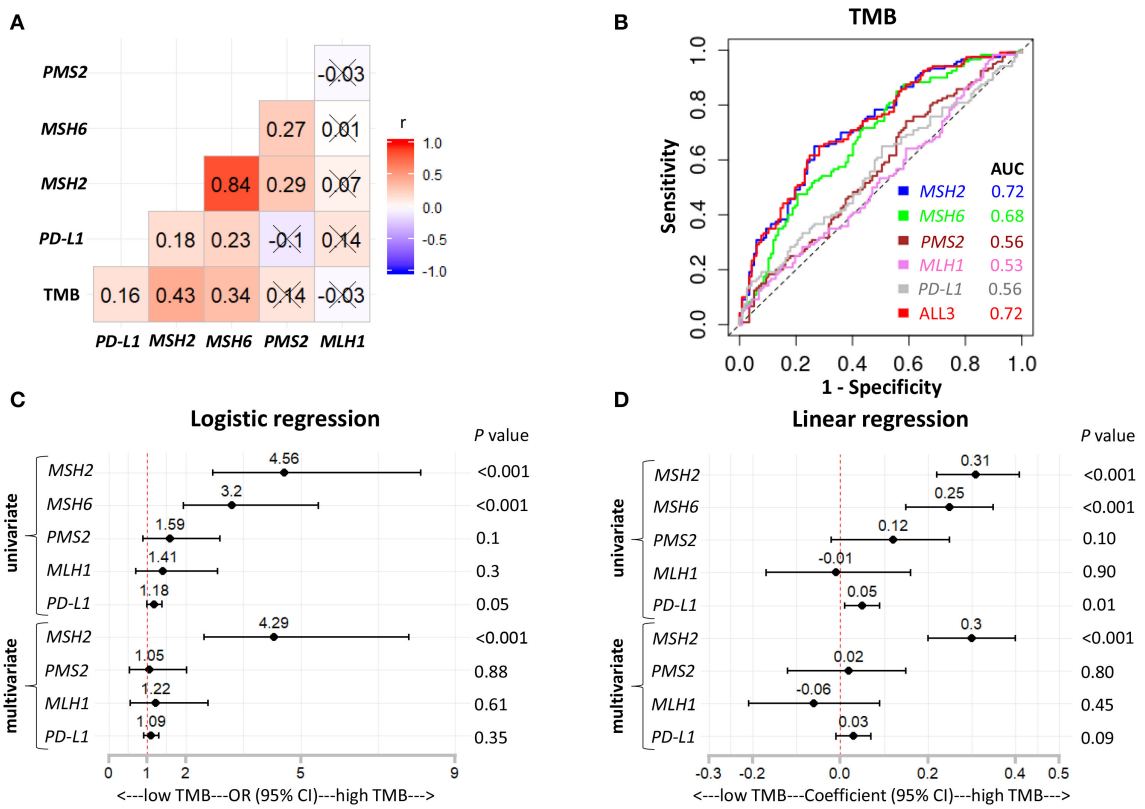


FIGURE 5 | *MSH2* expression outperformed other indicated genes for predicting TMB in the *Validation-LUAD* cohort. **(A)** Pairwise correlations between TMB and the indicated gene expression. Cross indicated no significant correlation ($P > 0.05$); r: Spearman correlation coefficient. **(B)** ROC curves for the association of *PD-L1, MSH2, MSH6, PMS2*, and *MLH1* mRNA expression with the TMB status. ALL3: the combination of *PD-L1, MSH2, MSH6, PMS2*, and *MLH1* mRNA expression. **(C,D)** Forest plots for univariate and multivariate logistic or linear analysis of TMB with the indicated gene expression. Of note, *MSH6* expression was not analyzed in multivariate regression model given its high correlation with *MSH2* expression. OR, odds ratio.

MSH2 Expression Outperformed Other MMR-Related Genes for Predicting Smoking Signature in LUAD

It is well-known that LUAD exhibiting high TMB is strongly associated with cigarette smoking, and smoking signature is also highly sensitive to ICB therapies in LUAD (1, 26). Consistent with previous studies (1, 26), we found that patients with high TMB significantly increased the fractions of smoking signature in LUAD (**Figure 6A**; **Figures S6A–D**). However, MMRd/MSI signature, as determined by NGS data, displayed significantly decreased proportions in patients with high TMB (**Figure 6B**; **Figures S6A–D**), suggesting that genome-assessed MMRd/MSI signature was not suitable as a potential predictor of increased TMB and improved ICB efficacy in LUAD.

Furthermore, using multivariate regression analysis, we demonstrated that *MSH2* expression was the most robust MMR feature positively associated with smoking signature while inversely correlated with MMRd/MSI signature in *Discovery-LUAD* cohort (**Figures 6C,D**), suggesting that high *MSH2* expression might also be a potential predictor of increased smoking signature in LUAD.

High Expression of *MSH2* Significantly Correlated With Increased *PD-L1* Expression and CD8+ T Cell Infiltration Within the Tumor Microenvironment

PD-L1 expression and the infiltration of CD8+ T cells are two important biomarkers for assessing the immunotherapeutic microenvironment in LUAD (1, 5, 10, 12). Therefore, we further examined the association of *MSH2* expression with *PD-L1* expression and CD8+ T cell infiltration within the tumor microenvironment. We stratified the *Validation-LUAD* samples into two groups according to the Youden index-associated cutoff value of *MSH2* expression (**Figure 7A**). We revealed that patients with high *MSH2* expression significantly increased *PD-L1* expression and CD8+ T cell infiltration while decreased the infiltration of T-regulatory cells (Tregs) (**Figures 7B–D**). It has been reported that tumor-associated macrophages (TAMs) were also important for assessing the efficacy of anti-PD-1/PD-L1 therapies (34). We found that inflammatory M1 macrophages, but not pro-tumor M2, were also significantly infiltrated into the tumor tissues with high *MSH2* expression (**Figures S7A,B**). These results suggested that patients with

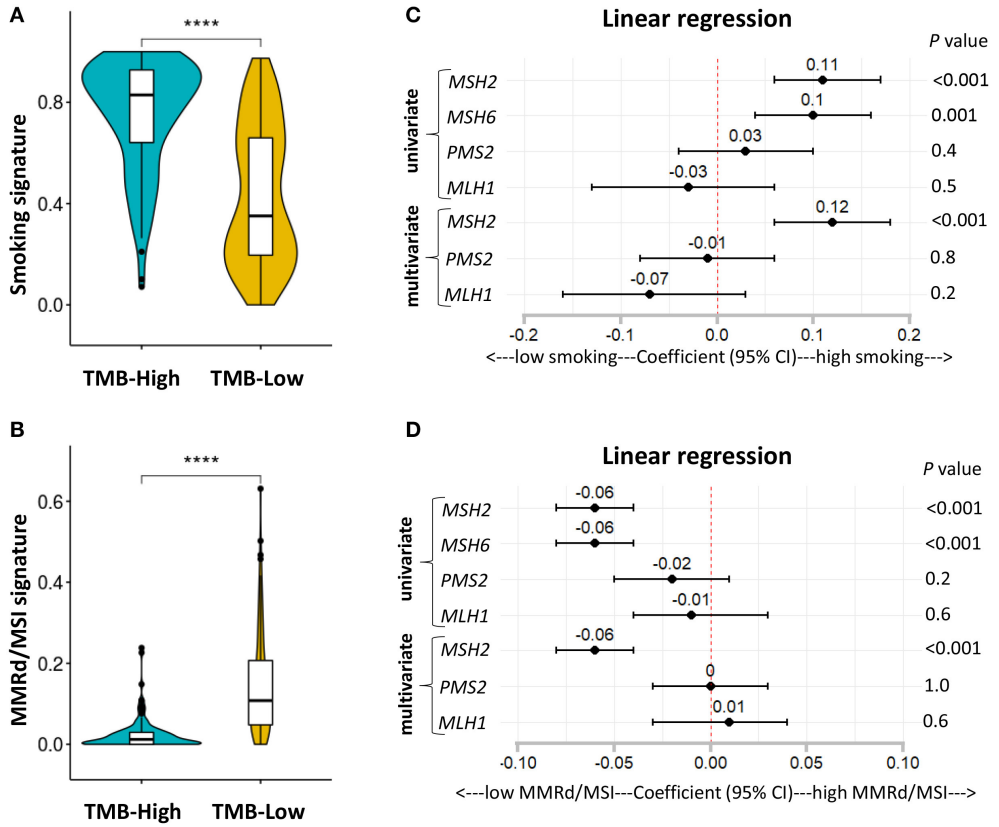


FIGURE 6 | *MSH2* expression positively correlated with smoking signature but negatively associated with MMRd signature in the *Discovery-LUAD* cohort. **(A,B)** Quantitative analysis of smoking and MMRd/MSI signature in two groups based on the TMB status. **(C)** Forest plot for univariate and multivariate linear analysis of smoking signature with the four MMR-related gene expression. **(D)** Forest plot for univariate and multivariate linear analysis of MMRd/MSI signature with the four MMR-related gene expression. Of note, *MSH6* expression was not analyzed in multivariate regression model given its tight correlation with *MSH2* expression. ****$P < 0.0001$.

high *MSH2* expression displayed a significant immunotherapy-responsive microenvironment in LUAD.

Of particular note, one LUAD patient who derived durable clinical benefit from anti-PD-1 therapy showed the strong staining of both PD-L1 expression and CD8+ T cell infiltration. Moreover, this patient also displayed the strongest staining of MSH2 expression among the four key MMR proteins, which directly supported our hypothesis that *MSH2* expression might be a potential surrogate biomarker of TMB to predict ICB response in LUAD [**Figures 7E,F**; raw IHC data retrieved from Dong et al. (10)].

DISCUSSION

ICB-based therapies targeting CTLA-4 or PD-1 have shown a promising future in multiple cancer types, but the molecular mechanism between them is completely different (35). Additionally, anti-PD-1 therapy performs much better than anti-CTLA-4 therapy on the efficacy, survival, and adverse events (5, 35). Therefore, this study mainly focused on LUAD for anti-PD-1 therapy.

TMB is one of the most important biomarkers for predicting ICB response in NSCLC (1–3, 7), and it also shows predictive efficacy for ICB therapies in other types of solid tumors (7, 36), but it still has some limitations (5, 6, 14, 19). For example, the cut-offs of TMB for identifying ICB responders are different for different tumor types, and the test platform for assessing TMB has also not been standardized (5, 14, 19). Thus, more studies are turning to develop surrogate biomarkers that highly correlate with the TMB status, such as genetic mutations of DNA damage response pathways and *TP53/KRAS* (10, 19, 37). However, these mutations are positive for only a minority of patients, and the broad detection of these TMB-related gene mutations in clinical practice remains challenging (19). Titin (*TTN*) is the longest gene within the whole genome, and its mutations have also been proposed as a surrogate TMB biomarker for predicting ICB response in solid tumors (19). However, *TTN* mutations are not the cause of high TMB in tumors, and its mutations also account for a small cohort of candidates (29.68%) (19, 38). In addition, targeted panel sequencing, such as MSK-IMPACT panel, has also been developed and approved by US Food and Drug Administration (FDA) to estimate TMB, but

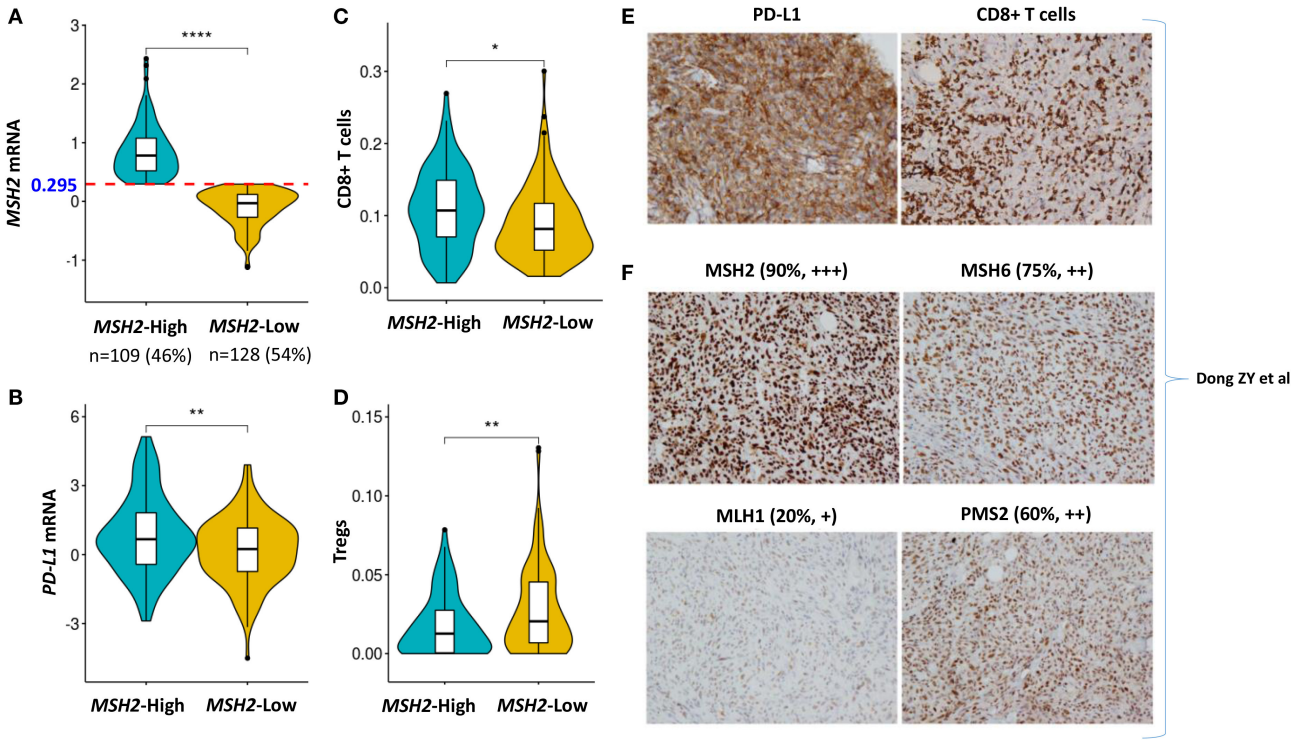


FIGURE 7 | High expression of *MSH2* highly correlated with increased *PD-L1* expression and CD8+ T cell infiltration in the *Validation-LUAD* cohort. **(A)** Samples in the *Validation-LUAD* cohort were stratified into two groups based on the Youden index-associated cutoff of *MSH2* expression from **Figure 5B**. *MSH2*-High: ≥0.295 ($n = 109$); *MSH2*-Low: <0.295 ($n = 128$); two samples whose expression profiles were missing. **(B–D)** Quantitative analysis of *PD-L1* expression and the infiltration of immune cells in individual tumor tissues based on *MSH2* expression. **(E,F)** IHC staining of PD-L1, CD8, and four MMR- related genes in one LUAD patient who derived durable clinical benefit from anti-PD-1 therapy. raw IHC data retrieved from Dong et al. (10). ****$P < 0.0001$, **$P < 0.01$, *$P < 0.05$.

its predictive accuracy for ICB response is significantly lower than WES in LUAD (**Figure 2B**), which suggests that it still needs more optimizations. Moreover, blood-based TMB (bTMB) is being developed to predict ICB response in NSCLC, while blood samples for detecting bTMB must contain the mutated circulating tumor DNA (ctDNA) that must be shed from the tumor, which has limited its clinical application (39).

Unlike previous studies that focused on genomic variations (10, 19, 37, 40), we turned to the transcriptomic landscape to explore the potential surrogates of TMB in LUAD. We revealed that *MSH2* was the most robust MMR gene whose expression significantly correlated with increased TMB in LUAD. Mechanistically, given the low mutation rates of MMR-related genes in LUAD, the transcriptome-based dysfunction of the MMR machinery is more likely to be the cause of high TMB. Therefore, this study mainly focuses on MMR-related gene expression. The other nine genes (*CDCA5*, *MCM10*, *GINS4*, *KIAA1524*, *KIF2C*, *NUF2*, *CDC20*, *CDC7*, *THOC4*; **Figure 3A**; **Table S2**) showing better performance than *MSH2* may also be potential TMB indicators in LUAD, which still needs more mechanistic investigations.

MMRd testing has primarily been developed and tested in patients with colorectal and endometrial cancer to predict ICB response given their relatively high positive rates (13, 30, 33, 41). Mechanistically, tumors with MMRd are often hypermutated and can result in microsatellite instability (MSI) within the genome. Therefore, MSI has been proposed as a marker of MMRd in previous studies (13, 30, 33). However, in contrast to the findings in colorectal and endometrial cancer (13, 26), we found that MMRd/MSI signature was significantly low in LUAD with high TMB (**Figure 6B**; **Figures S6A–D**), suggesting that genome-based MMRd/MSI might not cause a high mutation load in LUAD. In addition, we observed that high *MSH2* expression showed a significantly inverse association with increased MMRd/MSI signature in LUAD (**Figure 6D**). Given that the relationships between MMRd/MSI and TMB are complex and different for different tumor types (13), more mechanistic investigations are required to illuminate these results in LUAD.

In previous studies (13, 30, 33), all four MMR proteins (MSH2, MSH6, PMS2, and MLH1) were always detected together by IHC to determine the MMRd status within a tumor. However, the IHC method was used to indirectly infer the mutation status of the four MMR genes (13, 30, 33). Of particular note, in LUAD, the IHC-based method has rarely been used to detect the MMRd status (13, 30, 33).

Importantly, our data revealed that a transcriptome-based, not genome-based, MMRd mechanism widely existed in LUAD,

which might partly illuminate the cause of high TMB in LUAD. The potential mechanism is that both MSH2 and MLH1 proteins are essential partners for forming the MMR machinery (13, 30). However, we observed that TMB significantly and positively correlated with *MSH2* expression but inversely correlated with *MLH1* expression in LUAD (**Figures 4E,F**; **Figures S2B,C**). Moreover, *MLH1* was the only gene whose expression was significantly down-regulated in LUAD tissues compared with the other three MMR genes (**Figures S8A–D**). *MLH1* expression may be suppressed by its promoter methylation in LUAD (33, 42, 43). These results suggested that down-regulated *MLH1* expression might impair the MMR machinery to repair the damaged genome and then caused more mutation load in LUAD.

One limitation of this study is that we only collected one LUAD sample showing the direct evidence that MSH2 expression alone could be a surrogate TMB biomarker to predict ICB response in LUAD. Because of the lack of public LUAD data, we could not directly validate this result in a large cohort. However, more prospective clinical trials are required to validate this correlation. Another limitation is that the TMB cutoff (TMB = 166.5) for stratifying LUAD samples was based on a small number of samples, which still needs large cohorts to determine. However, given the intratumoral heterogeneity across different cancer subtypes (1–3, 7, 16, 17), our data are much more homogeneous and thus more reliable to find the specific biomarkers benefiting the specific patients. Additionally, we also recruited another independent cohort (*Validation-LUAD* cohort) to validate our conclusion and proposed a mechanistic connection between *MSH2* expression and increased TMB in LUAD.

Since ICB therapies are associated with specific adverse events, it is profound to identify predictive biomarkers to select patients who are more likely to derive the maximum benefits from ICB therapies. Therefore, more multi-omic datasets are indispensable to explore and improve the efficacy of immunotherapies. It is possible that *MSH2* expression can be applied jointly with other factors to acquire a greater prediction performance, which is already suggested that combining multiple biomarkers are more robust than a single analyte for predicting ICB efficacy (7, 10, 12, 17, 29, 39).

In summary, our data suggest that *MSH2* expression highly correlates with increased TMB and the immunotherapy-responsive microenvironment in LUAD. Prospective clinical trials are required to further confirm these results.

DATA AVAILABILITY STATEMENT

TCGA-LUAD datasets were extracted from the UCSC Xena multi-omics database platform (https://xenabrowser.net/datapages/).

AUTHOR CONTRIBUTIONS

MJ initiated and performed the analysis. MJ, LY, and QY wrote and revised the manuscript. TC supervised the project.

FUNDING

This work was supported by the start-up package from ShanghaiTech University (TC).

ACKNOWLEDGMENTS

The authors thank the TCGA project and other groups for providing invaluable datasets for statistical analyses.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2020.00168/full#supplementary-material

REFERENCES

1. Rizvi NA, Hellmann MD, Snyder A, Kvistborg P, Makarov V, Havel JJ, et al. Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science.* (2015) 348:124–8. doi: 10.1126/science.aaa1348
2. Hellmann MD, Nathanson T, Rizvi H, Creelan BC, Sanchez-Vega F, Ahuja A, et al. genomic features of response to combination immunotherapy in patients with advanced non-small-cell lung cancer. *Cancer Cell.* (2018) 33:843–52.e4. doi: 10.1016/j.ccell.2018.03.018
3. Rizvi H, Sanchez-Vega F, La K, Chatila W, Jonsson P, Halpenny D, et al. Molecular determinants of response to anti-programmed cell death (PD)-1 and anti-programmed death-ligand 1 (PD-L1) blockade in patients with non-small-cell lung cancer profiled with targeted next-generation sequencing. *J Clin Oncol.* (2018) 36:633–41. doi: 10.1200/JCO.2017.75.3384
4. Mahoney KM, Rennert PD, Freeman GJ. Combination cancer immunotherapy and new immunomodulatory targets. *Nat Rev Drug Discov.* (2015) 14:561–84. doi: 10.1038/nrd4591
5. Havel JJ, Chowell D, Chan TA. The evolving landscape of biomarkers for checkpoint inhibitor immunotherapy. *Nat Rev Cancer.* (2019) 19:133–50. doi: 10.1038/s41568-019-0116-x
6. Chan TA, Yarchoan M, Jaffee E, Swanton C, Quezada SA, Stenzinger A, et al. Development of tumor mutation burden as an immunotherapy biomarker: utility for the oncology clinic. *Ann Oncol.* (2019) 30:44–56. doi: 10.1093/annonc/mdy495
7. Samstein RM, Lee CH, Shoushtari AN, Hellmann MD, Shen R, Janjigian YY, et al. Tumor mutational load predicts survival after immunotherapy across multiple cancer types. *Nat Genet.* (2019) 51:202–6. doi: 10.1038/s41588-018-0312-8
8. Herbst RS, Soria JC, Kowanetz M, Fine GD, Hamid O, Gordon MS, et al. Predictive correlates of response to the anti-PD-L1 antibody MPDL3280A in cancer patients. *Nature.* (2014) 515:563–7. doi: 10.1038/nature14011
9. Tang H, Wang Y, Chlewicki LK, Zhang Y, Guo J, Liang W, et al. Facilitating T cell infiltration in tumor microenvironment overcomes resistance to PD-L1 blockade. *Cancer Cell.* (2016) 30:500. doi: 10.1016/j.ccell.2016.08.011
10. Dong ZY, Zhong WZ, Zhang XC, Su J, Xie Z, Liu SY, et al. Potential predictive value of TP53 and KRAS mutation status for response to PD-1 blockade immunotherapy in lung adenocarcinoma. *Clin Cancer Res.* (2017) 23:3012–24. doi: 10.1016/j.jtho.2016.11.504
11. Diaz LA Jr., Le DT. PD-1 blockade in tumors with mismatch-repair deficiency. *N Engl J Med.* (2015) 373:1979. doi: 10.1056/NEJMc1510353

12. Topalian SL, Taube JM, Anders RA, Pardoll DM. Mechanism-driven biomarkers to guide immune checkpoint blockade in cancer therapy. *Nat Rev Cancer.* (2016) 16:275–87. doi: 10.1038/nrc.2016.36
13. Luchini C, Bibeau F, Ligtenberg MJL, Singh N, Nottegar A, Bosse T, et al. ESMO recommendations on microsatellite instability testing for immunotherapy in cancer, and its relationship with PD-1/PD-L1 expression and tumour mutational burden: a systematic review-based approach. *Ann Oncol.* (2019) 30:1232–43. doi: 10.1093/annonc/mdz116
14. Addeo A, Banna GL, Weiss GJ. Tumor mutation burden-from hopes to doubts. *JAMA Oncol.* (2019) 5:934–5. doi: 10.1001/jamaoncol.2019.0626
15. Tumeh PC, Harview CL, Yearley JH, Shintaku IP, Taylor EJ, Robert L, et al. PD-1 blockade induces responses by inhibiting adaptive immune resistance. *Nature.* (2014) 515:568–71. doi: 10.1038/nature13954
16. Cristescu R, Mogg R, Ayers M, Albright A, Murphy E, Yearley J, et al. Pan-tumor genomic biomarkers for PD-1 checkpoint blockade-based immunotherapy. *Science.* (2018) 362:eaar3593. doi: 10.1126/science.aar3593
17. Jiang P, Gu S, Pan D, Fu J, Sahu A, Hu X, et al. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat Med.* (2018) 24:1550–8. doi: 10.1158/2326-6074.CRICIMTEATIAACR18-B077
18. Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang TH, et al. The immune landscape of cancer. *Immunity.* (2018) 48:812–30.e14. doi: 10.1016/j.immuni.2018.03.023
19. Jia Q, Wang J, He N, He J, Zhu B. Titin mutation associated with responsiveness to checkpoint blockades in solid tumors. *JCI Insight.* (2019) 4:e127901. doi: 10.1172/jci.insight.127901
20. Herbst RS, Morgensztern D, Boshoff C. The biology and management of non-small cell lung cancer. *Nature.* (2018) 553:446–54. doi: 10.1038/nature25183
21. Cancer Genome Atlas Research N. Comprehensive molecular profiling of lung adenocarcinoma. *Nature.* (2014) 511:543–50. doi: 10.1038/nature13385
22. Goldman M, Craft B, Hastie M, Repečka K, Kamath A, McDade F, et al. The UCSC Xena platform for public and private cancer genomics data visualization and interpretation. *bioRxiv [Preprint].* (2019). doi: 10.1101/326470
23. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol.* (2013) 31:213–9. doi: 10.1038/nbt.2514
24. Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res.* (2018) 28:1747–56. doi: 10.1101/gr.239244.118
25. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA.* (2005) 102:15545–50. doi: 10.1073/pnas.0506580102
26. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. *Nature.* (2013) 500:415–21. doi: 10.1038/nature12477
27. Kim J, Mouw KW, Polak P, Braunstein LZ, Kamburov A, Kwiatkowski DJ, et al. Somatic ERCC2 mutations are associated with a distinct genomic signature in urothelial tumors. *Nat Genet.* (2016) 48:600–6. doi: 10.1038/ng.3557
28. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods.* (2015) 12:453–7. doi: 10.1038/nmeth.3337
29. Ayers M, Lunceford J, Nebozhyn M, Murphy E, Loboda A, Kaufman DR, et al. IFN-gamma-related mRNA profile predicts clinical response to PD-1 blockade. *J Clin Invest.* (2017) 127:2930–40. doi: 10.1172/JCI91190
30. de la Chapelle A, Hampel H. Clinical relevance of microsatellite instability in colorectal cancer. *J Clin Oncol.* (2010) 28:3380–7. doi: 10.1200/JCO.2009.27.0652
31. Carbone DP, Reck M, Paz-Ares L, Creelan B, Horn L, Steins M, et al. First-line nivolumab in stage IV or recurrent non-small-cell lung cancer. *N Engl J Med.* (2017) 376:2415–26. doi: 10.1056/NEJMoa1613493
32. Vokes EE, Ready N, Felip E, Horn L, Burgio MA, Antonia SJ, et al. Nivolumab versus docetaxel in previously treated advanced non-small-cell lung cancer (CheckMate 017 and CheckMate 057): 3-year update and outcomes in patients with liver metastases. *Ann Oncol.* (2018) 29:959–65. doi: 10.1093/annonc/mdy041
33. Baretti M, Le DT. DNA mismatch repair in cancer. *Pharmacol Ther.* (2018) 189:45–62. doi: 10.1016/j.pharmthera.2018.04.004
34. Villanueva MT. Cancer immunotherapy: macrophages steal the show. *Nat Rev Drug Discov.* (2017) 16:455. doi: 10.1038/nrd.2017.126
35. Chen L, Han X. Anti-PD-1/PD-L1 therapy of human cancer: past, present, and future. *J Clin Invest.* (2015) 125:3384–91. doi: 10.1172/JCI80011
36. Yarchoan M, Hopkins A, Jaffee EM. Tumor mutational burden and response rate to pd-1 inhibition. *N Engl J Med.* (2017) 377:2500–1. doi: 10.1056/NEJMc1713444
37. Wang J, Wang Z, Zhao J, Wang G, Zhang F, Zhang Z, et al. Co-mutations in DNA damage response pathways serve as potential biomarkers for immune checkpoint blockade. *Cancer Res.* (2018) 78:6486–96. doi: 10.1158/0008-5472.CAN-18-1814
38. Maruvka YE, Haradhvala NJ, Getz G. Analyzing frequently mutated genes and the association with tumor mutation load. *JAMA Oncol.* (2019) 5:577. doi: 10.1001/jamaoncol.2019.0127
39. Gandara DR, Paul SM, Kowanetz M, Schleifman E, Zou W, Li Y, et al. Blood-based tumor mutational burden as a predictor of clinical benefit in non-small-cell lung cancer patients treated with atezolizumab. *Nat Med.* (2018) 24:1441–8. doi: 10.1038/s41591-018-0134-3
40. Lyu GY, Yeh YH, Yeh YC, Wang YC. Mutation load estimation model as a predictor of the response to cancer immunotherapy. *NPJ Genom Med.* (2018) 3:12. doi: 10.1038/s41525-018-0051-x
41. Latham A, Srinivasan P, Kemel Y, Shia J, Bandlamudi C, Mandelker D, et al. Microsatellite instability is associated with the presence of lynch syndrome pan-cancer. *J Clin Oncol.* (2019) 37:286–95. doi: 10.1200/JCO.18.00283
42. Seng TJ, Currey N, Cooper WA, Lee CS, Chan C, Horvath L, et al. DLEC1 and MLH1 promoter methylation are associated with poor prognosis in non-small cell lung carcinoma. *Br J Cancer.* (2008) 99:375–82. doi: 10.1038/sj.bjc.6604452
43. Simpkins SB, Bocker T, Swisher EM, Mutch DG, Gersell DJ, Kovatich AJ, et al. MLH1 promoter methylation and gene silencing is the primary cause of microsatellite instability in sporadic endometrial cancers. *Hum Mol Genet.* (1999) 8:661–6. doi: 10.1093/hmg/8.4.661

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

SYSTEMATIC REVIEW
published: 03 March 2020
doi: 10.3389/fonc.2020.00290

Diagnostic Accuracy of Droplet Digital PCR and Amplification Refractory Mutation System PCR for Detecting EGFR Mutation in Cell-Free DNA of Lung Cancer: A Meta-Analysis

Caichen Li†, Qihua He†, Hengrui Liang†, Bo Cheng, Jianfu Li, Shan Xiong, Yi Zhao, Minzhang Guo, Zhichao Liu, Jianxing He and Wenhua Liang**

China State Key Laboratory of Respiratory Disease and National Clinical Research Center for Respiratory Disease, The First Affiliated Hospital of Guangzhou Medical University, Guangzhou, China

OPEN ACCESS

Edited by:
Christian Rolfo, University of Maryland Medical System, United States

Reviewed by:
Laura Mezquita, Institut Gustave Roussy, France
Marzia Del Re, University of Pisa, Italy

***Correspondence:**
Jianxing He
drjianxing.he@gmail.com
Wenhua Liang
liangwh1987@163.com

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Thoracic Oncology, a section of the journal Frontiers in Oncology

Received: *17 September 2019*
Accepted: *19 February 2020*
Published: *03 March 2020*

Citation:
Li C, He Q, Liang H, Cheng B, Li J, Xiong S, Zhao Y, Guo M, Liu Z, He J and Liang W (2020) Diagnostic Accuracy of Droplet Digital PCR and Amplification Refractory Mutation System PCR for Detecting EGFR Mutation in Cell-Free DNA of Lung Cancer: A Meta-Analysis. Front. Oncol. 10:290. doi: 10.3389/fonc.2020.00290

Background: Epidermal growth factor receptor (EGFR) mutation testing in plasma cell-free DNA (cfDNA) from advanced lung cancer patients is an emerging clinical tool. This meta-analysis was designed to determine the diagnostic accuracy of two common PCR systems, droplet digital PCR (ddPCR) and amplification refractory mutation system PCR (ARMS-PCR), for detecting EGFR mutation in cfDNA.

Materials and methods: A systematic search was carried out based on PubMed, Web of science, Embase and the Cochrane library. Data from eligible studies were extracted and pooled to calculate the sensitivity, specificity, diagnostic odds ratio (DOR), area under the summary receiver-operating characteristic curve (AUROC), using tissue biopsy results as the standard method. Subgroup analyses were performed regarding EGFR mutation type, tumor stage, and EGFR-TKI treatment.

Results: Twenty-five studies involving 4,881 cases were included. The plasma testing sensitivity, specificity, DOR, and AUROC, compared with the matched tumor tissues, were 72.1%, 95.6%, 38.5, 0.89 for ddPCR, and 65.3%, 98.2%, 52.8, 0.71 for ARMS-PCR, respectively, through indirect comparison, significant differences were found in sensitivity ($P = 0.003$) and specificity ($P = 0.007$). Furthermore, significant difference was found in sensitivity between tumor stage subgroups (IIIB–IV subgroup vs. IA–IV subgroup) in ARMS-PCR (73.7 vs. 64.2%, $P = 0.008$), but not in ddPCR (72.5 vs. 71.2%, $P = 0.756$).

Conclusions: This study demonstrates that ddPCR and ARMS-PCR have a high specificity with a practical sensitivity for detecting EGFR mutation in cfDNA, which supports their application as a supplement or a conditional-alternative to tissue biopsy in clinical practice for genotyping. It seems that ddPCR has a higher sensitivity than ARMS-PCR, especially in early stages.

Keywords: lung cancer, droplet digital PCR (ddPCR), amplification refractory mutation system PCR (ARMS-PCR), cell free DNA (cfDNA), epidermal growth factor receptor (EGFR)

INTRODUCTION

Lung cancer remains the most frequently diagnosed cancer and the leading cause of cancer-related mortality worldwide, with 85% of patients having non-small-cell lung cancer (NSCLC) (1–3). Fortunately, accurate gene analysis of epidermal growth factor receptor (EGFR) mutation in advanced NSCLC patients has provided them great opportunities to receive optimal treatments. Successful analysis of genotyping plays an important role in this process (4, 5). Conventionally, detection of EGFR mutation status in tumor tissue is the standard approach, which can be obtained by tissue biopsy or surgery (6). However, tissue samples are not always available or sufficient in quantity for genotyping. Furthermore, tissue biopsy-related complications are common, such as pneumothorax and hemoptysis (7).

Liquid biopsy is emerging as an important clinical tool and has significant potential to offer a supplement or a conditional alternative to tissue biopsy for tumor genotyping (6). Liquid biopsy offers the advantages of being non-invasive, easily accessible, and can be performed repeatedly (8). Presently, cell-free DNA is available for liquid biopsy (9). Mature testing platforms of EGFR mutation include next generation sequencing technologies (NGS), digital platforms [droplet digital PCR (ddPCR), Beads, Emulsions, Amplification, and Magnetic (BEAMing)] and real time PCR [Cobas, Amplification Refractory Mutation System (ARMS-PCR)]. Thress et al. demonstrated that the Cobas EGFR Mutation Test and BEAMing dPCR had high sensitivity (82–87%) and specificity (97%) for EGFR-sensitizing mutations (10). Moreover, Feng et al. indicated that the sensitivity of ddPCR was similar with ARMS in plasma EGFR detection (80.4 vs. 76.5%), as was the specificity (89.3 vs. 100%) (11). These findings showed the high sensitivity and specificity of PCR platforms, suggesting that EGFR mutations can be accurately detected in cfDNA. In addition, the PCR-based methods had the advantages of being both rapid and inexpensive, and suitable for detection of specific point mutations (12).

Several studies have reported promising results detecting EGFR mutation from cfDNA of patients with lung cancer using ddPCR and ARMS-PCR (11, 13–15). The question of interest is whether these tissue-free methods are sufficiently accurate to be considered a supplement or even alternative to tissue genotyping. Therefore, we conducted this meta-analysis to determine the diagnostic accuracy of the ddPCR system and the ARMS-PCR system for detecting EGFR mutation in cfDNA, using tissue biopsy results as the standard detection modality.

MATERIALS AND METHODS

This meta-analysis was conducted according to the PRISMA Checklist.

Literature and Search Strategy

All potentially relevant studies were retrieved through search of PubMed, Web of science, Embase and the Cochrane library databases up to Dec 1, 2019, using a combination of key words "lung cancer," "EGFR," "droplet digital PCR," and "amplification refractory mutation system PCR." No search limitations were set. The previous published articles and reviews were inspected to identify studies not included by the initial search. This study is registered with PROSPERO, number CRD42019120049.

Inclusion and Exclusion Criteria

Eligible studies should meet the following criteria: (i) included patients with lung cancer diagnosed by histopathology or cytologically; (ii) studied diagnostic accuracy of ddPCR or ARMS-PCR for detecting EGFR sensitivity mutation based on cfDNA or ctDNA; (iii) EGFR mutation statuses were compared with tumor tissues.

Studies were excluded if (i) data was insufficient to calculate the sensitivity or specificity for this meta-analysis, (ii) they were review articles, abstracts, case reports, commentary articles, editorials, expert opinions, non-comparative studies, letters, unrelated to research topics, or duplicate reports.

Data Extraction

Data were extracted independently by two reviewers (Li C.C. and Liang H.R.), and conflicts were resolved by a third reviewer (He Q.H.). For the selected studies, the name of first author, year of publication, country of origin, sample size, basic characteristics of studied population, clinical stage, tumor histology, percentage of TKI-naïve, and TKI-treated patients, techniques used for EGFR mutation detection for both tissue sample and cfDNA, true positive (TP), false positive (FP), false negative (NP), and true negative (TN) were collected from eligible studies. Subgroup analyses, and comparison of two PCR platforms were conducted according to EGFR mutation type, tumor stage and EGFR-TKI treatment, respectively.

Quality Assessment

Quality assessment of diagnostic accuracy studies 2 (QUADAS-2) is a tool used to evaluate the quality of primary diagnostic accuracy studies, including patient selection, index tests, reference standard, and flow and timing. QUADAS-2 was evaluated by two reviewers (CL and HL).

Statistical Analysis

Sensitivity, specificity, diagnostic odds ratio (DOR) and the area under the summary receiver-operating characteristic curve (AUROC) were pooled. The value of DOR is calculated by the positive likelihood ratio (PLR)/the negative likelihood ratio (NLR), and its value ranges from 0 to infinity, with higher value indicating better discriminatory test performance.

We use Cochrane's Q and the I^2 statistic to examine the heterogeneity. $P \leq 0.05$ and $I^2 \geq 50\%$ mean that significant heterogeneity existed in pooled statistics. In addition, publication bias was detected by the Deek's funnel plot asymmetry test and $P < 0.05$ indicated the presence of publication bias.

The analysis was performed with STATA 13.0 software (STATA corporation, College Station, TX, USA) with the MIDAS module and Meta-Disc 1.4 (Ramón y Cajal Hospital in Madrid, Spain).

RESULTS

Study Selection and Quality Assessment

A total of 396 records were screened according to the search strategy up to Dec 1, 2019. Finally, 25 full-text articles were identified and reviewed. Of the included articles, 14 studied ddPCR (10, 11, 15–26) and 16 studied ARMS-PCR (10, 11, 13–15, 20, 22, 27–35). Specifically, five articles made direct comparisons between ddPCR and ARMS-PCR (10, 11, 15, 20, 22). **Figure 1** summarized the flow chart. The quality assessment of each study is summarized in **Table S1**.

Characteristics of Included Studies

Twenty-five studies involving 4,881 cases were identified and included for analysis. The majority of patients were Asians with advanced NSCLC. To assess ddPCR performance of cfDNA-based EGFR mutation detection, a total of 1,105 samples were used for testing EGFR mutation and compared with the result of

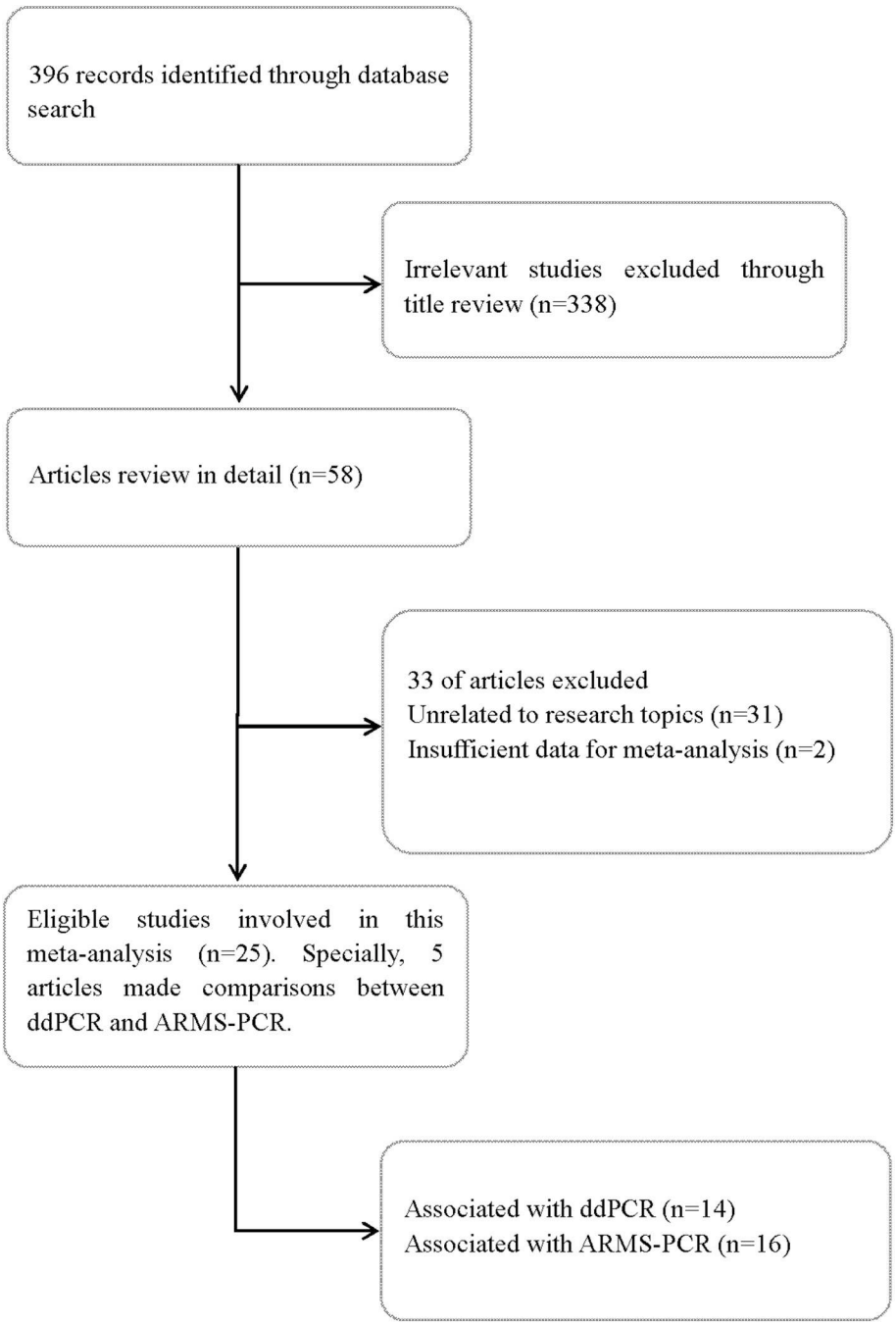


FIGURE 1 | Flow diagram detailing the search strategy of the included studies in this meta-analysis.

tissue biopsy. Similarly, a total of 3,950 samples were tested and compared with tissue biopsy to assess diagnostic performance of ARMS-PCR. **Table 1** summarized the characteristics of all relevant studies. It should be noted that, some studies only presented the results of mutation in exon 19 deletion and L858R, rather than total mutations level of EGFR. Therefore, we added the sample number of mutations of exon 19 deletion and L858R together to get an overall result of EGFR mutation in each study.

Overall Accuracy of the ddPCR and ARMS-PCR Test

The plasma testing sensitivity, specificity, DOR, and AUROC, compared with the matched tumor tissues, were 72.1% (95% CI, 68.2–75.7%), 95.6% (95% CI, 93.5–97.1%), 38.5 (95% CI, 22.3–66.4), 0.89 (95% CI, 0.83–0.95) for ddPCR, and 65.3% (95% CI, 62.9–67.6%), 98.2% (95% CI, 97.6–98.7%), 52.8 (95% CI, 26.3–106.1), 0.71 (95% CI, 0.52–0.91) for ARMS-PCR, respectively (**Figure 2**). There was no publication bias for outcome measures with asymmetrical appearance on funnel plot analysis (**Figure 3**), and $P > 0.05$ in Deek's test.

Comparison of ddPCR and ARMS-PCR in Different Subgroups

Subsequently, results of the two platforms in different EGFR-sensitizing mutations, tumor stages and EGFR-TKI treatment status were assessed by stratified analysis (**Table 2**). Significant difference in sensitivity was found between tumor stage subgroups (IIIB–IV subgroup vs. IA–IV subgroup) in ARMS-PCR (73.7 vs. 64.2%, $P = 0.008$), but not in ddPCR (72.5 vs. 71.2%, $P = 0.756$).

Indirect and Direct Comparison of ddPCR and ARMS-PCR

Twenty five full-text articles were included in the indirect comparison and the detailed characteristics of clinical stage of the enrolled patients are summarized in **Table S2**. In studies indirectly comparing the two PCR systems, there was a significant difference in sensitivity ($P = 0.003$) and in specificity ($P = 0.007$) (**Table 3**). We performed indirect comparison about sensitivity between ddPCR and ARMS-PCR systems, better sensitivities for ddPCR were observed in stage IIIB–IV (73.4 vs. 62.9%, $P = 0.012$), TKI-naïve (72.7 vs. 64.5%, $P = 0.040$), and TKI-treated (75.6 vs. 65.5%, $P = 0.035$) patients. Compared to ARMS-PCR, more favorable specificity was found in the TKI-treated

TABLE 1 | Characteristics of included studies.

Study	Country	Sample size	Age	Female (%)	Smoker (%)	Histology	Clinical stage	TKI naïve (%)
Ishii et al. (16)	Japan	18	50-81	89	6	NSCLC	Recurrence	0
Lee et al. (18)	Korea	81	32-81	62	37	NSCLC	IV/recurrence	0
Sacher et al. (19)	US	180	18+	62	NA	NSCLC	IIIB/IV/recurrence	0, 100[†]
Thress et al. (10)	UK	38	NA	NA	NA	NSCLC	Advanced	0
Feng et al. (11)	China	79	30–75	54	32	NSCLC	Advanced	100
Xu et al. (15)	China	20	37–76	50	30	NSCLC	I–IV	40
Zhang et al. (22)	China	122	30–85	47	42	NSCLC	III–IV	100
Wang et al. (20)	China	65	32–85	38	48	LC	I—IV/uncertain	NA
Yu et al. (21)	China	22	35–74	54	NA	NSCLC	IIIB–IV	86
Zhang et al. (23)	China	215	NA	41	44	NSCLC	IIIB–IV	100
Zhu et al. (17)	China	86	28–81	35	NA	NSCLC	IIIB–IV	100
Zhu et al. (24)	China	51	60.89 ± 1.48	39	55.8	NSCLC	I–IV	56.9
Li et al. (32)	China	109	NA	53	33	NSCLC	IIIB–IV	96.3
Cui et al. (34)	China	180	37–77	48	NA	NSCLC	IIIB–IV	70
Douillard et al. (14)	France	1060	32–82	71	39	LC	IIIA–IV	100
Duan et al. (30)	China	94	58 ± 11	35	51	LC	IIA–IV	100
Li et al. (29)	China	164	32–81	41.5	48.8	LC	IIB–IV/recurrence	58.5
Liu et al. (28)	China	86	28–81	35	55	NSCLC	IIIB–IV	NA
Ma et al. (31)	China	219	26–81	34	49	LC	IIIA–IV	100
Su et al. (35)	China	107	29–81	58	15	NSCLC	I–IV	73.8
Wan et al. (13)	China	2463	NA	NA	NA	LC	I–IV	0
Xu et al. (27)	China	51	25–77	39	37	NSCLC	IIIB–IV	100
Zhou et al. (33)	China	447	27–86	45	47	LC	I–IV	98.2
Jiang et al. (25)	China	50	NA	NA	NA	NSCLC	NA	100
Guo et al. (26)	China	201	NA	52.2	NA	NSCLC	I–IV	66.2

[†] *Patients of this study have divided into two groups according to TKI used. TKI, tyrosine kinase inhibitors; LC, lung cancer; NSCLC, non-small-cell lung cancer.*

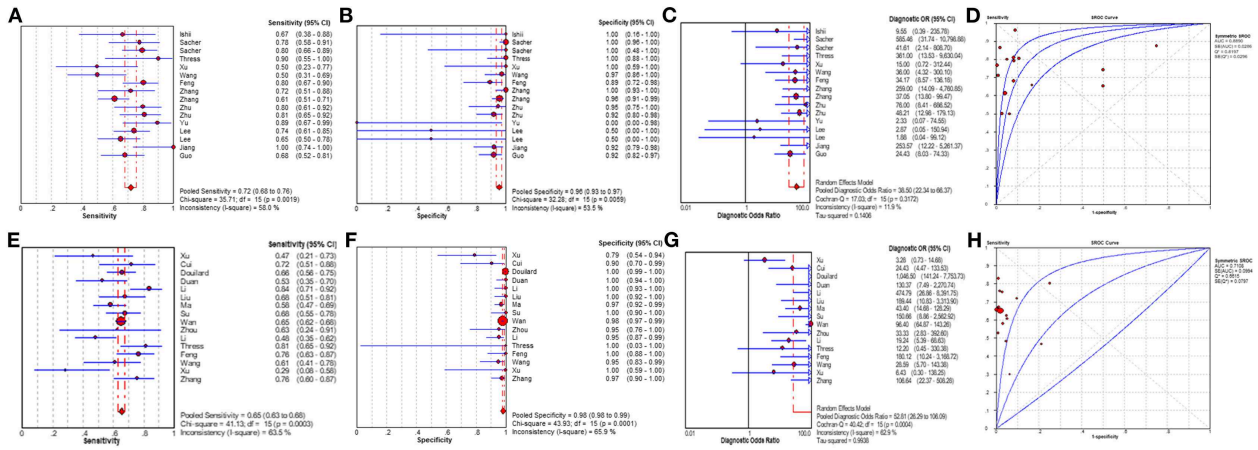


FIGURE 2 | The results of meta-analysis. **(A)** sensitivity, **(B)** specificity, **(C)** diagnostic odds ratio, and **(D)** SROC curve for ddPCR; **(E)** sensitivity, **(F)** specificity, **(G)** diagnostic odds ratio, and **(H)** SROC curve for ARMS-PCR. Two articles of ddPCR had two status including prior treatment group and disease progression group. Add 0.5 to all cells of the studies with zero.

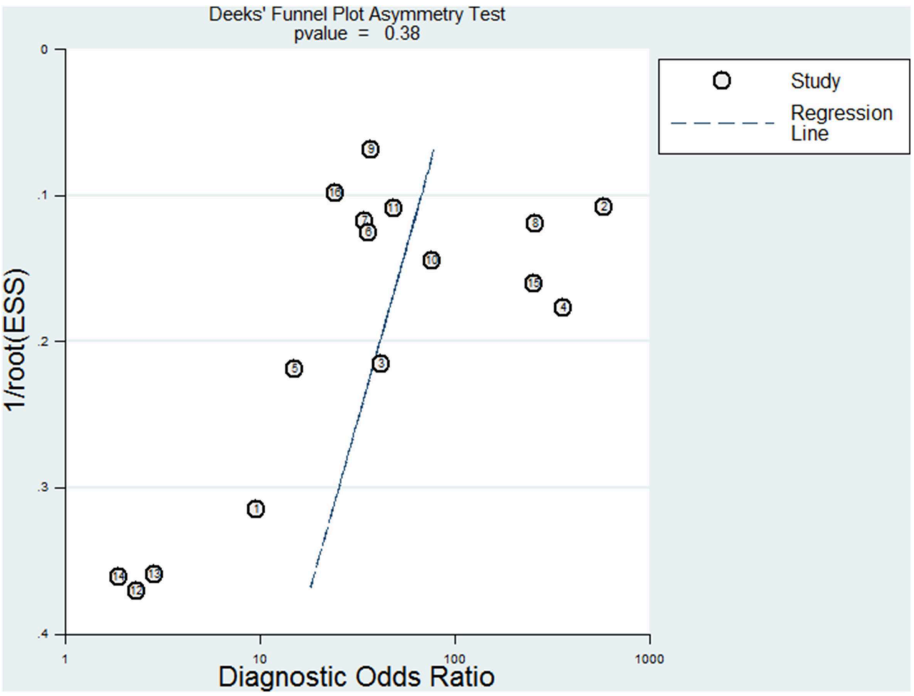


FIGURE 3 | Assessment of publication bias by Deek's funnel plot asymmetry test in ddPCR system.

subgroup using ddPCR (93.5 vs. 98.0%, $P = 0.038$). In studies simultaneously comparing the two platforms, however, we observed no significant difference in specificity (97.3 vs. 98.7%, $P = 0.473$) and sensitivity (69.3 vs. 69.0%, $P = 0.960$) between ddPCR and ARMS-PCR, regardless of EGFR mutation type and EGFR-TKI treatment.

DISCUSSION

Precise detection of EGFR mutation in lung cancer can allow clinicians to assign patients to highly specific treatment, especially for those with EGFR-sensitizing mutations as a series of clinical trials has proven (4, 5). Many retrospective studies have reported that patients with ctDNA-based EGFR mutation status have better clinical outcomes with EGFR-TKIs than those without EGFR mutation (14, 36). In a prospective clinical trial reported by Wang et al., detection of EGFR mutation in ctDNA was a selection method to provide patients with appropriate first-line gefitinib treatment, providing more evidence to guide treatment decisions for those patients with advanced lung cancer who cannot obtain tumor tissue samples (37). It is interesting and meaningful to answer whether these tissue-free methods

TABLE 2 | The results of meta-analysis.

Index method	Included studies	Sensitivity	I^2	*P*	Specificity	I^2	*P*	DOR	I^2	*P*	AUROR
ddPCR[†]	14	72.1% (95% CI, 68.2–75.7%)	58%	0.002	95.6% (95% CI, 93.5–97.1%)	53.5%	0.006	38.5 (95% CI, 22.3–66.4)	11.9%	0.317	0.89 (95% CI, 0.83–0.95)
ARMS-PCR[†]	16	65.3% (95% CI, 62.9–67.6%)	63.5%	0.001	98.2% (95% CI, 97.6–98.7%)	65.9%	0.001	52.8 (95% CI, 26.3–106.1)	62.9%	0.001	0.71 (95% CI, 0.52–0.91)
SUBGROUP ANALYSIS OF EXON 19 DELETION IN TWO SYSTEMS[‡]											
ddPCR	11	72.9% (95% CI, 67.2–78.2%)	46.0%	0.035	99.1% (95% CI, 98.2–99.7%)	18.0%	0.262	179.6 (95% CI, 85.9–375.5)	0%	0.997	0.97 (95% CI, 0.94–1.00)
ARMS-PCR	11	66.3% (95% CI, 60.9–71.3%)	44.3%	0.056	99.3% (95% CI, 98.6–99.7%)	68.9%	0.001	113.7 (95% CI, 39.9–323.4)	53.4%	0.018	0.65 (95% CI, 0.25–1.00)
SUBGROUP ANALYSIS OF L858R IN TWO SYSTEMS[§]											
ddPCR	12	69.7% (95% CI, 63.3–75.5%)	45.6%	0.032	98.2% (95% CI, 97.0–99.0%)	31.3%	0.125	96.5 (95% CI, 53.2–175.2)	0%	0.872	0.96 (95% CI, 0.91–1.00)
ARMS-PCR	11	61.6% (95% CI, 54.9–68.0%)	51.1%	0.025	99.3% (95% CI, 98.7–99.7%)	33.8%	0.128	110.1 (95% CI, 49.7–243.8)	26.3%	0.193	0.79 (95% CI, 0.42–1.00)
SUBGROUP ANALYSIS OF STAGE IIIB–IV AND RECURRENCE IN TWO SYSTEMS											
ddPCR^	8	73.4% (95% CI, 68.8–77.6%)	47.3%	0.048	95.7% (95% CI, 92.9–97.6%)	62.7%	0.004	34.4 (95% CI, 15.3–77.3)	27.5%	0.191	0.85 (95% CI, 0.75–0.95)
ARMS-PCR	6	62.9% (95% CI, 55.9–69.6%)	80.9%	0.001	95.6% (95% CI, 91.9–98.0%)	70.4%	0.005	24.5 (95% CI, 6.1–98.1)	65.0%	0.014	0.69 (95% CI, 0.16, 1.00)
SUBGROUP ANALYSIS OF STAGE IA–IV IN TWO SYSTEMS											
ddPCR^	3	64.7% (95% CI, 55.2–73.3%)	59.5%	0.060	94.4% (95% CI, 88.9–97.7%)	0%	0.538	29.8 (95% CI, 12.6–70.6)	0%	0.79	0.92 (95% CI, 0.77–1.00)
ARMS-PCR^	4	65.1% (95% CI, 62.2–68.0%)	0%	0.929	98.0% (95% CI, 97.2–98.7%)	13.9%	0.323	88.7 (95% CI, 60.9–129.2)	0%	0.419	0.49 (95% CI, 0–1.00)
SUBGROUP ANALYSIS OF TKI-NAIVE IN TWO SYSTEMS[¶]											
ddPCR^	6	72.7% (95% CI, 66.6–78.2%)	69.1%	0.006	96.6% (95% CI, 94.1–98.2%)	65.9%	0.012	62.7 (95% CI, 28.1–140.1)	14.3%	0.323	0.94 (95% CI, 0.87–1.00)
ARMS-PCR^	6	64.5% (95% CI, 59.1–69.6%)	56.8%	0.041	98.7% (95% CI, 97.7–99.4%)	83.1%	0.001	74.5 (95% CI, 14.9–373.8)	79.5%	0.001	0.59 (95% CI, 0.22–0.97)
SUBGROUP ANALYSIS OF TKI-TREATED IN TWO SYSTEMS											
ddPCR	8	75.6% (95% CI, 69.0–81.5)	45.2%	0.090	93.5% (95% CI, 88.4–96.8%)	48.1%	0.073	29.7 (95% CI, 14.3–61.4)	0%	0.462	0.87 (95% CI, 0.79–0.96)
ARMS-PCR	8	65.5% (95% CI, 62.8–68.1%)	75.9%	0.001	98.0% (95% CI, 97.2–98.6%)	24.3%	0.235	45.4 (95% CI, 18.9–108.9)	48.3%	0.060	0.86 (95% CI, 0.66–1.00)

[†]Specially, five articles made direct comparisons between ddPCR and ARMS-PCR; [‡]Two articles of ddPCR did not give the results of exon 19 deletion mutation or L858R, separately; five articles of ARMS-PCR did not give the results of exon 19 deletion mutation or L858R, separately; [§]One article of ddPCR only presented L858R data; ^Three article of ddPCR and six articles of ARMS-PCR did not meet the requirements of subgroup analysis, separately. [¶]One article of ddPCR and two articles of ARMS-PCR did not give the results of TKI used, separately. DOR, diagnostic odds ratio; AUROC, area under the summary receiver-operating characteristic curve; ddPCR, droplet digital PCR; ARMS-PCR, amplification refractory mutation system PCR; L858, exon 21 Leu858Arg; TKI, tyrosine kinase inhibitors.

TABLE 3 | The results of direct and indirect comparison of ddPCR and ARMS-PCR.

Results of comparison		Platforms	*P*-value
Overall results	Direct comparison	Sensitivity: ddPCR vs. ARMS-PCR = 69.3 vs. 69.0%	0.960
		Specificity: ddPCR vs. ARMS-PCR = 97.3 vs. 98.7%	0.473
	Indirect comparison	Sensitivity: ddPCR vs. ARMS-PCR = 72.1 vs. 65.3%	0.003
		Specificity: ddPCR vs. ARMS-PCR = 95.6 vs. 98.2%	0.007
Subgroup results	Direct comparison under the same platforms	Sensitivity: ddPCR (19del) vs. ddPCR (L858R) = 73.2 vs. 68.4%	0.282
		Specificity: ddPCR (19del) vs. ddPCR (L858R) = 99.3 vs. 98.8%	0.430
		Sensitivity: ddPCR (IIIB–IV) vs. ddPCR (IA–IV) = 72.5 vs. 71.2%	0.756
		Specificity: ddPCR (IIIB–IV) vs. ddPCR (IA–IV) = 93.5 vs. 98.8%	0.010
		Sensitivity: ddPCR (TKI-naive) vs. ddPCR (TKI-treated) = 71.7 vs. 74.1%	0.567
		Specificity: ddPCR (TKI-naive) vs. ddPCR (TKI-treated) = 96.5 vs. 95.4%	0.733
		Sensitivity: ARMS-PCR (19del) vs. ARMS-PCR (L858R) = 66.3 vs. 61.6%	0.271
		Specificity: ARMS-PCR (19del) vs. ARMS-PCR (L858R) = 99.3 vs. 99.3%	1.000
		Sensitivity: ARMS-PCR (IIIB–IV) vs. ARMS-PCR (IA–IV) = 73.7 vs. 64.2%	0.008
		Specificity: ARMS-PCR (IIIB–IV) vs. ARMS-PCR (IA–IV) = 96.3 vs. 98.3%	0.234
		Sensitivity: ARMS-PCR (TKI-naive) vs. ARMS-PCR (TKI-treated) = 63.9 vs. 65.4%	0.766
		Specificity: ARMS-PCR (TKI-naive) vs. ARMS-PCR (TKI-treated) = 96.7 vs. 98.5%	0.112
	Indirect comparison under the different platforms	Sensitivity: ddPCR (19del) vs. ARMS-PCR (19del) = 72.9 vs. 66.3%	0.087
		Specificity: ddPCR (19del) vs. ARMS-PCR (19del) = 99.1 vs. 99.3%	0.673
		Sensitivity: ddPCR (L858R) vs. ARMS-PCR (L858R) = 69.7 vs. 61.6%	0.080
		Specificity: ddPCR (L858R) vs. ARMS-PCR (L858R) = 98.2 vs. 99.3%	0.053
		Sensitivity: ddPCR (IIIB–IV) vs. ARMS-PCR (IIIB–IV) = 73.4 vs. 62.9%	0.012
		Specificity: ddPCR (IIIB–IV) vs. ARMS-PCR (IIIB–IV) = 95.7 vs. 95.6%	0.959
		Sensitivity: ddPCR (IA–IV) vs. ARMS-PCR (IA–IV) = 64.7 vs. 65.1%	0.934
		Specificity: ddPCR (IA–IV) vs. ARMS-PCR (IA–IV) = 94.4 vs. 98.0%	0.114
		Sensitivity: ddPCR (TKI-naive) vs. ARMS-PCR (TKI-naive) = 72.7 vs. 64.5%	0.040
		Specificity: ddPCR (TKI-naive) vs. ARMS-PCR (TKI-naive) = 96.6 vs. 98.7%	0.064
		Sensitivity: ddPCR (TKI-treated) vs. ARMS-PCR (TKI-treated) = 75.6 vs. 65.5%	0.035
		Specificity: ddPCR (TKI-treated) vs. ARMS-PCR (TKI-treated) = 93.5 vs. 98.0%	0.038

ddPCR, droplet digital PCR; ARMS-PCR, amplification refractory mutation system PCR; 19del, exon 19 deletion; L858, exon 21 Leu858Arg; TKI, tyrosine kinase inhibitors.

are sufficiently accurate to be considered a supplement or even alternative to tissue genotyping. Accordingly, this meta-analysis was conducted to assess the diagnostic accuracy of ddPCR system and ARMS-PCR system for detecting EGFR mutation in cfDNA.

In this meta-analysis, using tissue test as reference, we found that both ddPCR and ARMS-PCR had high diagnostic accuracy when testing in plasma cfDNA. By direct comparison, there was no significant difference between ddPCR and ARMS-PCR in overall accuracy. However, significant difference could be found in sensitivity and in specificity by indirect comparison. The direct comparison of results of the two platforms reported here suggested ddPCR had a higher sensitivity than ARMS-PCR in subgroup analysis of stage IA–IV. Combining the result of stratified analysis of tumor stage in sensitivity in ARMS-PCR, which demonstrated that ARMS-PCR had a higher sensitivity in the pure advanced lung cancer subgroup compared with early stage patients. We suspected that the sensitivity of ddPCR might be higher than ARMS-PCR in early stages, which warrants more data specific to early stage lung cancers. After indirect comparison, significant difference was also found in sensitivity between ddPCR and ARMS-PCR in the IIIB–IV subgroup, TKI-naïve subgroup and TKI-treated subgroup. Obvious higher specificity for ARMS-PCR was also observed in the TKI-treated subgroup. The indirect comparison of results of the two platforms suggested ddPCR had a higher sensitivity and ARMS-PCR had a higher specificity in some situations. However, the above results showed the discordance of sensitivity and specificity in two PCR platforms. The difference of results between direct and indirect comparison may be caused by insufficient sample sizes as only five articles had data for direct comparison. Furthermore, studies demonstrated the sensitivity of ARMS-PCR was 0.1% (38) and the sensitivity of ddPCR was 0.001% (39) detected in plasma, ddPCR showed improved limits of detection compared to ARMS-PCR, which may give rise to diverse results. Studies also indicated the abundance of ctDNA in patients with advanced stage varied from 0.1 to 53.2%, and was lower (<0.01%) in patients with early stage cancer (40, 41). Thus, it may sometimes show different diagnostic results in ddPCR and AMRS-PCR. In addition, by stratified analysis of EGFR-sensitizing mutations, we found that the exon 19 deletion testing sensitivity seemed higher than L858R in both testing systems, this is probably because tumor mutation burden (TMB) or ctDNA in plasma from the tumor of exon

19 deletion was higher than L858R, resulting in an increase of cfDNA in plasma.

Based on the excellent diagnostic performances of ddPCR and ARMS-PCR, we have reason to believe that it is rational to use these tissue-free methods as a supplement or an alternative option to tissue genotyping. Of note, both methods are relatively quick and inexpensive to detect the allelic frequency of mutations in cfDNA, but they cannot provide a comprehensive molecular profile of cancer. Besides, the sensitivity of the PCR systems could be limited if the proportion of tumor DNA in cfDNA is low. Owing to the high specificity, a patient with a negative result due to low percentage of mutant cfDNA could retest or the diagnosis could be determined in other ways. When a positive result is found, the patient may receive EGFR-TKI therapy, and should be followed up to evaluate the therapeutic effect. In addition, to validate the effectiveness and accuracy of liquid biopsy, prospective study based on the above test platforms for detecting EGFR mutation in cfDNA to compare lung cancer patients with healthy people as control is required and necessary, so as to set up an optimal cut-off point and reduce false positives. At the same time, we can also increase sample sizes to identify the diagnostic accuracy of ddPCR and ARMS-PCR.

To date, liquid biopsy is a complement to the tissue biopsy. If we want to use the result of genotyping of liquid biopsy directly in patients whose tissue samples are not available, we need to focus on the result of specificity first. When the specificity of liquid biopsy is increased to be consistent with tissue biopsy, it is reasonable to use liquid biopsy as an alternative to tissue biopsy in clinical practice for genotyping. For this reason, we should be cautious of false positives, though research has reported that cfDNA analysis does not involve formaldehyde fixation which can reduce false positive results due to deamination (42). The reasons for false positives can be divided into detection causes and non-detection causes. Detection causes mainly include: (i) Determination of cut-off values for EGFR mutations that were defined too low, (ii) Single tissue biopsy specimens were difficult to reflect the genetic characteristics of the whole tumor on intratumor heterogeneity, which meant even the result of cfDNA may be correct sometimes, false positive results by tissue biopsy conduced an opposite conclusion, (iii) Non-specific annealing of PCR primers could result in a false positive when the concentration of wild-type template was much higher than mutant template (20). Furthermore, the time interval between tissue samples acquired first and plasma samples acquired later may also cause false positives due to the tumor burdens becoming more severe as the disease progressed. Non-detection causes are mainly reflected in the following: (i) germline mutation, (ii) non-tumorous EGFR mutation, (iii) subclone EGFR. Germline mutation was caused by the change of family gene, contributing to the generation of family background in this population, which led to false positive results. The incorrect results of non-tumorous EGFR mutation were similar to germline mutation. If liquid biopsy technology can reduce false positive and increase specificity further, it would greatly benefit, not only tissue genotyping but also the longitudinal surveillance of clonal evolution (43).

We acknowledge several limitations to our study. First, it should be noted that, some studies only presented the results of mutations in exon 19 deletion and L858R, rather than total mutation level of EGFR. Therefore, we added the sample number of mutations of exon 19 deletion and L858R together to get an approximate result of EGFR mutation in each study. Second, most publications were retrospective studies, which may improve diagnostic accuracy artificially by setting a cut-off value, prospective clinical trials are needed for further investigation. Third, the diagnostic methods of ddPCR and ARMS-PCR, such as different extraction methods of DNA, and different types of primers and probes were not analyzed in this study. Last but not the least, in studies directly comparing the two PCR systems, the sample size was not large enough and the literature reports were limited. Through indirect comparison we were able to overcome sample selection bias to some extent however, the power of the test was not strong enough.

CONCLUSION

This study demonstrates that both ddPCR and ARMS-PCR have a high specificity with a practical sensitivity for detecting EGFR mutation in cfDNA of advanced lung cancer patients, which supports their application as a supplement or a conditional-alternative to tissue biopsy in clinical practice for genotyping. In addition, ddPCR-based plasma genotyping may be applied in clinical use more often with minimal false positives.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation, to any qualified researcher.

AUTHOR CONTRIBUTIONS

WL, JH, and CL conceived the study, and take responsibility for the integrity of the data and accuracy of the data analysis. QH and HL did the literature research, performed study selection, data extraction, and synthesis. BC, JL, and YZ participated in the analysis and interpretation of the data. SX, MG, and ZL wrote the draft review paper. WL, ZL, and CL revised the manuscript critically for important intellectual content and redrafted some of its section. All authors contributed to manuscript revision, read and approved the submitted version.

FUNDING

This work was supported by the following funding: The grant 2016YFC0905400 from the National Key R&D Program of China; China National Science Foundation (Grant Nos. 81871893, 81501996); Key Project of Guangzhou Scientific Research Project (Grant No. 201804020030); High-level university construction project of Guangzhou medical university

(Grant Nos. 20182737, 201721007, 201715907, 2017160107); National key R&D Program (Grant Nos. 2017YFC0907903, 2017YFC0112704).

ACKNOWLEDGMENTS

We thank Lindsey Hamblin for assistance with the language revision. The abstract of this study has been displayed as a poster in the European Lung Cancer Conference (ELCC), which was entitled as *Diagnostic Accuracy of Droplet Digital PCR (ddPCR) and Amplification Refractory Mutation System PCR (ARMS-PCR) for Detecting EGFR Mutation in Cell-Free DNA of Advanced Lung Cancer: A Meta-analysis.* Thanks for the exhibition opportunity given by the conference officials.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2020.00290/full#supplementary-material

REFERENCES

1. 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. doi: 10.3322/caac.21492
2. Jemal A, Center MM, DeSantis C, Ward EM. Global patterns of cancer incidence and mortality rates and trends. *Cancer Epidemiol Biomark Prev*. (2010) 19:1893–907. doi: 10.1158/1055-9965.EPI-10-0437
3. Jemal A, Siegel R, Xu J, Ward E. Cancer statistics, 2010. *CA Cancer J Clin*. (2010) 60:277–300. doi: 10.3322/caac.20073
4. Maemondo M, Inoue A, Kobayashi K, Sugawara S, Oizumi S, Isobe H, et al. Gefitinib or chemotherapy for non-small-cell lung cancer with mutated EGFR. *N Engl J Med*. (2010) 362:2380–8. doi: 10.1056/NEJMoa0909530
5. Rosell R, Carcereny E, Gervais R, Vergnenegre A, Massuti B, Felip E, et al. Erlotinib versus standard chemotherapy as first-line treatment for European patients with advanced EGFR mutation-positive non-small-cell lung cancer (EURTAC): a multicentre, open-label, randomised phase 3 trial. *Lancet Oncol*. (2012) 13:239–46. doi: 10.1016/S1470-2045(11)70393-X
6. Lindeman NI, Cagle PT, Aisner DL, Arcila ME, Beasley MB, Bernicker EH, et al. Updated molecular testing guideline for the selection of lung cancer patients for treatment with targeted tyrosine kinase inhibitors: guideline from the college of American pathologists, the international association for the study of lung cancer, and the association for molecular pathology. *J Thorac Oncol*. (2018) 13:323–58. doi: 10.1016/j.jtho.2017.12.001
7. Heerink WJ, de Bock GH, de Jonge GJ, Groen HJ, Vliegenthart R, Oudkerk M. Complication rates of CT-guided transthoracic lung biopsy: meta-analysis. *Eur Radiol*. (2017) 27:138–48. doi: 10.1007/s00330-016-4357-8
8. Siravegna G, Marsoni S, Siena S, Bardelli A. Integrating liquid biopsies into the management of cancer. *Nat Rev Clin Oncol*. (2017) 14:531–48. doi: 10.1038/nrclinonc.2017.14
9. Reckamp KL, Melnikova VO, Karlovich C, Sequist LV, Camidge DR, Wakelee H, et al. A highly sensitive and quantitative test platform for detection of NSCLC EGFR mutations in urine and plasma. *J Thorac Oncol*. (2016) 11:1690–700. doi: 10.1016/j.jtho.2016.05.035
10. Thress KS, Brant R, Carr TH, Dearden S, Jenkins S, Brown H, et al. EGFR mutation detection in ctDNA from NSCLC patient plasma: a cross-platform comparison of leading technologies to support the clinical development of AZD9291. *Lung Cancer*. (2015) 90:509–15. doi: 10.1016/j.lungcan.2015.10.004
11. Feng WN, Gu WQ, Zhao N, Pan YM, Luo W, Zhang H, et al. Comparison of the SuperARMS and droplet digital PCR for detecting EGFR mutation in ctDNA from NSCLC patients. *Transl Oncol*. (2018) 11:542–5. doi: 10.1016/j.tranon.2018.02.007
12. Gorgannezhad L, Umer M, Islam MN, Nguyen NT, Shiddiky MJA. Circulating tumor DNA and liquid biopsy: opportunities, challenges, and recent advances in detection technologies. *Lab Chip*. (2018) 18:1174–96. doi: 10.1039/C8LC00100F
13. Wan R, Wang Z, Lee JJ, Wang S, Li Q, Tang F, et al. Comprehensive analysis of the discordance of EGFR mutation status between tumor tissues and matched circulating tumor DNA in advanced non-small cell lung cancer. *J Thorac Oncol*. (2017) 12:1376–87. doi: 10.1016/j.jtho.2017.05.011
14. Douillard JY, Ostoros G, Cobo M, Ciuleanu T, Cole R, McWalter G, et al. Gefitinib treatment in EGFR mutated caucasian NSCLC: circulating-free tumor DNA as a surrogate for determination of EGFR status. *J Thorac Oncol*. (2014) 9:1345–53. doi: 10.1097/JTO.0000000000000263
15. Xu T, Kang X, You X, Dai L, Tian D, Yan W, et al. Cross-Platform comparison of four leading technologies for detecting EGFR mutations in circulating tumor DNA from non-small cell lung carcinoma patient plasma. *Theranostics*. (2017) 7:1437–46. doi: 10.7150/thno.16558
16. Ishii H, Azuma K, Sakai K, Kawahara A, Yamada K, Tokito T, et al. Digital PCR analysis of plasma cell-free DNA for non-invasive detection of drug resistance mechanisms in EGFR mutant NSCLC: correlation with paired tumor samples. *Oncotarget*. (2015) 6:30850–8. doi: 10.18632/oncotarget.5068
17. Zhu G, Ye X, Dong Z, Lu YC, Sun Y, Liu Y, et al. Highly sensitive droplet digital PCR method for detection of EGFR-Activating mutations in plasma cell-free DNA from patients with advanced non-small cell lung cancer. *J Mol Diagn*. (2015) 17:265–72. doi: 10.1016/j.jmoldx.2015.01.004
18. Lee JY, Qing X, Xiumin W, Yali B, Chi S, Bak SH, et al. Longitudinal monitoring of EGFR mutations in plasma predicts outcomes of NSCLC patients treated with EGFR TKIs: korean lung cancer consortium (KLCC-12-02). *Oncotarget*. (2016) 7:6984–93. doi: 10.18632/oncotarget.6874
19. Sacher AG, Paweletz C, Dahlberg SE, Alden RS, O'Connell A, Feeney N, et al. Prospective validation of rapid plasma genotyping for the detection of egfr and kras mutations in advanced lung cancer. *JAMA Oncol*. (2016) 2:1014–22. doi: 10.1001/jamaoncol.2016.0173
20. Wang L, Guo Q, Yu W, Qiao L, Zhao M, Zhang C, et al. Quantification of plasma EGFR mutations in patients with lung cancers: comparison of the performance of ARMS-Plusdroplet digital PCR. *Lung Cancer*. (2017) 114:31–7. doi: 10.1016/j.lungcan.2017.10.007
21. Yu Q, Huang F, Zhang M, Ji H, Wu S, Zhao Y, et al. Multiplex picoliter-droplet digital PCR for quantitative assessment of EGFR mutations in circulating cell-free DNA derived from advanced non-small cell lung cancer patients. *Mol Med Rep*. (2017) 16:1157–66. doi: 10.3892/mmr.2017.6712
22. Zhang X, Chang N, Yang G, Zhang Y, Ye M, Cao J, et al. A comparison of ARMS-Plus and droplet digital PCR for detecting EGFR activating mutations in plasma. *Oncotarget*. (2017) 8:112014–23. doi: 10.18632/oncotarget.22997
23. Zhang Y, Xu Y, Zhong W, Zhao J, Chen M, Zhang L, et al. Total DNA input is a crucial determinant of the sensitivity of plasma cell-free DNA EGFR mutation detection using droplet digital PCR. *Oncotarget*. (2017) 8:5861–73. doi: 10.18632/oncotarget.14390
24. Zhu YJ, Zhang HB, Liu YH, Zhu YZ, Chen J, Li Y, et al. Association of mutant EGFR L858Rexon 19 concentration in circulating cell-free DNA using droplet digital PCR with response to EGFR-TKIs in NSCLC. *Oncol Lett*. (2017) 14:2573–9. doi: 10.3892/ol.2017.6425
25. Jiang XW, Liu W, Zhu XY, Xu XX. Evaluation of EGFR mutations in NSCLC with highly sensitive droplet digital PCR assays. *Mol Med Rep*. (2019) 20:593–603. doi: 10.3892/mmr.2019.10259
26. Guo QM, Wang L, Yu WJ, Qiao LH, Zhao MN, Hu XM, et al. Detection of plasma EGFR mutations in NSCLC Patients with a validated ddPCR lung cfDNA assay. *J Cancer*. (2019) 10:4341–9. doi: 10.7150/jca.31326
27. Xu F, Wu J, Xue C, Zhao Y, Jiang W, Lin L, et al. Comparison of different methods for detecting epidermal growth factor receptor mutations in peripheral blood and tumor tissue of non-small cell lung cancer as a predictor of response to gefitinib. *Onco Targets Ther*. (2012) 5:439–47. doi: 10.2147/OTT.S37289

28. Liu X, Lu Y, Zhu G, Lei Y, Zheng L, Qin H, et al. The diagnostic accuracy of pleural effusion and plasma samples versus tumour tissue for detection of EGFR mutation in patients with advanced non-small cell lung cancer: comparison of methodologies. *J Clin Pathol*. (2013) 66:1065–9. doi: 10.1136/jclinpath-2013-201728
29. Li X, Ren R, Ren S, Chen X, Cai W, Zhou F, et al. Peripheral blood for epidermal growth factor receptor mutation detection in non-small cell lung cancer patients. *Transl Oncol*. (2014) 7:341–8. doi: 10.1016/j.tranon.2014.04.006
30. Duan H, Lu J, Lu T, Gao J, Zhang J, Xu Y, et al. Comparison of EGFR mutation status between plasma and tumor tissue in non-small cell lung cancer using the scorpion ARMS method and the possible prognostic significance of plasma EGFR mutation status. *Int J Clin Exp Pathol*. (2015) 8:13136–45.
31. Ma M, Shi C, Qian J, Teng J, Zhong H, Han B. Comparison of plasma and tissue samples in epidermal growth factor receptor mutation by ARMS in advanced non-small cell lung cancer. *Gene*. (2016) 591:58–64. doi: 10.1016/j.gene.2016.06.053
32. Li Y, Xu H, Su S, Ye J, Chen J, Jin X, et al. Clinical validation of a highly sensitive assay to detect EGFR mutations in plasma cell-free DNA from patients with advanced lung adenocarcinoma. *PLoS ONE*. (2017) 12:e0183331. doi: 10.1371/journal.pone.0183331
33. Zhou Y, Yang Y, Yang C, Chen Y, Yang C, Du Y, et al. Epidermal growth factor receptor (EGFR) mutations in non-small cell lung cancer (NSCLC) of Yunnan in southwestern China. *Oncotarget*. (2017) 8:15023–33. doi: 10.18632/oncotarget.14706
34. Cui S, Ye L, Wang H, Chu T, Zhao Y, Gu A, et al. Use of SuperARMS EGFR mutation detection kit to detect EGFR in plasma cell-free DNA of patients with lung adenocarcinoma. *Clin Lung Cancer*. (2018) 19:e313–22. doi: 10.1016/j.cllc.2017.12.009
35. Su F, Zheng K, Fu Y, Wu Q, Tang Y, Wang W, et al. [Influence of different therapies on EGFR mutants by circulating cell-free DNA of lung adenocarcinoma and prognosis]. *Zhongguo Fei Ai Za Zhi*. (2018) 21:389–96. doi: 10.3779/j.issn.1009-3419.2018.05.06
36. Bai H, Mao L, Wang HS, Zhao J, Yang L, An TT, et al. Epidermal growth factor receptor mutations in plasma DNA samples predict tumor response in Chinese patients with stages IIIB to IV non-small-cell lung cancer. *J Clin Oncol*. (2009) 27:2653–9. doi: 10.1200/JCO.2008.17.3930
37. Wang Z, Cheng Y, An T, Gao H, Wang K, Zhou Q, et al. Detection of EGFR mutations in plasma circulating tumour DNA as a selection criterion for first-line gefitinib treatment in patients with advanced lung adenocarcinoma (BENEFIT): a phase 2, single-arm, multicentre clinical trial. *Lancet Respir Med*. (2018) 6:681–90. doi: 10.1016/S2213-2600(18)30264-9
38. Spindler KL, Pallisgaard N, Vogelius I, Jakobsen A. Quantitative cell-free DNA, KRAS, and BRAF mutations in plasma from patients with metastatic colorectal cancer during treatment with cetuximab and irinotecan. *Clin Cancer Res*. (2012) 18:1177–85. doi: 10.1158/1078-0432.CCR-11-0564
39. Hindson BJ, Ness KD, Masquelier DA, Belgrader P, Heredia NJ, Makarewicz AJ, et al. High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. *Anal Chem*. (2011) 83:8604–10. doi: 10.1021/ac202028g
40. Xu S, Lou F, Wu Y, Sun DQ, Zhang JB, Chen W, et al. Circulating tumor DNA identified by targeted sequencing in advanced-stage non-small cell lung cancer patients. *Cancer Lett*. (2016) 370:324–31. doi: 10.1016/j.canlet.2015.11.005
41. Yong E. Cancer biomarkers: written in blood. *Nature*. (2014) 511:524–6. doi: 10.1038/511524a
42. Denis MG, Vallee A, Theoleyre S. EGFR T790M resistance mutation in non small-cell lung carcinoma. *Clin Chim Acta*. (2015) 444:81–5. doi: 10.1016/j.cca.2015.01.039
43. Siravegna G, Mussolin B, Buscarino M, Corti G, Cassingena A, Crisafulli G, et al. Clonal evolution and resistance to EGFR blockade in the blood of colorectal cancer patients. *Nat Med*. (2015) 21:795–801. doi: 10.1038/nm0715-827b

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

ORIGINAL RESEARCH
published: 11 March 2020
doi: 10.3389/fonc.2020.00236

Harmonization of Next-Generation Sequencing Procedure in Italian Laboratories: A Multi-Institutional Evaluation of the SiRe® Panel

*Umberto Malapelle[1], Francesco Pepe[1], Pasquale Pisapia[1], Roberta Sgariglia[1], Mariantonia Nacchio[1], Caterina De Luca[1], Rosanna Lacalamita[2], Stefania Tommasi[2], Rosamaria Pinto[2], Grazia Palomba[3], Giuseppe Palmieri[3], Davide Vacirca[4], Massimo Barberis[4], Irene Bottillo[5], Paola Grammatico[5], Lucia Rosalba Grillo[6], Valerio Costa[7], Riccardo Smeraglio[1], Dario Bruzzese[1] and Giancarlo Troncone[1]**

[1] Department of Public Health, University of Naples Federico II, Naples, Italy, [2] Molecular Diagnostics and Pharmacogenetics Unit, IRCCS Istituto Tumori "Giovanni Paolo II," Bari, Italy, [3] Unit of Cancer Genetics, Institute of Biomolecular Chemistry (ICB), National Research Council (CNR), Sassari, Italy, [4] Pathology Unit, European Institute of Oncology, Milan, Italy, [5] Laboratory of Medical Genetics, Department of Molecular Medicine, San Camillo-Forlanini Hospital, Sapienza University, Rome, Italy, [6] Department of Pathology, San Camillo-Forlanini Hospital, Sapienza University, Rome, Italy, [7] Institute of Genetics and Biophysics (CNR), Naples, Italy

OPEN ACCESS

Edited by:
Alfredo Addeo,
Geneva University Hospitals (HUG), Switzerland

Reviewed by:
Rafael Rosell,
Catalan Institute of Oncology, Spain
Nicola Normanno,
Istituto Nazionale Tumori IRCCS Fondazione G. Pascale, Italy

***Correspondence:**
Giancarlo Troncone
giancarlo.troncone@unina.it

Specialty section:
This article was submitted to Thoracic Oncology, a section of the journal Frontiers in Oncology

Received: 04 November 2019
Accepted: 12 February 2020
Published: 11 March 2020

Citation:
Malapelle U, Pepe F, Pisapia P, Sgariglia R, Nacchio M, De Luca C, Lacalamita R, Tommasi S, Pinto R, Palomba G, Palmieri G, Vacirca D, Barberis M, Bottillo I, Grammatico P, Grillo LR, Costa V, Smeraglio R, Bruzzese D and Troncone G (2020) Harmonization of Next-Generation Sequencing Procedure in Italian Laboratories: A Multi-Institutional Evaluation of the SiRe® Panel. Front. Oncol. 10:236. doi: 10.3389/fonc.2020.00236

Background: Next-generation sequencing (NGS) needs to be validated and standardized to ensure that cancer patients are reliably selected for target treatments. In Italy, NGS is performed in several institutions and harmonization of wet and dry procedures is needed. To this end, a consortium of five different laboratories, covering the most part of the Italian peninsula, was constituted. A narrow gene panel (SiRe®) covering 568 clinically relevant mutations in six different genes (*EGFR*, *KRAS*, *NRAS*, *BRAF*, *cKIT*, and *PDGFRα*) with a predictive role for therapy selection in non-small cell lung cancer (NSCLC), gastrointestinal stromal tumor, colorectal carcinoma (CRC), and melanoma was evaluated in each participating laboratory.

Methods: To assess the NGS inter-laboratory concordance, the SiRe® panel, with a related kit and protocol for library preparation, was used in each center to analyze a common set of 20 NSCLC and CRC routine samples. Concordance rate, in terms of mutation detected and relative allelic frequencies, was assessed. Then, each institution prospectively analyzed an additional set of 40 routine samples (for a total of 160 specimens) to assess the reproducibility of the NGS run parameters in each institution.

Results: An inter-laboratory agreement of 100% was reached in analyzing the data obtained from the 20 common sample sets; the concordance rate of allelic frequencies distribution was 0.989. The prospective analysis of the run metric parameters obtained by each center locally showed that the analytical performance of the SiRe® panel in the different institutions was highly reproducible.

Conclusions: The SiRe® panel represents a robust diagnostic tool to harmonize the NGS procedure in different Italian laboratories.

Keywords: colon cancer, lung cancer, predictive molecular pathology, next-generation sequencing, biomarkers

INTRODUCTION

In this era of precision oncology, predictive molecular pathology is key to assess actionable genetic targets in cancer patients (1–3). Thus, a large and steadily increasing number of predictive biomarkers need to be taken into account to personalize the therapeutic strategy in different solid tumors and in different patients (1–3). As an example, patients with metastatic colorectal cancer (mCRC), whose tumors are mutated in exons 2-3-4 of either Kirsten rat sarcoma (*KRAS*) or of neuroblastoma RAS viral oncogene homolog (*NRAS*) genes, are not eligible for target therapy with monoclonal antibodies against epidermal growth factor receptor (EGFR) protein (4–7). In addition, in mCRC, the National Comprehensive Cancer Network (NCCN) guidelines (Version 2.2018) recommend to genotype the patients for v-Raf murine sarcoma viral oncogene homolog B (*BRAF*) mutations, whose adverse prognostic role is well-established (8). Similarly, in non-small cell lung cancer (NSCLC) patients, the updated molecular testing guideline issued by the College of American Pathologists, the International Association for the Study of Lung Cancer, and the Association for Molecular Pathology defines *EGFR*, anaplastic lymphoma kinase (*ALK*), and ROS proto-oncogene 1 receptor tyrosine kinase as the "must test" genes to select patients for treatment with tyrosine kinase inhibitors (9–13). Even more recently, the American Society of Clinical Oncology established that also *BRAF* needs to be tested in all patients with advanced NSCLC as a positive predictive biomarker (14). In this rapidly evolving scenario, several are the technologies adopted to perform a molecular analysis. Among these, next-generation sequencing (NGS) represents a fascinating and versatile technology which is able to simultaneously analyze mutational hotspots in different gene targets for different cancer patients. However, only a laboratory with skilled and experienced personnel can reliably validate and implement NGS; moreover, the DNA quality and quantity derived from formalin-fixed and paraffin-embedded samples can be suboptimal and the number of samples classified as "inadequate" for molecular analysis is not negligible (15–19). As a result, when mutated alleles are only present in the subclonal neoplastic population, NGS might yield results which are not entirely consistent among the different institutions (20, 21).

To date, a large number of gene panels are commercially available for different clinical purposes; however, most of these panels are quite large and their use in routine practice is not cost-effective. Conversely, smaller gene panels seems to be more suitable than larger panels, especially when DNA input is less abundant and when the tested samples are represented by small tissue biopsies, as it often occurs in metastatic NSCLC patients (17, 18). To meet these challenges, in a previous study we have designed, developed, and validated, for both tissue samples and liquid biopsy specimens, a narrow NGS gene panel (SiRe®) that covers 568 clinically relevant mutations in six genes (*EGFR*, *KRAS*, *NRAS*, *BRAF*, *cKIT*, and *PDGFRα*) involved in NSCLC, gastrointestinal stromal tumor, mCRC, and melanoma (19–23).

In the current study, we aim to evaluate the performance of the SiRe® NGS panel in a multi-institutional study, thanks to a consortium constituted by five different Italian laboratories experienced in the application of NGS in a predictive molecular pathology setting.

MATERIALS AND METHODS

SiRe® Gene Panel

The SiRe® NGS panel was developed by the Department of Public Health of the University of Naples Federico II (Fed II) to assess 568 mutations in six different genes (*EGFR*, *KRAS*, *NRAS*, *BRAF*, *cKIT*, and *PDGFRα*) as previously described (**Figure 1**). This panel, together with all reagents and a dedicated protocol (**Supplementary File 1**) required to produce gene libraries, was distributed to four different institutions (**Figure 2**).

Study Design

This study was designed to evaluate the multi-institutional performance of the SiRe® gene panel-based NGS assay (**Supplementary File 1**) by both a retrospective and a prospective analysis. The retrospective analysis verified the concordance rate among the different institutions (inter-laboratory concordance, **Table 1**), whereas the prospective part evaluated the SiRe® gene panel-based NGS assay performance in the routine daily practice.

Briefly, the coordinating center (University of Naples Federico II) selected from its archives a set of 20 colon or lung cases; this set of cases was blinded and dispatched to the participating institutions that provided the coordinator center with the run metric parameters and with the complete list of all the mutations detected and their relative allelic fraction (AF) by 10 working days. Signal processing, base calling, and coverage analysis were carried out in a blinded way in each institution by using the SiRe® bed files on the Torrent Suite (Thermofisher). Variants were automatically annotated using a variant caller plug-in at specific optimized parameters of the SiRe® panel, as previously reported (19). The obtained BAM files were also visually inspected by an experienced user on the Golden Helix Genome Browser v.2.0.7 (Bozeman, MT, USA). Details relative to the DNA extraction procedures and to the NGS platform employed are reported in the **Supplementary Table 1**.

To assess the SiRe® gene panel-based NGS assay performance in the routine daily practice, each institution analyzed a distinct set of 40 colon or lung cancer tissue samples (**Supplementary Table 2**), reporting the analytical success rate, the median number of reads for the sample, the median read length, the median number of mapped reads, the percentage of reads on target, the average reads for amplicon, and the uniformity of coverage. Data were compared with the previously obtained "in-house" validation data set from the coordinator center to assess the analytical performance of the SiRe® kit in different clinical settings.

All of the analyzed cases were reviewed by experienced pathologists and featured at least 20% of neoplastic cells.

Written informed consent was obtained from all patients and documented in accordance with the general authorization to process personal data for scientific research purposes from "The Italian Data Protection Authority" (http://www.garanteprivacy.it/web/guest/home/docweb/-/docwebdisplay/export/2485392). All information regarding human material

FIGURE 1 | Upper **(A)** and frontal **(B)** parts of the SiRe® kit box and the related reagent tubes distributed in the internal **(C)** part. Permission to publish the figure was obtained by Genedin s.r.l. (a spin-off of the Department of Public Health, University of Naples Federico II).

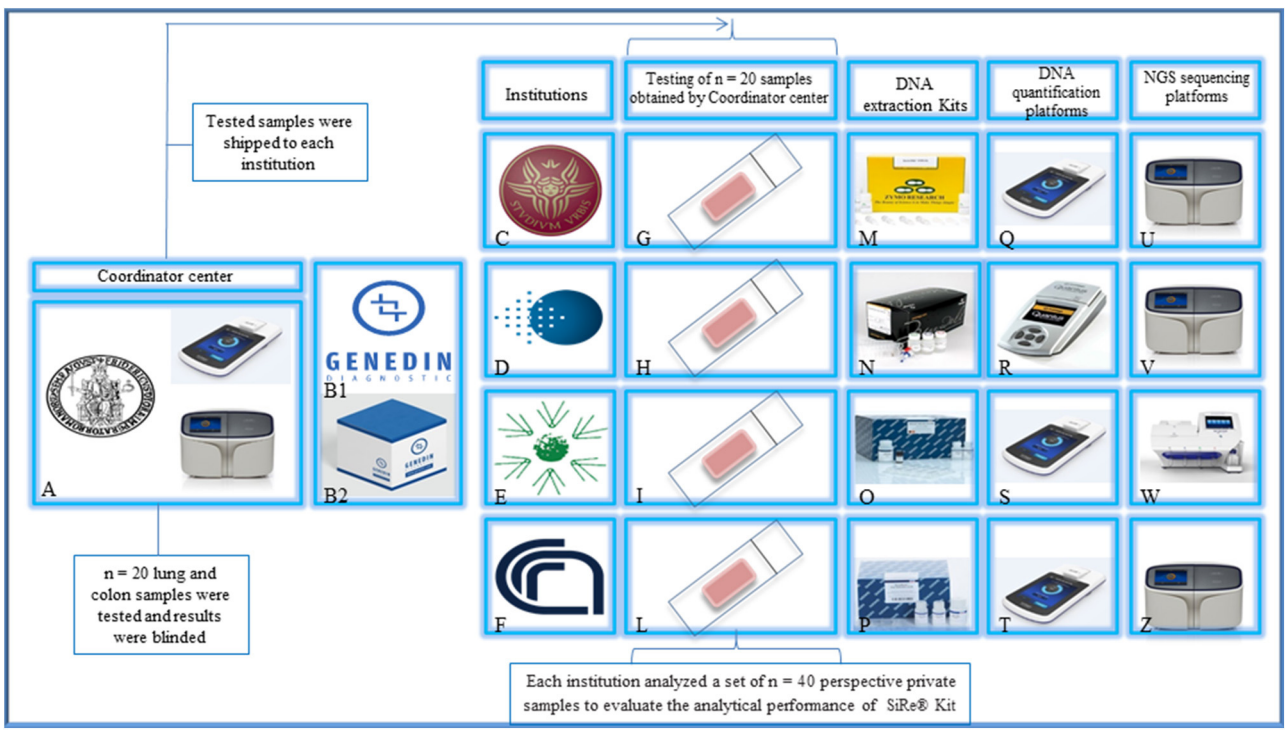


FIGURE 2 | Study design. The Department of Public Health of the University of Naples Federico II **(A)** and Genedin s.r.l. (a spin-off of the Department of Public Health, University of Naples Federico II) **(B1)** develop the SiRe® NGS panel kit **(B2)**. The kit was distributed to $n = 4$ different institutions [**(C)** University La Sapienza—Rome, **(D)** Istituto Oncologico Europeo—Milan, **(E)** Istituto Tumori Giovanni Paolo II—Bari, **(F)** Consiglio Nazionale delle Ricerche—Sassari] to analyze a series of $n = 160$ colon/lung routine samples **(G–L)** ($n = 40$ for each institution) following the local NGS workflow, for DNA extraction **(M–P)**, quantification **(Q–T)**, and sequencing on IonTorrent platforms **(U–Z)**, after an alignment phase on a common set of $n = 20$ shared samples. Permission to publish the figure was obtained by Genedin s.r.l. (a spin-off of the Department of Public Health, University of Naples Federico II).

TABLE 1 | Mutations detected on a common sample set ($n = 20$) by the University of Naples Federico II with relative allelic frequencies and relative results obtained among all the participating institutions.

	Fed II			Participating institutions		
Sample	**Gene**	**Mutation**	**Frequencies (%)**	**Gene**	**Mutation**	**Frequencies (%)**
1	*PIK3CA*	p.R88Q	15.9	*PIK3CA*	p.R88Q	15.9; 15.9; 15.6; 15.8
	PDGFRα	p.V824V	47.8	*PDGFR*	p.V824V	47.1; 51.2; 47.7; 42.6
2	*KRAS*	p.G12C	18.9	*KRAS*	p.G12C	18.9; 16.8; 22.8; 18.9
3	*KRAS*	p.G12V	42.0	*KRAS*	p.G12V	41.3; 39.9; 37.1; 42.3
4	*NRAS*	p.G12D	36.1	*NRAS*	p.G12D	38.9; 36.3; 39.8; 34.9
5	*KRAS*	p.A146T	31.0	*KRAS*	p.A146T	31.1; 31.1; 23.7; NA
	KIT	p.M541L	53.4	*KIT*	p.M541L	53.1; 53.1; 47.1; NA
6	*KRAS*	p.G13D	57.9	*KRAS*	p.G13D	57.4; 57.4; 58.0; 65.3
7	*KRAS*	p.G12C	36.4	*KRAS*	p.G12C	35.1; 35.1; 40.5; 32.4
	KIT	p.M541L	58.4	*KIT*	p.M541L	58.4; 58.4; 60.3; 58.3
8	*KRAS*	p.G12D	18.1	*KRAS*	p.G12D	18.4; 18.4; 18.7; 21.8
	PIK3CA	p.E545K	11.0	*PIK3CA*	p.E545K	11.5; 11.5; 9.9; 9.3
	KIT	p.M541L	100.0	*KIT*	p.M541L	99.8; 99.8; 100.0; 99.7
9	*KRAS*	p.G12V	22.0	*KRAS*	p.G12V	22.3; 22.3; 24.0; 22.0
10	*KRAS*	p.G12V	11.0	*KRAS*	p.G12V	11.0; 11.0; 7.5; 9.3
	PIK3CA	p.E542K	10.8	*PIK3CA*	p.E542K	10.7; 10.7; 11.8; 10.5
11	*KRAS*	p.G12D	14.9	*KRAS*	p.G12D	14.4; 19.3; 17.6; 15.7
12	–	WT	–	–	WT	–
13	*KRAS*	p.G12S	27.8	*KRAS*	p.G12S	27.1; 31.5; 27.0; 36.5
14	*PDGFRα*	p.V824V	99.1	*PDGFR*	p.V824V	99.1; 99.2; 99.1; 99.4
15	–	WT	–	–	WT	–
16	*KIT*	p.M541L	50.4	*KIT*	p.M541L	46.8; 51.2; 50.3; 54.0
17	*KRAS*	p.G13D	45.2	*KRAS*	p.G13D	44.6; 46.5; 45.3; 46.2
	PDGFR	p.V824V	18.9	*PDGFR*	p.V824V	19.7; 13.7; 18.4; 20.6
18	*KRAS*	p.G12A	6.1	*KRAS*	p.G12A	6.3; 5.8; 5.5; 6.2
	PDGFRα	p.V824V	45.8	*PDGFR*	p.V824V	43.9; 47.6; 46.5; 48.1
19	–	WT	–	–	WT	–
20	*EGFR*	p.E746_S752>V	43.8	*EGFR*	p.E746_S752>V	45.4; 51.1; 45.5; 40.9

BRAF, V-Raf murine sarcoma viral oncogene homolog B1; EGFR, epidermal growth factor receptor; KIT, KIT proto-oncogene, receptor tyrosine kinase; KRAS, V-Ki-Ras2 Kirsten rat sarcoma 2 viral oncogene homolog; NA, not assessed; NRAS, neuroblastoma RAS viral (V-Ras) oncogene homolog; PDGFRα, platelet-derived growth factor; PIK3CA, phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha; WT, wild type; Fed II, University of Naples Federico II. For the participating institutions, the frequencies are reported from left to right: University La Sapienza—Rome, Istituto Oncologico Europeo—Milan, Consiglio Nazionale delle Ricerche—Sassari, Tumori Giovanni Paolo II—Bari.

was managed using anonymous numerical codes, and all samples were handled in compliance with the Helsinki Declaration (https://www.wma.net/fr/news-post/en-matiere-de-transfert-des-taches-la-securite-des-patients-et-la-qualite-des-soins-devraient-etre-primordiales/). According to the aforementioned national guidelines, the double-blinded study did not require an Ethical Committee approval since it did not affect the clinical management of the involved patients' samples.

Data Analysis

The mutations and their relative allelic frequencies concordance rate were assessed by using an intra-class correlation coefficient (ICC), while concordance between each institution and Federico II was evaluated using the Linn's concordance correlation coefficient (CCC) and further explored using Bland-Altman plots. CCC is a reproducibility index which allows assessing both precision and accuracy by evaluating the degree to which individual pairs fall on the line of perfect concordance (i.e., the 45° line through the origin). A Bland–Altman plot shows the average of two measures on the x-axis and their difference on the y-axis; it allows the evaluation of both biases—that which occurs when the average of the differences between the two paired measurements is significantly different from 0 and the so-called proportional bias that refers to a significant trend between the difference and the magnitude of the measurements, i.e., when the difference in values increases or decreases in proportion to the average values.

RESULTS

Inter-laboratory Agreement

As reported in "Materials and Methods," each institution received from the coordinating center (University of Naples Federico II) a set of 20 cases, providing the coordinator center with the run

metric parameters and with the complete list of all the mutations detected and their relative AFs by 10 working days. Then, the samples were aliquoted and shipped to each institution. The mutations detected by the University of Naples Federico II with relative AFs were considered as the gold standard and reported in **Table 1**. The obtained results, using the SiRe® kit, among all the institutions were fully concordant, reaching an inter-laboratory agreement of 100.00% (**Table 1**).

Prospective Evaluation of the SiRe® Panel's Analytical Performance in a Routine Setting

Following the retrospective analysis aimed to assess the inter-laboratory concordance rate, each institution prospectively analyzed a distinct set of 40 cases, including colon or lung cancer tissue samples. All cases (160/160) were successfully analyzed and the success rate was 100.00% (**Supplementary Table 2**). The median number of reads for the sample was 306,332.38 (ranging from 191 to 976,243), the median number of read length was 128.20 bp (ranging from 60 to 168 bp), the median number of mapped reads was 302,203.01 (ranging from 191 to 908,628), the mean percentage of reads on target was 88.45% (ranging from 14.47 to 99.53%), the average reads for amplicon was 5,380.93 (ranging from seven to 20,385), and the uniformity of coverage was 93.73% (ranging from 71.95 to 100%), in line with the data previously obtained from our group in the "in-house" validation experiments of the SiRe® panel (23).

Regarding the mutant allelic frequencies distribution, a high level of agreement was reached; in particular, the ICC was 0.989 (95% C.I.: 0.981–0.994), and comparing the mutant allelic frequencies distribution with the gold standard, the Linn's concordance correlation coefficient was high for *KRAS* mutation (CCC: 0.977; 95.00% C.I.: 0.963–0.986), *PDGFRα* mutation (CCC: 0.989; 95% C.I.: 0.974–0.996), and *cKIT* mutations (CCC: 0.954; 95% C.I.: 0.888–0.982).

DISCUSSION

NGS represents a fascinating and versatile technology for the simultaneous analysis of different genes in different cancer patients. However, NGS requires many different laboratory steps, from DNA extraction, libraries preparation, sequencing procedure, and data interpretation, which can lead to results not being always fully consistent and reproducible in different laboratories, a limit inherent to many laboratory-developed tests. Thus, it is widely held that a reliable and cost-effective validation and implementation of this procedure in routine practice would benefit from a high degree of collaboration among skilled and experienced molecular biologists belonging to different institutions. In particular, networking is crucial to meet the challenges related to routine clinical sample processing as, in many cases, issues involve a suboptimal quantity and quality of nucleic acids (15–19). Furthermore, besides driving detection of mutations that are evenly distributed in neoplastic tissue, resistant genomic alteration features heterogeneity in the molecular landscape of many cancers; the detection of distinct mutations in different subclonal neoplastic populations can only be addressed by robust and reliable gene panels, ensuring a uniform coverage of the target regions. In particular, small gene panels, such as the SiRe® NGS panel, filling an intermediate space between allelic-specific PCR approaches and targeted re-sequencing have several advantages (17, 18). In fact, the SiRe® panel has previously been designed and validated for both tissue samples and liquid biopsy specimens to cover 568 clinically relevant mutations in six genes (*EGFR*, *KRAS*, *NRAS*, *BRAF*, *cKIT*, and *PDGFRα*) involved in NSCLC, gastrointestinal stromal tumor, mCRC, and melanoma, meeting the clinical indication for drug prescription from the European Medical Agency. In the validation study, a high analytical sensitivity (0.005%) with a 0.01% lower limit of detection was reported (19). We also developed a SiRe®-specific preparation protocol to enable the pooling of two 16-sample libraries in each run. Thus, using this well-standardized procedure, we were able to sequence simultaneously up to 32 paired plasma/serum samples in <3 h on the IonTorrent platforms, with a consequent reduction in the total consumable cost, limiting the expense to 98 euro for the simultaneous analysis of six different genes, which is comparable with the cost of the most commercially available real-time PCR-based kits (19–23).

In different European countries, the implementation of NGS in routine diagnostic procedures, beyond pre-analytical and technical factors, strongly depends on more general considerations relative to the healthcare systems in which predictive molecular pathology is practiced. While in North American countries, tumor samples are outsourced in large reference laboratories, thanks to the well-resourced, reimbursement-based systems ensuring the repayment of extensive tumor sequencing. In Italy, NGS is practiced in many different laboratories close to the patients' homes, each of them using different wet and dry procedures. Needless to say, harmonization of different laboratory practices and of mutational databases is strongly needed to improve and homogenize the assessment of genomic biomarkers. To this end, this current study was carried out to assess the feasibility of adopting the same NGS panel in the context of a multi-institution study, thanks to a consortium constituted by five different highly experienced Italian laboratories whose geographical location covers a large part of the Italian peninsula. Our data, generated from the retrospective analysis in each participating laboratory of a set of 20 samples, reached a high interlaboratory agreement level, not only relative to the mutation detection but also in relation to allelic frequencies estimation. Considering the prospective phase of this study, to assess the analytical performance of the SiRe® panel in different routine settings, promising results were obtained; in fact, an overall success rate of 100.00% was reported, with the median values of the run metric parameters confirming that the data obtained in the "in-house" validation data of the SiRe® panel can be successfully reproduced in four different institutions.

The main limitation of the SiRe® NGS panel relies in the limited number of analyzed genes. In particular, as discussed above, in the design and development of our panel, we only focused our attention on the clinically relevant mutations in six genes. In addition, this panel is able to identify point mutations and indel alterations. Thus, further improvements are required

to increase the clinical performance of the SiRe® NGS panel. In particular, it is necessary to expand the reference range of the SiRe® NGS gene panel, focusing our attention on other clinically relevant genes and on additional alterations such as copy number variations and gene fusions.

In conclusion, considering altogether the results obtained from the current multi-institution study, the SiRe® NGS panel represents a robust diagnostic tool for mutational analysis in a predictive molecular pathology routine setting, which is useful in harmonizing the NGS procedures in different Italian laboratories.

DATA AVAILABILITY STATEMENT

The datasets analyzed for this study can be found in the SRA repository (https://www.ncbi.nlm.nih.gov/sra) (PRJNA604200).

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

UM and GT conceived the study. FP, PP, RSg, MN, CD, RL, RP, GPalo, DV, IB, and RSm performed the experiments and contributed as molecular pathologists. UM, ST, GPalm, MB, PG, LG, and GT supervised the experiments. DB performed the statistical analysis. UM, FP, PP, and GT wrote the manuscript. VC contributed as biostatistic. All of the authors critically reviewed the paper and approved the final version of the manuscript.

FUNDING

This study was funded by the Department of Public Health, University of Naples Federico II.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2020.00236/full#supplementary-material

Supplementary Table 1 | Details relative to the DNA extraction procedures and to the NGS platform employed by participating institutions.

Supplementary Table 2 | Run metric parameters of all analyzed samples by participating institutions.

Supplementary File 1 | Smart protocol for manual library preparation on Ion Torrent platforms. Permission to publish the protocol was obtained by Genedin s.r.l. (a spin-off of the Department of Public Health, University of Naples Federico II).

REFERENCES

1. Salto-Tellez M. More than a decade of molecular diagnostic cytopathology leading diagnostic and therapeutic decision-making. *Arch Pathol Lab Med.* (2018) 142:443–5. doi: 10.5858/apra.2017-0258-ED
2. Wall DP, Tonellato PJ. The future of genomics in pathology. *F1000 Med Rep.* (2012) 4:14. doi: 10.3410/M4-14
3. Walk EE. The role of pathologists in the era of personalized medicine. *Arch Pathol Lab Med.* (2009) 133:605–10. doi: 10.1043/1543-2165-133.4.605
4. Douillard JY, Siena S, Cassidy J, Tabernero J, Burkes R, Barugel M, et al. Randomized, phase III trial of panitumumab with infusional fluorouracil, leucovorin, and oxaliplatin (FOLFOX4) versus FOLFOX4 alone as first-line treatment in patients with previously untreated metastatic colorectal cancer: the PRIME study. *J Clin Oncol.* (2010) 28:4697–705. doi: 10.1200/JCO.2009.27.4860
5. Bokemeyer C, Bondarenko I, Hartmann JT, de Braud F, Schuch G, Zubel A, et al. Efficacy according to biomarker status of cetuximab plus FOLFOX-4 as first-line treatment for metastatic colorectal cancer: the OPUS study. *Ann Oncol.* (2011) 22:1535–46. doi: 10.1093/annonc/mdq632
6. Van Cutsem E, Köhne CH, Hitre E, Zaluski J, Chang Chien CR, Makhson A, et al. Cetuximab and chemotherapy as initial treatment for metastatic colorectal cancer. *N Engl J Med.* (2009) 360:1408–17. doi: 10.1056/NEJMoa0805019
7. Sorich MJ, Wiese MD, Rowland A, Kichenadasse G, McKinnon RA, Karapetis CS. Extended RAS mutations and anti-EGFR monoclonal antibody survival benefit in metastatic colorectal cancer: a meta-analysis of randomized, controlled trials. *Ann Oncol.* (2015) 26:13–21. doi: 10.1093/annonc/mdu378
8. Benson AB III, Venook AP, Al-Hawary MM, Cederquist L, Chen YJ, Ciombor KK, et al. NCCN guidelines insights: colon cancer, version 2.2018. *J Natl Compr Canc Netw.* (2018) 16:359–69. doi: 10.6004/jnccn.2018.0021
9. Lindeman NI, Cagle PT, Aisner DL, Arcila ME, Beasley MB, Bernicker EH, et al. Updated molecular testing guideline for the selection of lung cancer patients for treatment with targeted tyrosine kinase inhibitors: guideline from the college of American pathologists, the international association for the study of lung cancer, and the association for molecular pathology. *Arch Pathol Lab Med.* (2018) 142:321–46. doi: 10.5858/arpa.2017-0388-CP
10. Mok TS, Wu YL, Thongprasert S, Yang CH, Chu DT, Saijo N, et al. Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. *N Engl J Med.* (2009) 361:947–57. doi: 10.1056/NEJMoa0810699
11. Rosell R, Carcereny E, Gervais R, Vergnenegre A, Massuti B, Felip E, et al. Erlotinib versus standard chemotherapy as first-line treatment for European patients with advanced EGFR mutation-positive non-small-cell lung cancer (EURTAC): a multicentre, open-label, randomised phase 3 trial. *Lancet Oncol.* (2012) 13:239–46. doi: 10.1016/S1470-2045(11)70393-X
12. Yang JC, Wu YL, Schuler M, Sebastian M, Popat S, Yamamoto N, et al. Afatinib versus cisplatin-based chemotherapy for EGFR mutation-positive lung adenocarcinoma (LUX-Lung 3 and LUX-Lung 6): analysis of overall survival data from two randomised, phase 3 trials. *Lancet Oncol.* (2015) 16:141–51. doi: 10.1016/S1470-2045(14)71173-8
13. Soria JC, Ohe Y, Vansteenkiste J, Reungwetwattana T, Chewaskulyong B, Lee KH, et al. Osimertinib in untreated EGFR-mutated advanced non-small-cell lung cancer. *N Engl J Med.* (2018) 378:113–25. doi: 10.1056/NEJMoa1713137
14. Kalemkerian GP, Narula N, Kennedy EB, Biermann WA, Donington J, Leighl NB, et al. Molecular testing guideline for the selection of patients with lung cancer for treatment with targeted tyrosine kinase inhibitors: American society of clinical oncology endorsement of the college of American pathologists/international association for the study of lung cancer/association for molecular pathology clinical practice guideline update. *J Clin Oncol.* (2018) 36:911–9. doi: 10.1200/JCO.2017.76.7293
15. Vigliar E, Malapelle U, de Luca C, Bellevicine C, Troncone G. Challenges and opportunities of next-generation sequencing: a cytopathologist's perspective. *Cytopathology.* (2015) 26:271–83. doi: 10.1111/cyt.12265
16. Malapelle U, Pisapia P, Sgariglia R, Vigliar E, Biglietto M, Carlomagno C, et al. Less frequently mutated genes in colorectal cancer: evidences from

next-generation sequencing of 653 routine cases. *J Clin Pathol.* (2016) 69:767–71. doi: 10.1136/jclinpath-2015-203403
17. Maxwell P, Hynes SO, Fuchs M, Craig S, McGready C, McLean F, et al. Practical guide for the comparison of two next-generation sequencing systems for solid tumour analysis in a universal healthcare system. *J Clin Pathol.* (2018) 72:225–31. doi: 10.1136/jclinpath-2017-204917
18. Karnes HE, Duncavage EJ, Bernadt CT. Targeted next-generation sequencing using fine-needle aspirates from adenocarcinomas of the lung. *Cancer Cytopathol.* (2014) 122:104–13. doi: 10.1002/cncy.21361
19. Malapelle U, Mayo de-Las-Casas C, Rocco D, Garzon M, Pisapia P, Jordana-Ariza N, et al. Development of a gene panel for next-generation sequencing of clinically relevant mutations in cell-free DNA from cancer patients. *Br J Cancer.* (2017) 116:802–10. doi: 10.1038/bjc.2017.8
20. Malapelle U, Mayo-de-Las-Casas C, Molina-Vila MA, Rosell R, Savic S, Bihl M, et al. Consistency and reproducibility of next-generation sequencing and other multigene mutational assays: a worldwide ring trial study on quantitative cytological molecular reference specimens. *Cancer Cytopathol.* (2017) 125:615–26. doi: 10.1002/cncy.21868
21. Pisapia P, Malapelle U, Roma G, Saddar S, Zheng Q, Pepe F, et al. Consistency and reproducibility of next-generation sequencing in cytopathology: a second worldwide ring trial study on improved cytological molecular reference specimens. *Cancer Cytopathol.* (2019) 127:285–96. doi: 10.1002/cncy.22134
22. Pisapia P, Pepe F, Smeraglio R, Russo M, Rocco D, Sgariglia R, et al. Cell free DNA analysis by SiRe® next generation sequencing panel in non small cell lung cancer patients: focus on basal setting. *J Thorac Dis.* (2017) 9(Suppl 13) S1383–90. doi: 10.21037/jtd.2017.06.97
23. Pepe F, De Luca C, Smeraglio R, Pisapia P, Sgariglia R, Nacchio M, et al. Performance analysis of SiRe next-generation sequencing panel in diagnostic setting: focus on NSCLC routine samples. *J Clin Pathol.* (2019) 72:38–45. doi: 10.1136/jclinpath-2018-205386

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

ORIGINAL RESEARCH
published: 15 April 2020
doi: 10.3389/fonc.2020.00424

BTK Has Potential to Be a Prognostic Factor for Lung Adenocarcinoma and an Indicator for Tumor Microenvironment Remodeling: A Study Based on TCGA Data Mining

*Ke-Wei Bi, Xu-Ge Wei, Xiao-Xue Qin and Bo Li**

Key Laboratory of Cell Biology, Department of Developmental Cell Biology, Ministry of Public Health and Key Laboratory of Medical Cell Biology, Ministry of Education, China Medical University, Shenyang, China

Tumor microenvironment (TME) plays a crucial role in the initiation and progression of lung adenocarcinoma (LUAD); however, there is still a challenge in understanding the dynamic modulation of the immune and stromal components in TME. In the presented study, we applied CIBERSORT and ESTIMATE computational methods to calculate the proportion of tumor-infiltrating immune cell (TIC) and the amount of immune and stromal components in 551 LUAD cases from The Cancer Genome Atlas (TCGA) database. The differentially expressed genes (DEGs) were analyzed by COX regression analysis and protein–protein interaction (PPI) network construction. Then, Bruton tyrosine kinase (BTK) was determined as a predictive factor by the intersection analysis of univariate COX and PPI. Further analysis revealed that BTK expression was negatively correlated with the clinical pathologic characteristics (clinical stage, distant metastasis) and positively correlated with the survival of LUAD patients. Gene Set Enrichment Analysis (GSEA) showed that the genes in the high-expression BTK group were mainly enriched in immune-related activities. In the low-expression BTK group, the genes were enriched in metabolic pathways. CIBERSORT analysis for the proportion of TICs revealed that B-cell memory and CD8+ T cells were positively correlated with BTK expression, suggesting that BTK might be responsible for the preservation of immune-dominant status for TME. Thus, the levels of BTK might be useful for outlining the prognosis of LUAD patients and especially be a clue that the status of TME transition from immune-dominant to metabolic activity, which offered an extra insight for therapeutics of LUAD.

Keywords: BTK, tumor microenvironment, ESTIMATE, CIBERSORT, tumor-infiltrating immune cells, lung adenocarcinoma

OPEN ACCESS

Edited by:
Umberto Malapelle, University of Naples Federico II, Italy

Reviewed by:
Francesco Passiglia, Paolo Giaccone University Hospital in Palermo, Italy
Hongguang Liu, Northeastern University, China
Xiaoni Kong, Shanghai Jiao Tong University, China

****Correspondence:***
Bo Li
bli28@cmu.edu.cn

Specialty section:
This article was submitted to Thoracic Oncology, a section of the journal Frontiers in Oncology

Received: *06 November 2019*
Accepted: *10 March 2020*
Published: *15 April 2020*

Citation:
Bi K-W, Wei X-G, Qin X-X and Li B (2020) BTK Has Potential to Be a Prognostic Factor for Lung Adenocarcinoma and an Indicator for Tumor Microenvironment Remodeling: A Study Based on TCGA Data Mining. Front. Oncol. 10:424. doi: 10.3389/fonc.2020.00424

INTRODUCTION

Lung cancer is the main cause of cancer-related death worldwide by reason of high recurrence rate, late detection, and poor prognosis. As a member of non-small cell lung cancer (NSCLC), LUAD accounts for ∼40% of all lung cancer cases. The current treatment for lung cancer, including surgical resection, chemotherapy, and radiation, had limitation on the improvement of patient's survival (1, 2).

Accordingly, it is urgently needed to explore the carcinogenesis and therapeutics of lung cancer.

Increasing evidence demonstrated the importance of the tumor microenvironment (TME) in the tumor development. Collaborative interactions between cancer cells and their supporting cells contributed to the malignant phenotypes of cancer, such as immortal proliferation, resisting apoptosis, and evading immune surveillance. Therefore, the TME significantly influences therapeutic response and clinical outcome in cancer patients (3, 4). Structural components of the TME are mainly resident stromal cells and recruited immune cells. While there was compelling evidence for the role of stromal cell contributing to tumor angiogenesis and extracellular matrix remodeling, but perhaps it is still not fully understood (5). Meanwhile, a few studies paid close attention to the impact of the immune cells in TME on tumor growth and progression. A growing body of studies showed that the tumor-infiltrating immune cell (TIC) in TME served as a promising indicator for the therapeutic effects (6). The tumor-infiltrating lymphocyte (TIL) was significantly correlated with the 5-year survival of NSCLC, and low lymphocyte abundance in cancer was identified as a poor prognostic indicator in early-stage NSCLC (7, 8). This relevance brought about the improvement of immune-based therapeutics, resulting in the application of immune checkpoint inhibitors for NSCLC patients (9, 10). A recent study elucidated the role of lung cancer lineage specifiers SOX2 and NKX2-1 in tumor cell fate and neutrophil recruitment, suggesting that the determination of tumor immune microenvironment might impact the nature of the tumor (11). A gene-expression profiling analysis showed that immune activation and immune escape in TME occur before lung cancer invasion (12). These results suggested that the adaptive immune response within TME might be of much importance at the earliest stage of lung cancer. Therefore, there is a challenge in performing precise genetic analysis that could appropriately indicate the dynamic modulation of the immune and stromal components in TME.

Transcriptome-sequencing patterns followed by functional genomics analysis have shed light on the roles of different types of cells during TME modulation. In the presented article, we applied ESTIMATE and CIBERSORT computational methods to calculate the TIC proportion and the ratio of immune and stromal components of LUAD samples from The Cancer Genome Atlas (TCGA) database and identified a predictive biomarker, Bruton tyrosine kinase (BTK). BTK was a non-receptor tyrosine kinase of the Tec family, locating in the downstream of signal transduction of B-cell antigen receptor (BCR). Upon phosphorylated, BTK triggered several signaling pathways, which resulted in the survival of leukemic cells in many B-cell malignancies (13, 14). Recently, some preclinical data had demonstrated that BTK was overexpressed in some solid tumors and their periphery cells in TME such as dendritic cells, macrophages, myeloid derived suppressor cells, and endothelial cells (15, 16), suggesting that BTK might play a role in TME. Here we embarked from differentially expressed genes (DEGs) generated by comparison between immune components and stromal components in LUAD samples and revealed that the BTK might be a potential indicator for the alteration of TME status in LUAD.

Abbreviations: TME, tumor microenvironment; LUAD, lung adenocarcinoma; PPI, protein–protein interaction; DEG, differentially expressed gene; BTK, Bruton tyrosine kinase; GSEA, Gene Set Enrichment Analysis; NSCLC, non–small cell lung cancer; TIC, tumor-infiltrating immune cell; TIL, tumor-infiltrating lymphocyte; ICI, immune checkpoint inhibitor.

RESULTS

Analysis Process of This Study

The analysis process of our study is shown in in **Figure 1**. To estimate the proportion of TICs and the amount of immune and stromal component in LUAD samples, transcriptome RNA-seq data of 551 cases were downloaded from TCGA database followed by calculating with CIBERSORT and ESTIMATE algorithms. DEGs shared by ImmuneScore and StromalScore were used to constructed protein–protein interaction (PPI) network and univariate COX regression analysis, and then intersection analysis was performed using the core nodes in PPI network and the top significant factors obtained from the analysis of univariate COX regression. BTK and CCR2 were obtained, and we focused on BTK for the subsequent series of analysis, including survival and clinicopathological characteristics correlation analysis, COX regression, Gene Set Enrichment Analysis (GSEA), and correlation with TICs.

Scores Were Correlated With the Survival of LUAD Patients

To establish the correlation of the estimated proportion of immune and stromal with the survival rate, Kaplan–Meier survival analysis was used for ImmuneScore, StromalScore, and ESTIMATEScore, respectively. The higher score estimated in ImmuneScore or StromalScore were represented for the larger amount of the immune or stromal components in TME. ESTIMATEScore was the sum of ImmuneScore and StromalScore denoting the comprehensive proportion of both components in TME. As shown in **Figure 2A**, the proportion of immune components had positive correlation with the overall survival rate. Despite StromalScore had no significant correlation with the overall survival rate (**Figure 2B**), ESTIMATEScore still showed positive correlation with the survival rate (**Figure 2C**). These results implied that the immune components in TME were more suitable for indicating the prognosis of LUAD patients.

Scores Were Associated With the Clinic–Pathological Staging of LUAD Patients

For determining the relationship between the proportion of immune and stromal components with the clinicopathological characteristics, we analyzed the corresponding clinical information of LUAD cases from TCGA database. As

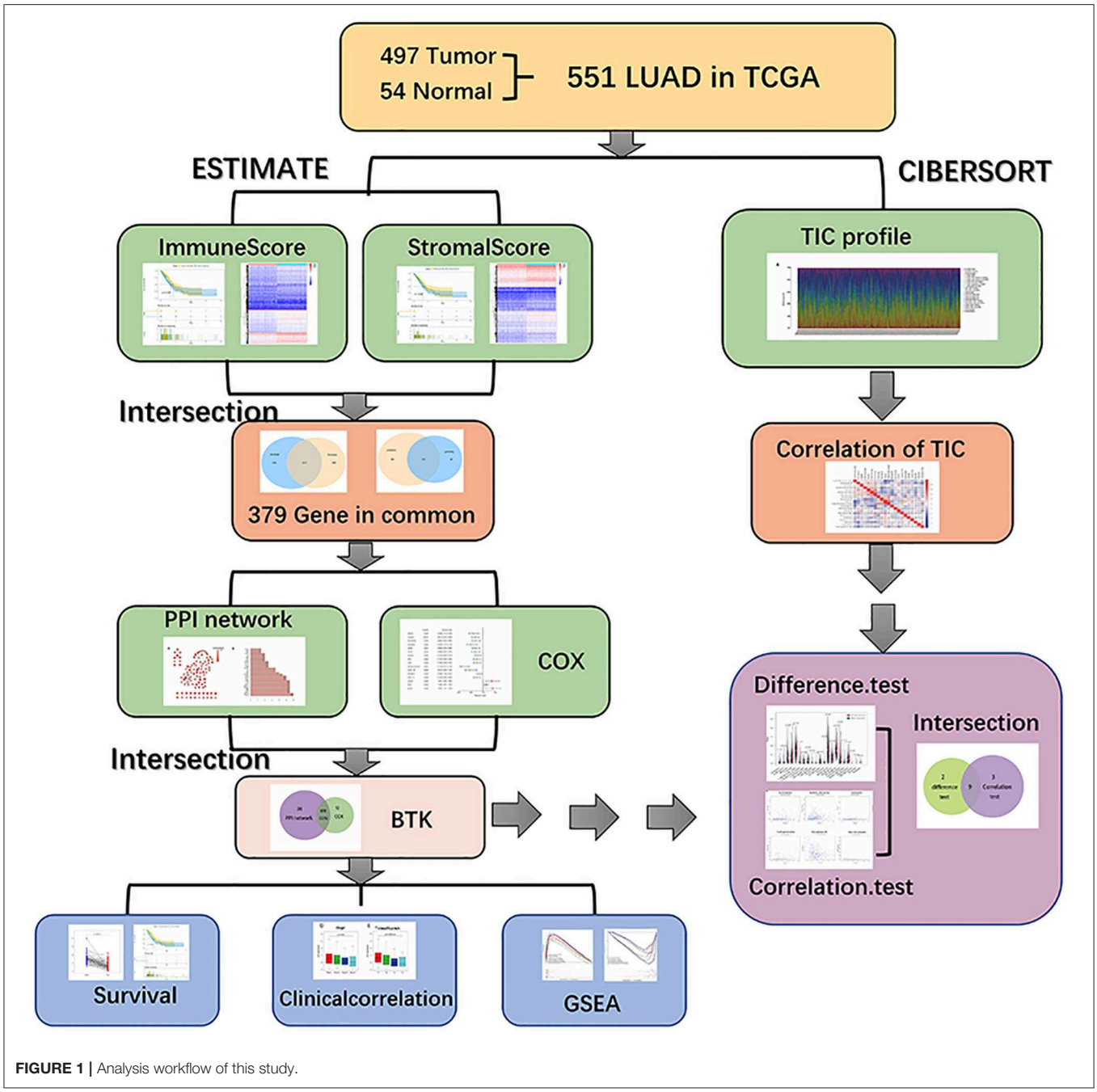


FIGURE 1 | Analysis workflow of this study.

shown in **Figure 3**, ImmuneScore showed the negative correlation with T classification of TMN stages (**Figure 3D**, $p = 0.003$); StromalScore was only negatively correlated to M classification of TMN stages (**Figure 3H**, $p = 0.007$), and ESTIMATEScore significantly declined accompany with the advance of TMN stages (**Figure 3F**, $p = 0.028$ and **Figure 3I**, $p = 0.021$). These results suggested that the ratio of immune and stromal components was associated with the progress of LUAD, such as invasion and metastasis.

DEGs Shared by ImmuneScore and StromalScore Were Predominantly Presented as the Enrichment of Immune-Related Genes

To ascertain the exact alterations of gene profile in TME regarding immune and stromal components, the comparison analysis between high- and low-score samples were carried out. Compared to the median, the total 776 DEGs were obtained from ImmuneScore (samples with high score vs. low score) Among

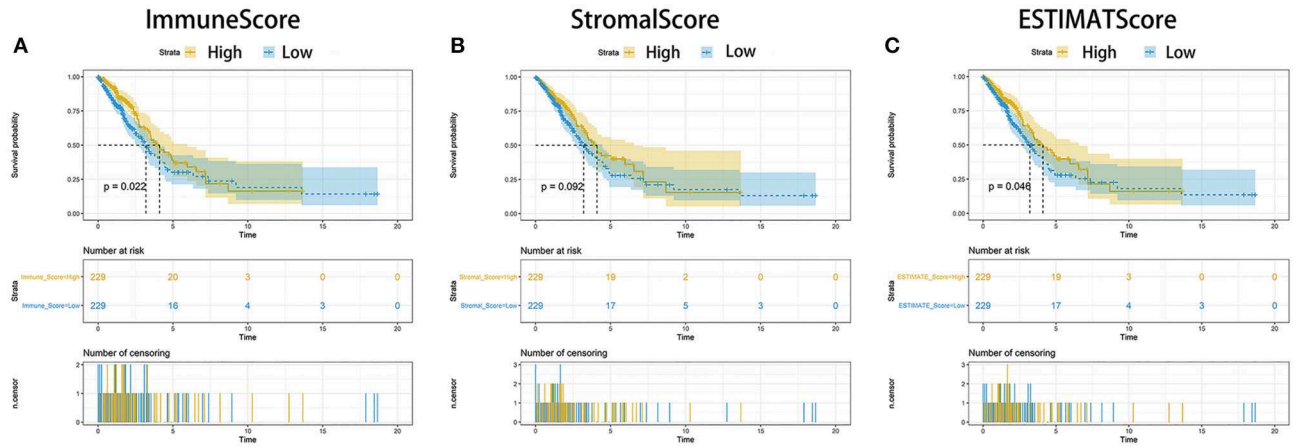


FIGURE 2 | Correlation of scores with the survival of patients with LUAD. **(A)** Kaplan–Meier survival analysis for LUAD patients grouped into high or low score in ImmuneScore determined by the comparison with the median. $p = 0.022$ by log-rank test. **(B)** Kaplan–Meier survival curve for StromalScore with $p = 0.092$ by log-rank test. **(C)** Survival analysis with Kaplan–Meier method for LUAD patients grouped by ESTIMATEScore ($p = 0.046$ by log-rank test).

them, 626 genes were up-regulated, and 150 genes were down-regulated (**Figures 4A,C,D**). Similarly, 783 DEGs were obtained from StromalScore, consisting of 665 up-regulated genes and 118 down-regulated genes (**Figures 4B–D**). The intersection analysis displayed by Venn plot showed a total of 317 up-regulated genes sharing by high score both in ImmuneScore and StromalScore and 62 down-regulated genes sharing by low score as well. These DEGs (total 379 genes) were possibly determinate factors for the status of TME. Results from gene ontology (GO) enrichment analysis indicated that the DEGs almost mapped to the immune-related GO terms, such as leukocyte proliferation and T-cell activation (**Figure 4E**). The Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis also displayed the enrichment of chemokine signaling pathway, cytokine–cytokine receptor interaction, and hematopoietic cell lineage (**Figure 4F**). Thus, the overall functions of DEGs seemed to map on immune-related activities, which implied that the involvement of immune factors was a predominant feature of TME in LUAD.

Intersection Analysis of PPI Network and Univariate COX Regression

To further explore the underlying mechanism, we constructed PPI network based on the STRING database using Cytoscape software [National Institute of General Medical Sciences (NIGMS) USA]. The interactions between 379 genes are shown in **Figure 5A**, and the bar plots were represented for the top 30 genes ranked by the number of nodes (**Figure 5B**). Univariate COX regression analysis for the survival of LUAD patients was performed to determine the significant factors among 379 DEGs (**Figure 5C**). And then, the intersection analysis between the leading nodes in PPI network and the top 16 factors ranked by the *p*-value of univariate COX regression was carried out, and only two factors, CCR2 and BTK, were overlapping from the above analyses (**Figure 5D**).

The Correlation of BTK Expression With the Survival and Classification of TNM Stages in LUAD Patients

BTK played a key role in the intracellular signaling of B lymphocytes. Ibrutinib, a BTK inhibitor, was effective for the treatment of patients with lymphocytic malignancies. Instead, there were no delightful results obtained from the treatment of patients with solid tumors, for example, NSCLC and breast cancer. In the presented study, all LUAD samples were grouped into BTK high-expression group and BTK low-expression group compared with the BTK median expression. The survival analysis showed that LUAD patients with BTK high expression had longer survival than that of BTK low expression (**Figure 6C**). After that, the analysis of BTK combined with clinical characteristics was performed (**Supplement Table 1**), and Wilcoxon rank sum test revealed that the expression of BTK in the tumor samples was significantly lower than that in the normal samples (**Figure 6A**). Similar results were observed in the pairing analysis between the normal and tumor tissues derived from the same patient (**Figure 6B**). The above results clearly indicated that the expression of BTK in TME was positive correlation with the prognosis of LUAD patients. In particular, the expressions of BTK were declined along with the progression of TNM stages (**Figures 6D–G**).

BTK Had Potential to Be an Indicator of TME Modulation

Given the levels of BTK were negatively correlated with the survival and TNM stages of LUAD patients, GSEA was implemented in the high-expression and the low-expression groups compared with the median level of BTK expression, respectively. As shown in **Figure 7A** and **Supplement Table 2**, the genes in BTK high-expression group were mainly enriched in immune-related activities, such as allograft rejection, complement, and interferon

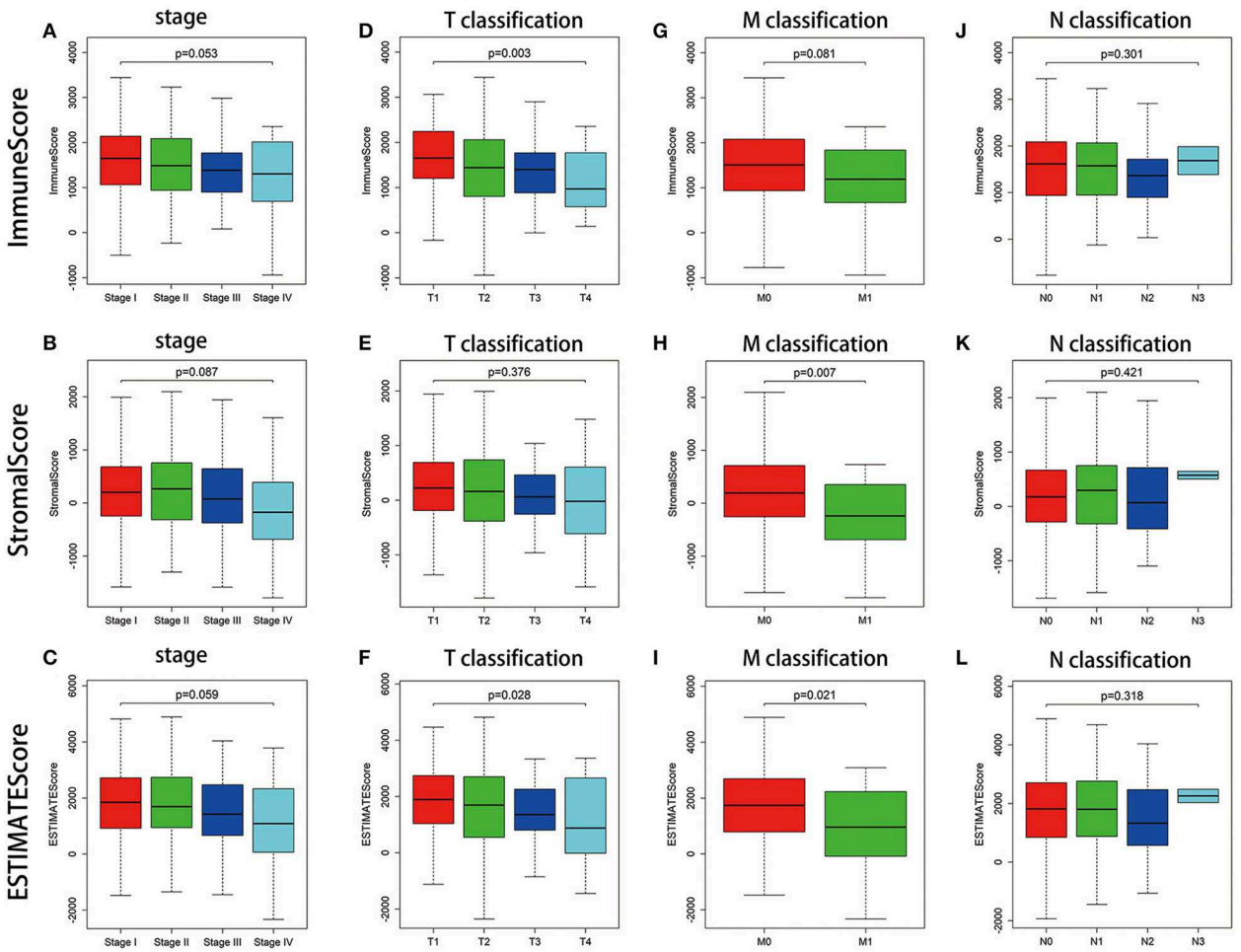


FIGURE 3 | Correlation of ImmuneScore and StromalScore with clinicopathological staging characteristics. **(A–C)** Distribution of ImmuneScore, StromalScore, and ESTIMATEScore in stage. The $p = 0.053$, 0.087, and 0.059, respectively, by Kruskal–Wallis rank sum test. **(D–F)** Distribution of three kinds of scores in T classification ($p = 0.003$, 0.376, 0.028 for ImmuneScore, StromalScore, and ESTIMATEScore, respectively, by Kruskal–Wallis rank sum test). **(G–I)** Distribution of scores in M classification ($p = 0.081$, 0.007, 0.021 for ImmuneScore, StromalScore, and ESTIMATEScore separately by Wilcoxon rank sum test). **(J–L)** Distribution of scores in N classification. Similar to the preceding, $p = 0.301$, 0.421, 0.318, respectively, with Kruskal–Wallis rank sum test.

response. As to BTK low-expression group, the genes were enriched in metabolic pathways, including glycolysis, oxidative phosphorylation, and typical tumor pathways (**Figure 7B** and **Supplement Table 2**). For C7 collection defined by MSigDB, the immunologic gene sets, multiple immune functional gene sets were enriched in the high BTK expression group (**Figure 7C** and **Supplement Table 2**). However, few gene sets were enriched in the low BTK expression group (**Figure 7D** and **Supplement Table 2**). These results suggested that BTK might be a potential indicator for the status of TME.

Correlation of BTK With the Proportion of TICs

To further confirm the correlation of BTK expression with the immune microenvironment, the proportion of tumor-infiltrating immune subsets was analyzed using CIBERSORT algorithm, and 21 kinds of immune cell profiles in LUAD samples were constructed (**Figure 8**). The results from the difference and correlation analyses showed that a total of eight kinds of TICs were correlated with the expression of BTK (**Figure 9**, **Supplement Figure 1**, and **Supplement Table 3**). Among them, five kinds of TICs were positively correlated with BTK expression, including B-cell memory, CD8+ T cells, monocytes, resting dendritic cells, and resting mast cells; three kinds of TICs were negatively correlated with BTK expression, including activated NK cells, macrophage M0, and activated mast cells. These results further supported that the levels of BTK affected the immune activity of TME.

DISCUSSION

In the presented study, we attempted to identify TME-related genes that contributed to the survival and the classification of TNM stages in LUAD patients from the TCGA database.

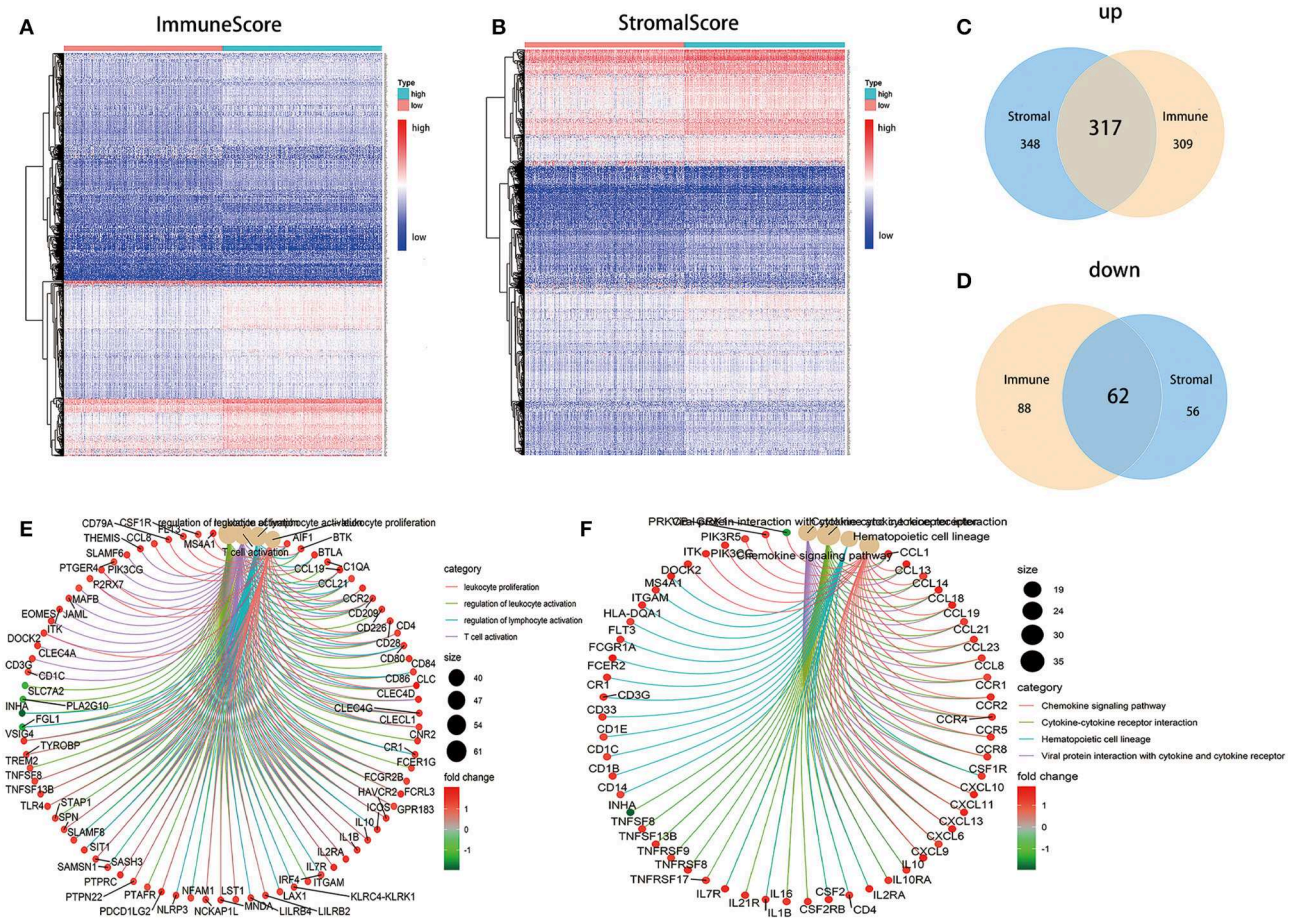


FIGURE 4 | Heatmaps, Venn plots, and enrichment analysis of GO and KEGG for DEGs. **(A)** Heatmap for DEGs generated by comparison of the high score group vs. the low score group in ImmuneScore. Row name of heatmap is the gene name, and column name is the ID of samples which not shown in plot. Differentially expressed genes were determined by Wilcoxon rank sum test with $q = 0.05$ and fold-change >1 after $\log_2$ transformation as the significance threshold. **(B)** Heatmap for DEGs in StromalScore, similar with **(A)**. **(C,D)** Venn plots showing common up-regulated and down-regulated DEGs shared by ImmuneScore and StromalScore, and $q < 0.05$ and fold-change >1 after $\log_2$ transformation as the DEGs significance filtering threshold. **(E,F)** GO and KEGG enrichment analysis for 379 DEGs, terms with p and $q < 0.05$ were believed to be enriched significantly.

BTK was identified to be involved in immune activities. Importantly, a series of bioinformatics analysis indicated that BTK might be an indicator for the status of TME in LUAD patients.

TME played a critical role in the initiation and progression of tumorigenesis. It is of great benefit to explore the potential therapeutic targets contributing to remodeling of TME and fostering transition of TME from tumor-friendly to tumor-suppressed. A large number of studies had shed light on the importance of immune microenvironment in tumorigenesis. Our results from the transcriptome analysis upon LUAD data in TCGA database implied that the immune components in TME contributed to the prognosis of patients. Particularly, the proportion of immune and stromal components in TME significantly correlated with the progression of LUAD, such as invasion and metastasis. These results highlighted the significance of exploring the interaction between tumor cells and immune cells, which provided novel insight for developing much more effective treatment regimen. Recently, a great advancement has been made in immunotherapy, and immune checkpoint inhibitors (ICIs) have been approved as a first-line drug for patients with advanced NSCLC (17). However, programmed death 1 ligand, the most important factor, seemed not like a good indicator to determine whether NSCLC patients were suitable for immunotherapy (18). As a heterogeneous tumor, NSCLC was consistently known as a non-immunogenic tumor. Nevertheless, recent studies discovered the existence of tumor antigen-specific cytotoxicity and clonal TIL expansion of particular sites in NSCLC, which was opposed to the previous understanding (19). Besides, the abundance of TILs was significantly correlated with the 5-year survival rate of NSCLC, and the low counts of preoperative lymphocyte have been known as a poor prognostic signal for early-stage NSCLC patients (7, 8). Despite promising efficacy in

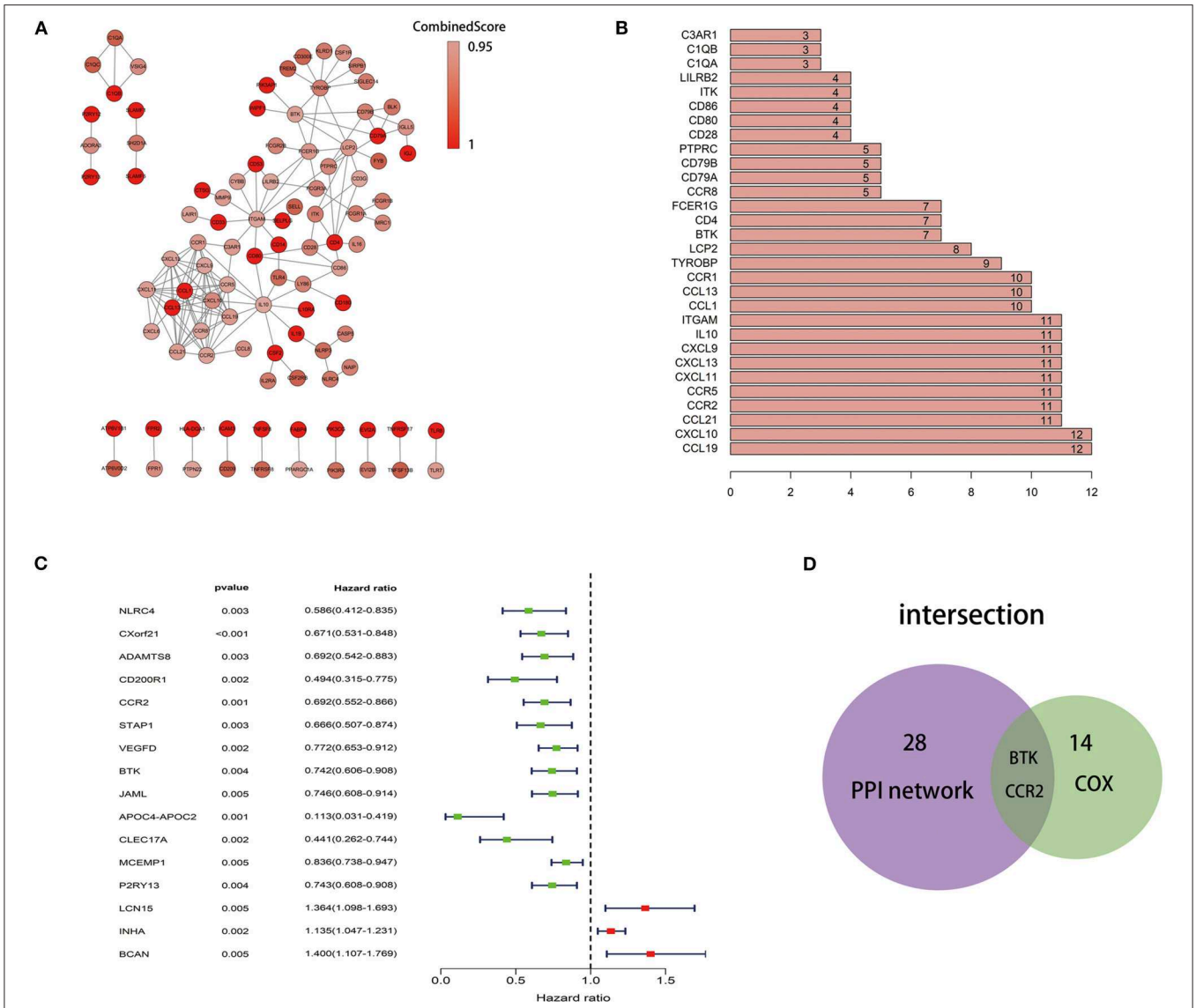


FIGURE 5 | Protein–protein interaction network and univariate COX. **(A)** Interaction network constructed with the nodes with interaction confidence value >0.95. **(B)** The top 30 genes ordered by the number of nodes. **(C)** Univariate COX regression analysis with 379 DEGs, listing the top significant factors with $p < 0.005$. **(D)** Venn plot showing the common factors shared by leading 30 nodes in PPI and top significant factors in univariate COX.

NSCLC treatment demonstrated by ICIs, a wide variety of immune-related adverse events could not easily be ignored (20). Therefore, the universality of ICIs as well as the susceptibility to immune-related adverse events was a tough problem, and it was necessary to investigate some novel candidates for the immunotherapy of NSCLC. Here, we embarked from the transcriptomic analysis of LUAD in TCGA database, which revealed that the decreased expression of BTK was significantly associated with the advanced clinicopathological characteristics (clinical stages and distant metastasis) and poor prognosis. Accordingly, it suggested that BTK might be a potential prognostic marker and a therapeutic target for TME in LUAD.

BTK is a non-receptor tyrosine kinase and a member of Tec kinase family. As a key component of the upstream in BCR signaling, BTK played a vital role in the proliferation and differentiation of B cells (13, 14). A small-molecule inhibitor of BTK, ibrutinib, had been used for the patients with hematological malignancies (21, 22). Recently, ibrutinib was expanded to treatment of some solid tumors, including pancreatic cancer, breast cancer, and NSCLC. However, it received no comforting effects as those of hematological malignancies (15, 23, 24). Our results suggested that the expression of BTK was decreased in the advancing stages of LUAD patients, which seemed to be inconsistent with hematological malignancies. Similarly, some studies had reported

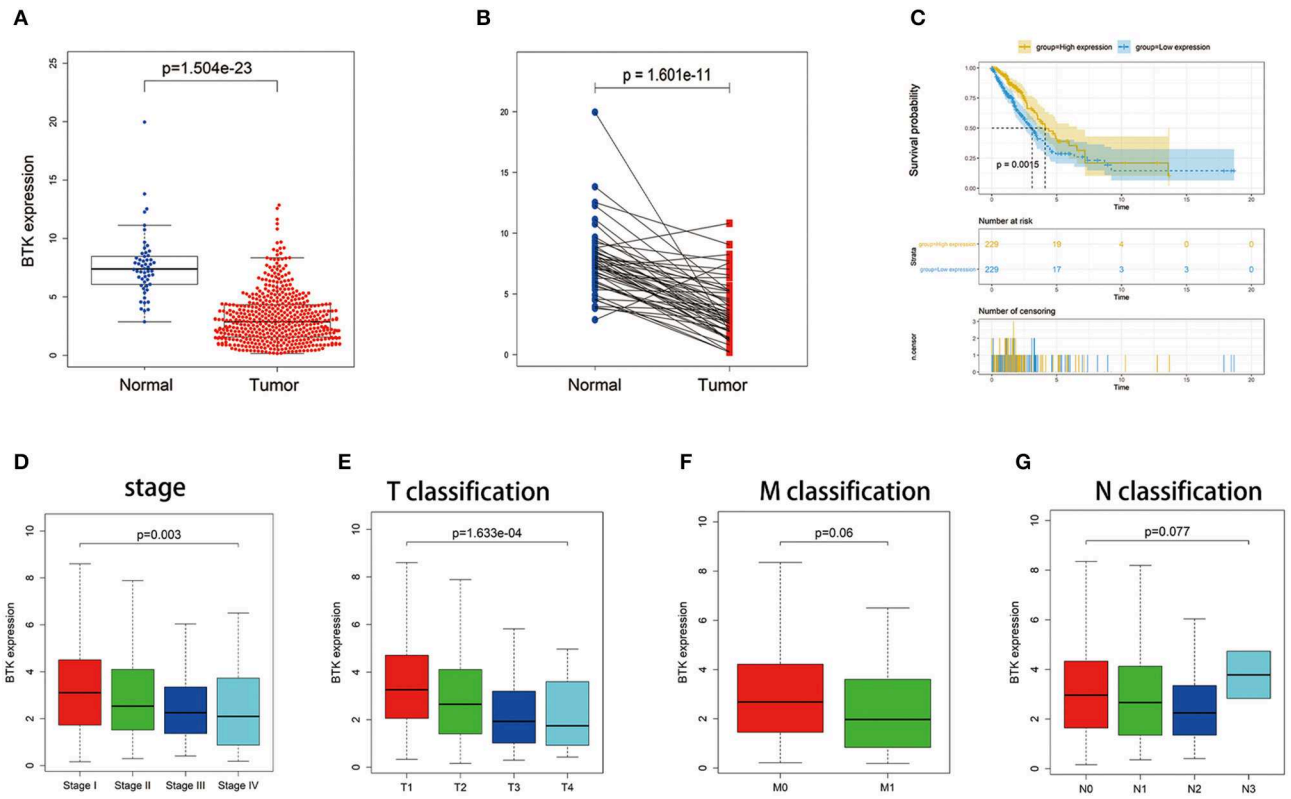


FIGURE 6 | The differentiated expression of BTK in samples and correlation with survival and clinicopathological staging characteristics of LUAD patients. **(A)** Differentiated expression of BTK in the normal and tumor sample. Analyses were performed across all normal and tumor samples with *p* value closing to zero by Wilcoxon rank sum test. **(B)** Paired differentiation analysis for expression of BTK in the normal and tumor sample deriving from the same one patient ($p = 1.601e{-}11$ by the Wilcoxon rank sum test). **(C)** Survival analysis for LUAD patients with different BTK expression. Patients were labeled with high expression or low expression depending on the comparison with the median expression level. $p = 0.0015$ by log-rank test. **(D–G)** The correlation of BTK expression with clinicopathological staging characteristics. Wilcoxon rank sum or Kruskal–Wallis rank sum test served as the statistical significance test.

that BTK might serve as a downstream effector in KRAS- and EGFR-activated signals in NSCLC, which might be explained by the existence of the distinct isoform of BTK in NSCLC (25, 26). Therefore, the discordance of response to BTK inhibitor implied that BTK seemed to play an antitumor role in LUAD. Further supports were derived from a series of studies, which indicated that BTK could modulate p53 activity to enhance tumor suppressor responses referring its antineoplastic properties (27–29). Thus, BTK might play a double-face role in tumor, either promoting survival or inducing apoptosis. Besides, it has been reported that BTK might be involved in regulating macrophage polarization in TME. Therefore, we further analyzed the relationship between BTK expression and TME. The GSEA results showed that immune-related signaling pathways, such as allograft rejection, complement, and interferon response, were significantly enriched in the BTK high-expression group. In the BTK low-expression group, metabolic pathways including glycolysis, oxidative phosphorylation, and typical tumor pathways were enriched. These results implied that BTK might participate in the status conversion of TME from immune-dominant to metabolic-dominant. Accumulated evidence had elucidated that BTK might be correlative to the metabolism. Ibrutinib promoted the uptake of glucose and glutamine and inhibited the synthesis of free fatty acid, which might be carried out via p53 signals (28, 30). To a certain degree, our data also showed that the balance between typical tumor pathways and vigorous glycolysis metabolism would affect the immunity status. The disorder of the balance could be reflected by the correlation of BTK low expression with metabolism. Further analysis of TIC supported this view. Accordingly, the downregulation of BTK along with the advancing stage of LUAD, the conversion of TME from immune-predominant to metabolic-dominant status, and the reduction of antitumor TICs supported that BTK might play an antitumor role in LUAD.

It was well-known that BTK was crucial for the functions of B lymphocytes. In the presented article, the CIBERSORT analysis for the proportion of TICs revealed that B-cell memory was positively correlated with BTK expression in LUAD patients. Regarding the role of B cells in cancer, there were some contradictory statements. A study indicated B cell as the promoter of carcinogenesis by inducing immunosuppression. However, the other study showed that CD40-activated B cell was a vigorous antigen-presenting cell and was able to induce

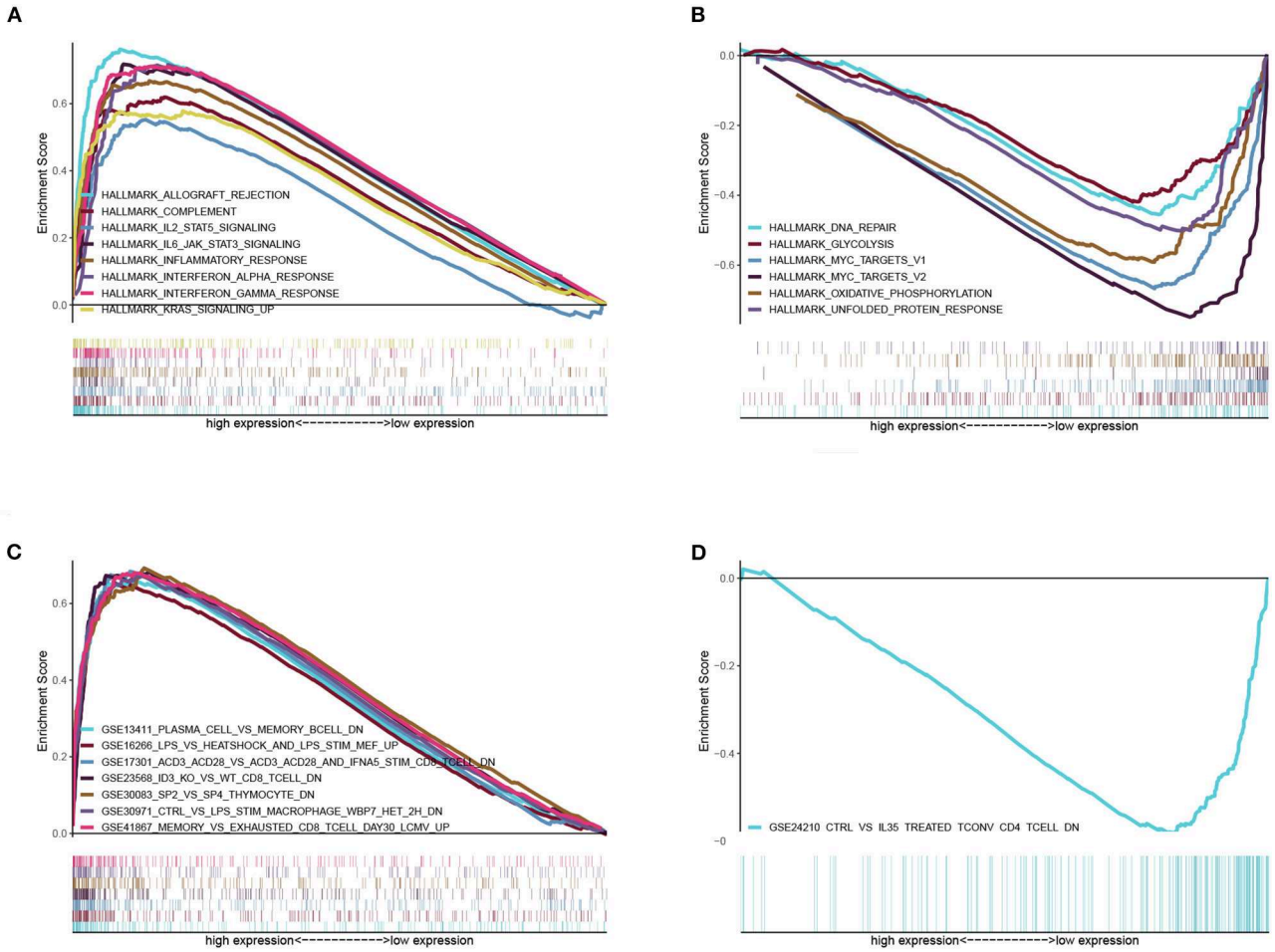


FIGURE 7 | GSEA for samples with high BTK expression and low expression. **(A)** The enriched gene sets in HALLMARK collection by the high BTK expression sample. Each line representing one particular gene set with unique color, and up-regulated genes located in the left approaching the origin of the coordinates, by contrast the down-regulated lay on the right of x-axis. Only gene sets with NOM $p < 0.05$ and FDR $q < 0.06$ were considered significant. And only several leading gene sets were displayed in the plot. **(B)** The enriched gene sets in HALLMARK by samples with low BTK expression. **(C)** Enriched gene sets in C7 collection, the immunologic gene sets, by samples of high BTK expression. Only several leading gene sets are shown in plot. **(D)** Enriched gene sets in C7 by the low BTK expression.

the effect of antitumor immunity (31). A recent study revealed a correlation between the amounts of tumor-infiltrating B cells and the survival of LUAD patients with specific mutation driver, which suggested that tumor-infiltrating B cells might be a symbol for the specific mutation in lung cancer cells (32). Therefore, the positive correlation between the amounts of B-cell memory and BTK expression in LUAD patients suggested that BTK might be responsible for the preservation of immune-active status in TME.

Using ESTIMATE algorithm, we determined the TME-related genes in LUAD through the functional enrichment analysis of LUAD samples in TCGA database. BTK was a potential prognostic factor for LUAD patients. More interestingly, BTK might be an indicator for the conversion of TME status from immune-dominant to metabolic-dominant. Therefore, further investigation should be conducted to clarify the accuracy of a combined analysis of BTK expression, the amounts of tumor-infltrating B-cell isoforms, and the types of mutation-driven prior to BTK inhibitor treatment for LUAD patients.

MATERIALS AND METHODS

Raw Data

Transcriptome RNA-seq data of 551 LUAD cases (normal samples, 54 cases; tumor samples, 497 cases) and the corresponding clinical data were downloaded from TCGA database (https://portal.gdc.cancer.gov/) with level 3.

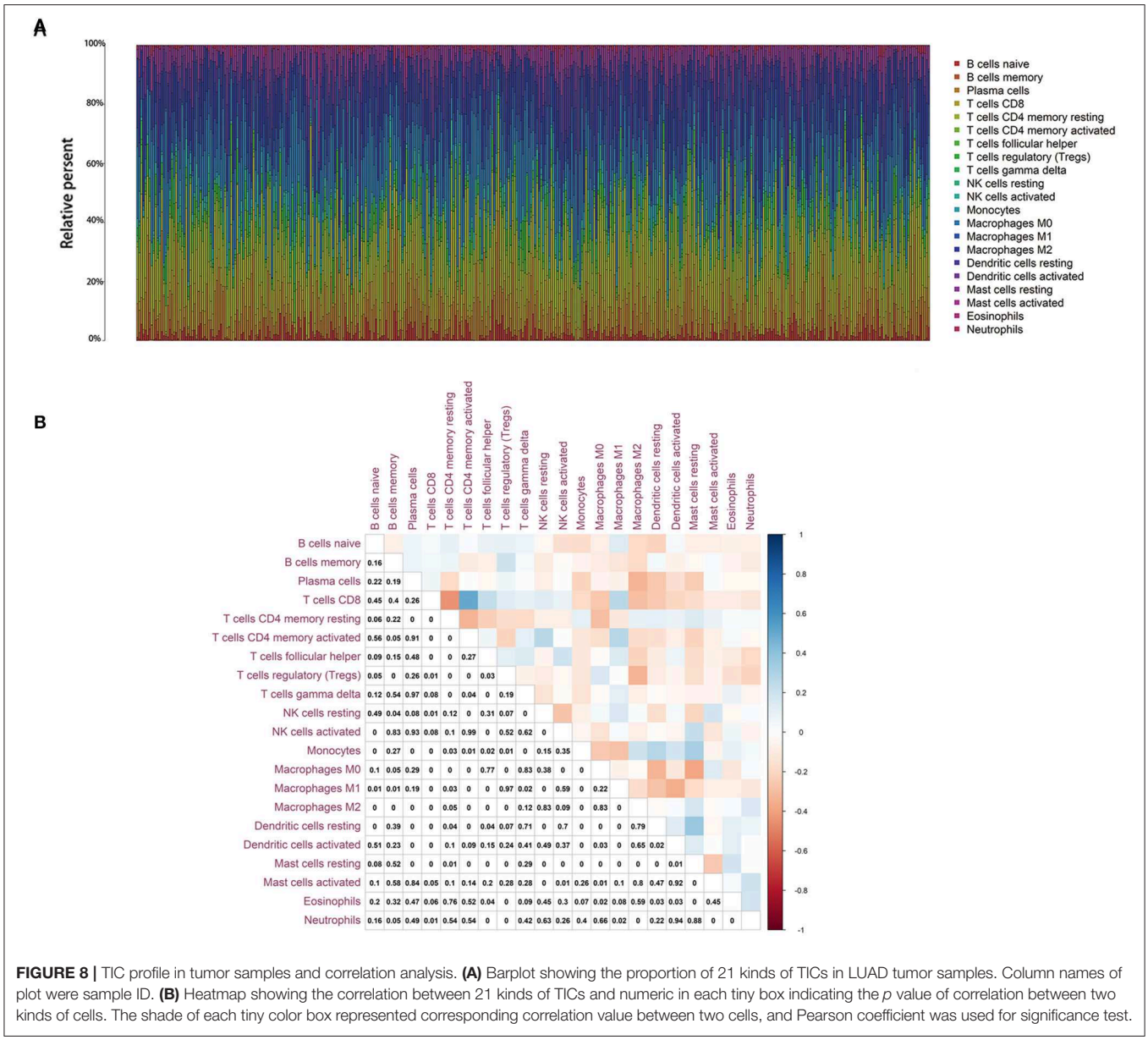


FIGURE 8 | TIC profile in tumor samples and correlation analysis. **(A)** Barplot showing the proportion of 21 kinds of TICs in LUAD tumor samples. Column names of plot were sample ID. **(B)** Heatmap showing the correlation between 21 kinds of TICs and numeric in each tiny box indicating the *p* value of correlation between two kinds of cells. The shade of each tiny color box represented corresponding correlation value between two cells, and Pearson coefficient was used for significance test.

Generation of ImmuneScore, StromalScore, and ESTIMATEScore

ESTIMATE algorithm by feat of R language version 3.5.1 loaded with estimate package (33) was used to estimate the ratio of immune-stromal component in TME for each sample, exhibited in the form of three kinds of scores: ImmuneScore, StromalScore, and ESTIMATEScore, which positively correlated with the ratio of immune, stromal, and the sum of both, respectively, which means the higher the respective score, the larger the ratio of the corresponding component in TME.

Survival Analysis

R language loaded with package survival and survminer was applied for the survival analysis. 458 tumor samples out of 497 cases had a detailed survival time record, with time span from 0 to 18.7 years, which were used for survival analysis. Kaplan–Meier method was used to plot the survival curve, and log rank as the statistical significance test; $p < 0.05$ was considered significant.

Generation of DEGs Between High-Score and Low-Score Groups Regarding ImmuneScore and StromalScore

497 tumor samples were labeled with high score or low score depending on the comparison to the median score in regarding ImmuneScore and StromalScore, respectively. Package limma was used to perform differentiation analysis of the gene expression, and DEGs were generated by the comparison between the high-score samples vs. the low-score samples. DEGs

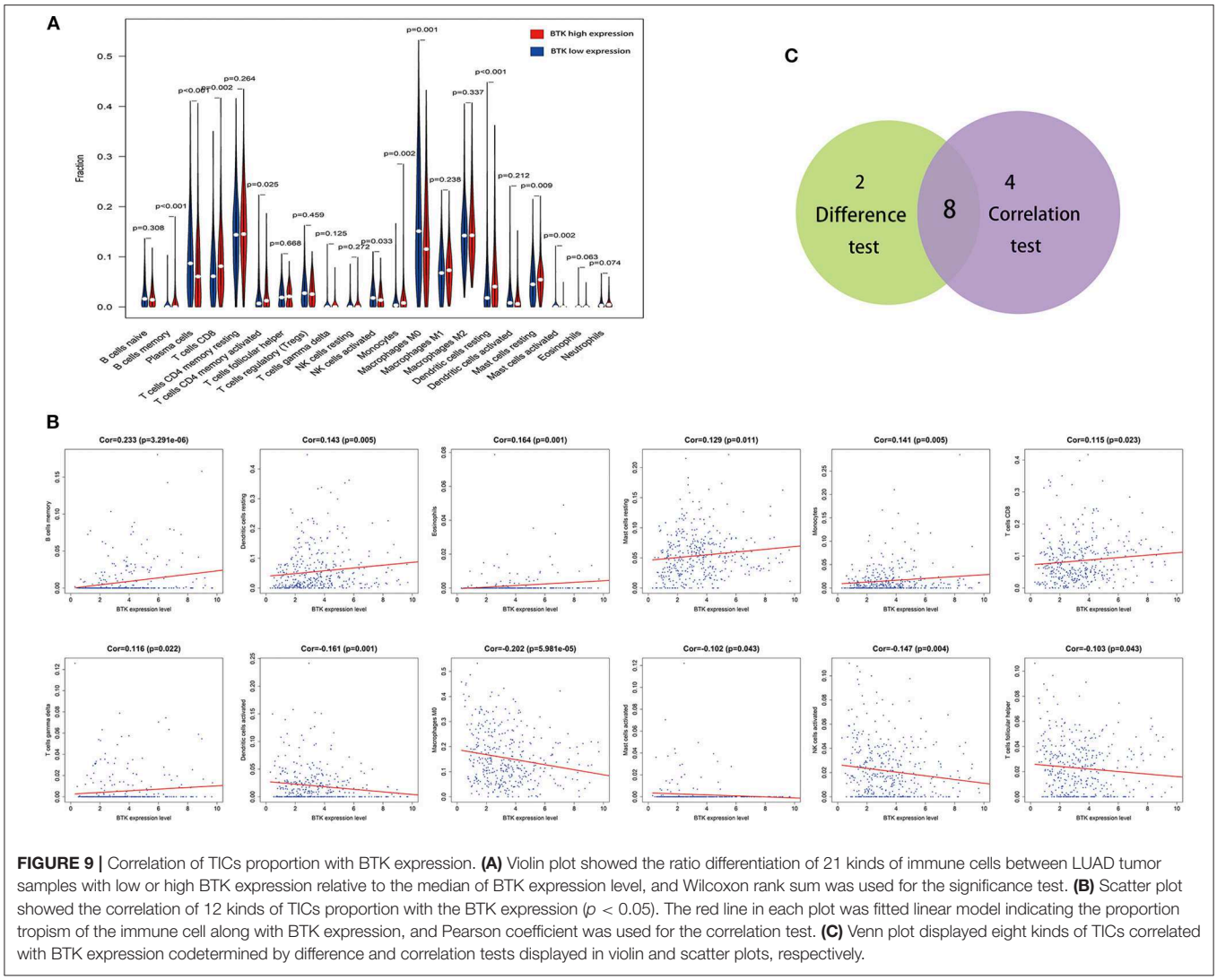


FIGURE 9 | Correlation of TICs proportion with BTK expression. **(A)** Violin plot showed the ratio differentiation of 21 kinds of immune cells between LUAD tumor samples with low or high BTK expression relative to the median of BTK expression level, and Wilcoxon rank sum was used for the significance test. **(B)** Scatter plot showed the correlation of 12 kinds of TICs proportion with the BTK expression ($p < 0.05$). The red line in each plot was fitted linear model indicating the proportion tropism of the immune cell along with BTK expression, and Pearson coefficient was used for the correlation test. **(C)** Venn plot displayed eight kinds of TICs correlated with BTK expression codetermined by difference and correlation tests displayed in violin and scatter plots, respectively.

with fold change larger than 1 after transformation of $\log_2$ (high-score group/low-score group) and false discovery rate (FDR) <0.05 were considered significant.

GO and KEGG Enrichment Analysis

GO and KEGG enrichment analyses using 379 DEGs were performed with R language with the aid of packages clusterProfiler, enrichplot, and ggplot2. Only terms with both *p*- and *q*-value of <0.05 were considered significantly enriched.

Heatmaps

Heatmaps of DEGs were produced by R language with package pheatmap.

Difference Analysis of Scores With Clinical Stages

The clinicopathological characteristics data corresponding to the LUAD samples were downloaded from TCGA. The analysis was performed by R language, and Wilcoxon rank sum or Kruskal–Wallis rank sum test as the significance test depending on the number of clinical stages for comparison.

PPI Network Construction

PPI network was constructed by STRING database, followed by reconstruction with Cytoscape of version 3.6.1. Nodes with confidence of interactive relationship larger than 0.95 were used for building network.

COX Regression Analysis

R language loaded with package survival was used for univariate COX regression. The top 16 genes ordered by *p* value from small to large in univariate COX were shown in the plot.

Gene Set Enrichment Analysis

Hallmark and C7 gene sets v6.2 collections were downloaded from Molecular Signatures Database as the target sets with which GSEA performed using the software gsea-3.0 downloaded from

Broad Institute. The whole transcriptome of all tumor samples was used for GSEA, and only gene sets with NOM $p < 0.05$ and FDR $q < 0.06$ were considered as significant.

TICs Profile

CIBERSORT computational method was applied for estimating the TIC abundance profile in all tumor samples, which followed by quality filtering that only 421 tumor samples with $p < 0.05$ were selected for the following analysis.

DATA AVAILABILITY STATEMENT

The datasets generated and analyzed during this study are available in the TCGA database (https://portal.gdc.cancer.gov).

AUTHOR CONTRIBUTIONS

K-WB, X-GW, and BL came up with the design and conception. The data analysis and visualization were conducted by K-WB, X-GW and X-XQ. The original writing of the draft and its editing were by K-WB and BL.

FUNDING

This work was supported by the National Natural Science Foundation of China (81672881).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2020.00424/full#supplementary-material

Supplement Figure 1 | TICs with non-significant correlation with the expression of BTK.

Supplement Table 1 | Clinicopathological characteristics statistics of LUAD patients from TCGA.

Supplement Table 2 | Enriched gene sets.

Supplement Table 3 | TICs co-determined by difference test and correlation test.

REFERENCES

1. Lemjabbar-Alaoui H, Hassan OU, Yang YW, Buchanan P. Lung cancer: biology and treatment options. *Biochim Biophys Acta.* (2015) 1856:189–210. doi: 10.1016/j.bbcan.2015.08.002
2. Hirsch FR, Scagliotti GV, Mulshine JL, Kwon R, Curran WJ Jr, Wu YL, et al. Lung cancer: current therapies and new targeted treatments. *Lancet.* (2017) 389:299–311. doi: 10.1016/S0140-6736(16)30958-8
3. Wood SL, Pernemalm M, Crosbie PA, Whetton AD. The role of the tumor-microenvironment in lung cancer-metastasis and its relationship to potential therapeutic targets. *Cancer Treat Rev.* (2014) 40:558–66. doi: 10.1016/j.ctrv.2013.10.001
4. Quail DF, Joyce JA. Microenvironmental regulation of tumor progression and metastasis. *Nat Med.* (2013) 19:1423–37. doi: 10.1038/nm.3394
5. Bussard KM, Mutkus L, Stumpf K, Gomez-Manzano C, Marini FC. Tumor-associated stromal cells as key contributors to the tumor microenvironment. *Breast Cancer Res.* (2016) 18:84. doi: 10.1186/s13058-016-0740-2
6. Gajewski TF, Schreiber H, Fu YX. Innate and adaptive immune cells in the tumor microenvironment. *Nat Immunol.* (2013) 14:1014–22. doi: 10.1038/ni.2703
7. Djenidi F, Adam J, Goubar A, Durgeau A, Meurice G, de Montpreville V, et al. CD8+CD103+ tumor-infiltrating lymphocytes are tumor-specific tissue-resident memory T cells and a prognostic factor for survival in lung cancer patients. *J Immunol.* (2015) 194:3475–86. doi: 10.4049/jimmunol.1402711
8. Kinoshita T, Muramatsu R, Fujita T, Nagumo H, Sakurai T, Noji S, et al. Prognostic value of tumor-infiltrating lymphocytes differs depending on histological type and smoking habit in completely resected non-small-cell lung cancer. *Ann Oncol.* (2016) 27:2117–23. doi: 10.1093/annonc/mdw319
9. Rizvi NA, Mazieres J, Planchard D, Stinchcombe TE, Dy GK, Antonia SJ, et al. Activity and safety of nivolumab, an anti-PD-1 immune checkpoint inhibitor, for patients with advanced, refractory squamous non-small-cell lung cancer (CheckMate 063): a phase 2, single-arm trial. *Lancet Oncol.* (2015) 16:257–65. doi: 10.1016/S1470-2045(15)70054-9
10. Carbone DP, Reck M, Paz-Ares L, Creelan B, Horn L, Steins M, et al. First-line nivolumab in stage IV or recurrent non-small-cell lung cancer. *N Engl J Med.* (2017) 376:2415–26. doi: 10.1056/NEJMoa1613493
11. Mollaoglu G, Jones A, Wait SJ, Mukhopadhyay A, Jeong S, Arya R, et al. The Lineage-defining transcription factors SOX2 and NKX2-1 determine lung cancer cell fate and shape the tumor immune microenvironment. *Immunity.* (2018) 49:764–79.e9. doi: 10.1016/j.immuni.2018.09.020
12. Mascaux C, Angelova M, Vasaturo A, Beane J, Hijazi K, Anthoine G, et al. Immune evasion before tumour invasion in early lung squamous carcinogenesis. *Nature.* (2019) 571:570–5. doi: 10.1038/s41586-019-1330-0
13. Mohamed AJ, Yu L, Backesjo CM, Vargas L, Faryal R, Aints A, et al. Bruton's tyrosine kinase (Btk): function, regulation, and transformation with special emphasis on the PH domain. *Immunol Rev.* (2009) 228:58–73. doi: 10.1111/j.1600-065X.2008.00741.x
14. Kim HO. Development of BTK inhibitors for the treatment of B-cell malignancies. *Arch Pharm Res.* (2019) 42:171–81. doi: 10.1007/s12272-019-01124-1
15. Molina-Cerrillo J, Alonso-Gordoa T, Gajate P, Grande E. Bruton's tyrosine kinase (BTK) as a promising target in solid tumors. *Cancer Treat Rev.* (2017) 58:41–50. doi: 10.1016/j.ctrv.2017.06.001
16. Yue C, Ma H, Zhou Y. Identification of prognostic gene signature associated with microenvironment of lung adenocarcinoma. *PeerJ.* (2019) 7:e8128. doi: 10.7717/peerj.8128
17. Han RH, Dunn GP, Chheda MG, Kim AH. The impact of systemic precision medicine and immunotherapy treatments on brain metastases. *Oncotarget.* (2019) 10:6739–53. doi: 10.18632/oncotarget.27328
18. Wojas-Krawczyk K, Kalinka E, Grenda A, Krawczyk P, Milanowski J. Beyond PD-L1 markers for lung cancer immunotherapy. *Int J Mol Sci.* (2019) 20:E1915. doi: 10.3390/ijms20081915
19. Khanna P, Blais N, Gaudreau PO, Corrales-Rodriguez L. Immunotherapy comes of age in lung cancer. *Clin Lung Cancer.* (2017) 18:13–22. doi: 10.1016/j.cllc.2016.06.006
20. Rhodin KE, Rucker AJ, Ready NE, D'Amico TA, Antonia SJ. The immunotherapeutic landscape in non-small cell lung cancer and its surgical horizons. *J Thorac Cardiovasc Surg.* (2019) 159:1616–23. doi: 10.1016/j.jtcvs.2019.08.138
21. Al-Toubah T, Schell M, Cives M, Zhou JM, Soares H, Strosberg JR. A phase II study of ibrutinib in advanced neuroendocrine neoplasms. *Neuroendocrinology.* (2019). doi: 10.1159/000502383. [Epub ahead of print].
22. Bond DA, Woyach JA. Targeting BTK in CLL: beyond ibrutinib. *Curr Hematol Malig Rep.* (2019) 14:197–205. doi: 10.1007/s11899-019-00512-0
23. Hong D, Rasco D, Veeder M, Luke JJ, Chandler J, Balmanoukian A, et al. A phase 1b/2 study of the bruton tyrosine kinase inhibitor ibrutinib and the PD-L1 inhibitor durvalumab in patients with pretreated solid tumors. *Oncology.* (2019) 97:102–11. doi: 10.1159/000500571
24. Zhang B, Wang L, Zhang Q, Yan Y, Jiang H, Hu R, et al. The Ibr-7 derivative of ibrutinib exhibits enhanced cytotoxicity against non-small cell

lung cancer cells via targeting of mTORC1/S6 signaling. *Mol Oncol.* (2019) 13:946–58. doi: 10.1002/1878-0261.12454
25. Giordano F, Vaira V, Cortinovis D, Bonomo S, Goedmakers J, Brena F, et al. p65BTK is a novel potential actionable target in KRAS-mutated/EGFR-wild type lung adenocarcinoma. *J Exp Clin Cancer Res.* (2019) 38:260. doi: 10.1186/s13046-019-1199-7
26. Gao W, Wang M, Wang L, Lu H, Wu S, Dai B, et al. Selective antitumor activity of ibrutinib in EGFR-mutant non-small cell lung cancer cells. *J Natl Cancer Inst.* (2014) 106:dju204. doi: 10.1093/jnci/dju204
27. Althubiti M, Rada M, Samuel J, Escorsa JM, Najeeb H, Lee KG, et al. BTK modulates p53 activity to enhance apoptotic and senescent responses. *Cancer Res.* (2016) 76:5405–14. doi: 10.1158/0008-5472.CAN-16-0690
28. Rada M, Althubiti M, Ekpenyong-Akiba AE, Lee KG, Lam KP, Fedorova O, et al. BTK blocks the inhibitory effects of MDM2 on p53 activity. *Oncotarget.* (2017) 8:106639–47. doi: 10.18632/oncotarget.22543
29. Rada M, Barlev N, Macip S. BTK: a two-faced effector in cancer and tumour suppression. *Cell Death Dis.* (2018) 9:1064. doi: 10.1038/s41419-018-1122-8
30. Galicia-Vazquez G, Aloyz R. Metabolic rewiring beyond Warburg in chronic lymphocytic leukemia: how much do we actually know? *Crit Rev Oncol Hematol.* (2019) 134:65–70. doi: 10.1016/j.critrevonc.2018.12.003
31. Wennhold K, Shimabukuro-Vornhagen A, von Bergwelt-Baildon M. B Cell-based cancer immunotherapy. *Transfus Med Hemother.* (2019) 46:36–46. doi: 10.1159/000496166
32. Isaeva OI, Sharonov GV, Serebrovskaya EO, Turchaninova MA, Zaretsky AR, Shugay M, et al. Intratumoral immunoglobulin isotypes predict survival in lung adenocarcinoma subtypes. *J Immunother Cancer.* (2019) 7:279. doi: 10.1186/s40425-019-0747-1
33. R Core Team. *R: A Language and Environment for Statistical Computing.* Vienna: R Foundation for Statistical Computing (2018). Available online at: https://www.R-project.org/

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

ORIGINAL RESEARCH
published: 21 April 2020
doi: 10.3389/fonc.2020.00571

Stratification of Patients With Stage IB NSCLC Based on the 8th Edition of the American Joint Committee on Cancer (AJCC) Staging Manual

*Lei-Lei Wu†, Xuan Liu†, Wen-Mei Jiang†, Wei Huang, Peng Lin, Hao Long, Lan-Jun Zhang and Guo-Wei Ma**

State Key Laboratory of Oncology in South China, Sun Yat-sen University Cancer Center, Collaborative Innovation Center for Cancer Medicine, Guangzhou, China

OPEN ACCESS

Edited by:
Paul Takam Kamga, Université de Versailles Saint-Quentin-en-Yvelines, France

Reviewed by:
Alex Friedlaender, Geneva University Hospitals (HUG), Switzerland
Elena Levantini, Harvard Medical School, United States

***Correspondence:**
Guo-Wei Ma
magw@sysucc.org.cn

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Thoracic Oncology, a section of the journal Frontiers in Oncology

Received: *05 November 2019*
Accepted: *30 March 2020*
Published: *21 April 2020*

Citation:
Wu L-L, Liu X, Jiang W-M, Huang W, Lin P, Long H, Zhang L-J and Ma G-W (2020) Stratification of Patients With Stage IB NSCLC Based on the 8th Edition of the American Joint Committee on Cancer (AJCC) Staging Manual. Front. Oncol. 10:571. doi: 10.3389/fonc.2020.00571

Objective: To assess the postoperative prognosis of patients with stage IB non-small cell lung cancer (NSCLC), using a prognostic model (PM).

Methods: Patients with stage IB of NSCLC from the two academic databases {the Surveillance, Epidemiology, and End Results [SEER-A, $N = 1,746$ (training cohort)], Sun Yat-sen University Cancer Center [SYSUCC, $N = 247$ (validation cohort)], and SEER-B ($N = 1,745$)} who had undergone lung surgery from 2001 to 2015 were enrolled. The primary clinical endpoint was cancer-specific survival (CSS). Covariate inclusion of prognostic indicators was carried out using a multivariable two-sided $P < 0.05$. We identified and integrated significant prognostic factors for survival in the training cohort to build a model that could be validated in the validation cohort. We used univariate analysis to evaluate the utilized ability of PM in the different races/ethnicities.

Results: CSS discrimination in the PM was comparable in both the training and validation cohorts [C index = 0.66(SEER-A), 0.67(SYSUCC), and 0.61(SEER-B), respectively]. Discretization with a fixed PM cutoff of 291.5 determined from the training dataset yielded low- and high-risk subgroups with disparate CSS in the validation cohort (training cohort: hazard ratio [HR] 2.724, 95% confidence intervals [CI], 2.074–3.577; validation cohort: SEER-B HR 1.679, 95% CI, 1.310–2.151, SYSUCC HR 3.649, 95% CI 2.203–6.043, all $P < 0.05$). Our five-factor PM was able to predict CSS; 48-month CSS was 87% in the low-risk subgroup vs. 69% in the high-risk subgroup for the training cohort, while in the validation cohort, they were 80 vs. 73%(SEER-B) and 84 vs. 60% (SYSUCC), respectively. In addition, the results showed that PM with all unadjusted HR > 1 was a significant risk prognostic indictor in white men ($P < 0.001$), Chinese people ($P < 0.001$), and other races ($P = 0.012$).

Conclusion: We established and validated a PM that may predict CSS for patients with IB NSCLC in different races/ethnicities, and thus, help clinicians screen subgroups with poor prognosis. In addition, further prospective studies and more cases from different regions are necessary to confirm our findings.

Keywords: NSCLC, stage IB, prognostic model, survival, treatment strategy

INTRODUCTION

Lung cancer remains the most common cause of cancer-related morbidity and mortality (1). In 2019, in the United States alone, the number of new cases is estimated to reach 228,150, and the death toll is projected to be 142,670 (2). Lung cancer is mainly classified into non-small cell lung cancer (NSCLC) and small cell lung cancer (2). More than 83% of lung cancers are NSCLC (2, 3). According to the 8th edition of the American Joint Committee on Cancer (AJCC) Staging Manual that was implemented in January 2017, the stratification effect on the overall survival (OS) rate is better than that in the 7th edition (4). Owing to the tendency of late diagnosis and tumor recurrence (5), the 5-year OS rate for NSCLC remains low at about 23% (2, 6). The decision of administering adjuvant treatment to patients with stage IB has been controversial. The National Comprehensive Cancer Network guidelines recommend postoperative chemotherapy in patients with high-risk factors, such as vascular invasion, visceral pleural invasion, unknown lymph nodes status, and tumor diameter >4 cm (7); the European Society for Medical Oncology guidelines recommend that adjuvant therapy be given to patients with a tumor diameter >4 cm (8) and the American Society of Clinical Oncology guidelines do not recommend routine treatment for stage IB patients (9). However, following the implementation of the 8th edition, patients with stage IB and tumor size >4 cm have been reassigned to stage IIA (4). According to the 8th AJCC Staging Manual, stage IB is defined by the following: (1) tumor size >3-4 cm, with or without visceral pleural invasion (PL1/PL2); (2) tumor size 0–3 cm, with visceral pleural invasion (PL1/PL2); (3) tumor size 0–3 cm, infringing the main bronchus but with a distance ≥2 cm from the carina or with local pneumonia or with local atelectasis (4).

Many studies have confirmed that tumor size ≥4 cm and visceral pleural invasion can worsen the prognosis of lung cancer patients (10–14).

The 5-year OS of patients with stage IA can be as high as 84%, while the 5-year OS of patients with stage IB is slightly poorer at 68%. With the improvement in early screening for lung cancer, the detection rate of stage I patients increases, and the proportion of patients with stage IB increases (4, 6). Therefore, it is more important to screen for high-risk factors of postoperative poor prognosis in patients with IB as per the 8th edition of the AJCC Staging Manual.

This study used the data of the lung cancer patients recorded in the database of the Surveillance, Epidemiology, and End Results (SEER) and Sun Yat-sen University Cancer Center (SYSUCC) to transform the 8th edition of AJCC Staging Manual based on the information provided. We further analyzed the postoperative prognosis of patients with stage IB NSCLC using a prognostic model (PM) and effectively stratified patients as per the AJCC Staging Manual. We believe that this study will provide important treatment-related information for clinicians and patients.

Abbreviations: NSCLC, non-small cell lung cancer; PM, prognostic model; SEER, the Surveillance, Epidemiology, and End Results; SYSUCC, Sun Yat-sen University Cancer Center; CSS, cancer-specific survival; HR, hazard ratio; CI, confidence intervals; AJCC, the American Joint Committee on Cancer; OS, overall survival; PL, pleural invasion; LNs, lymph nodes; AC, adenocarcinoma; SCC, squamous cell carcinoma; BAC, bronchial alveolar carcinoma; NT, neuroendocrine tumor; TSPI, both tumor size of > 3 cm and pleural invasion.

MATERIALS AND METHODS

Study Cohort

The study cohort comprised 3,491 patients from the SEER database who underwent lung surgery from January 2010 to December 2015 and 247 patients from SYSUCC who underwent lung surgery from January 2001 to December 2014. Patients who met the following inclusion criteria were enrolled in the study: (1) histopathologic confirmation of NSCLC diagnosis; (2) no distant metastasis to the lymph nodes (LNs) or other organs; (3) pathologically confirmed stage IB as per the 8th edition of the AJCC Staging Manual. Patients were excluded if they (1) had received adjuvant and neoadjuvant cytotoxic chemotherapy or radiotherapy or immune checkpoint inhibitors or underwent other immune therapy regimens; or (2) had a past or current history of another malignancy. According to the patients' records, we translated the pathological staging into the 8th edition of AJCC. The process of patient screening is shown in **Figures 1**, **2**. All patient records were anonymized before analyses. We included information regarding the following patient information: sex, race, age at diagnosis, surgical approach, tumor differentiation, histologic type, number of LNs removed, positive number of LNs, tumor location, tumor extension status, tumor size, pleural invasion (PL), pT stage, pN stage, pM stage, pTNM stage, chemotherapy, and radiation. Patients from the SEER database were randomized into a training cohort (SEER-A) and a validation cohort (SEER-B). SEER-A included 1,746 patients, while the validation cohort included 1,745 patients (SEER-B) and 247 patients (SYSUCC). We obtained approval to use SYSUCC data from the Research Data Deposit of Sun Yat-sen University Cancer Center (Approval number: RDDA2019001261). The primary clinical endpoint was CSS.

Surgery

According to record in the SEER database and SYSUCC, the main approaches for lung surgery included lobectomy, pneumonectomy, sleeve resection, and sublobectomy (wedge resection and segmental resection). In the SEER database, the average number of LNs removed during surgery was 9.97 ± 0.13, and the median number of LNs was 8.0. However, in the SYSUCC data, the average number of LNs removed during surgery was 20.98 ± 0.79, and the median number of LNs was 19.0

Histologic Type

Patients exhibited the following histologic types: adenocarcinoma (AC), squamous cell carcinoma (SCC), carcinoid tumor, bronchial alveolar carcinoma (BAC), and neuroendocrine tumor (NT).

Follow-Up

The survival time and status information was available for these patients. In the SEER database, follow-up duration ranged from 0.0–83.0 months, with an average of 37.0 ± 0.36 months; in

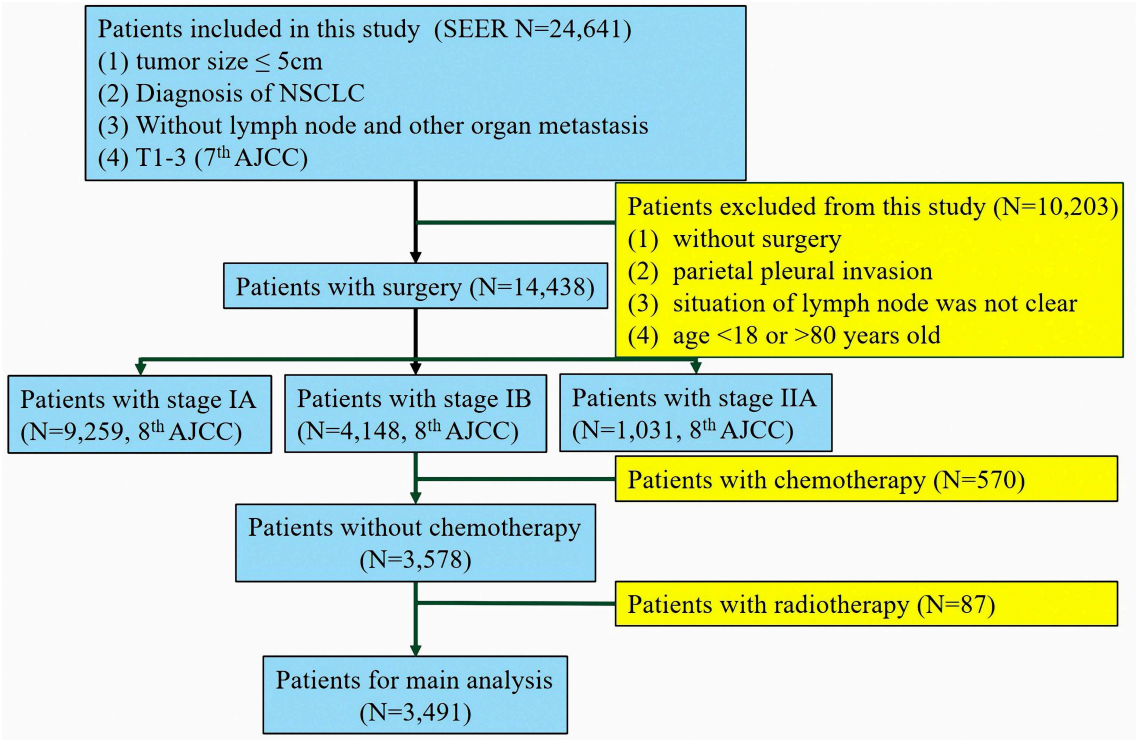


FIGURE 1 | Flow chart of the patient screening process in the the Surveillance, Epidemiology, and End Results.

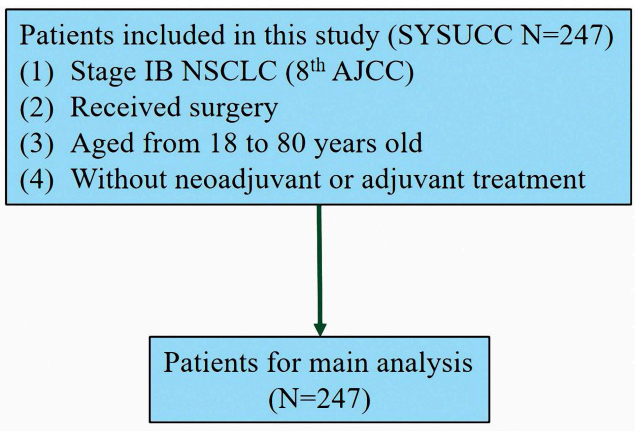


FIGURE 2 | The diagram of the patient screening process in the Sun Yat-sen University Cancer Center.

the SYSUCC, follow-up duration ranged from 1.0–202.0 months, with an average of 68.6 ± 2.29 months.

Statistical Analyses

Statistical analyses were performed using SPSS Statistics 25.0 software (IBM SPSS, Inc., Chicago, IL, USA), X-tile software (15), R version 3.5.2 and Graph pad Prism 5. Hazard ratios (HR) with 95% confidence intervals (CIs) were calculated using multivariate regression analysis. Correlations between groups and clinicopathological characteristics were assessed using the χ^2 test. We then considered information regarding pleural invasion and tumor size and defined patients with both tumor size >3 cm and pleural invasion (TSPI) as TSPI positive, and the other patients as TSPI negative. Multivariate analysis was performed to evaluate the influence of gender, age at diagnosis, race, tumor location, tumor differentiation, surgical approach, histologic type, tumor size and pleura invasion on CSS. A two-sided $p < 0.05$ was considered statistically significant. The most valuable prognostic factors identified using univariate analysis were confirmed with multivariate analysis. Multivariate Cox regression analysis was used to exclude other confounding factors affecting survival. Prognostic indicators were included as covariates in our multivariate analysis with a two-sided *P*-value threshold of <0.05. Similarly, Kaplan–Meier analysis and log-rank tests were used to compare survival curves between groups. Cases were censored when cancer-related death occurred or at the end of follow-up. CSS was selected as the primary clinical endpoint as it was considered the most clinically relevant factor. We adopted a model development and validation approach, using a randomized method to extract the training and validation data sets.

Patient demographics and clinical characteristics were reported for the training cohort. The PM for CSS was constructed by using the linear predictor of the finalized model derived from the training data set. The training cohort was dichotomized into a low-risk and high-risk subgroups using X-tile to determine the

cutoff value of PM. A risk score cutoff was defined for classifying patients in the validation cohorts. Concordance C index was generated for discrimination of the multivariable PM.

In the validation cohorts, the PM was applied to calculate the risk score, and patient discretization into the low- and high-risk subgroups was based on the same cutoffs defined in the training datasets.

To investigate the effect of stratification, we screened patients from the SEER database with stage IA and IIA (stages were translated into the 8th edition AJCC), which included 9,259 and 1,031 patients, respectively. We then compared the survival between patients in stage IA, low-risk stage IB, high-risk stage IB, and stage IIA.

RESULTS

Patient Characteristics

Clinical characteristics of patients in the SEER database are listed in **Table 1**. Among the 3,491 patients, 1,630 (46.7%) were men and 1,861 (53.3%) were women; 2,878 (82.4%) were white, 314 (9.0%) were black, and 288 (8.2%) were of other races. Patients' age ranged from 22–80 years (median, 68 years). In this cohort, the 1-, 3-, and 4-year CSS rates were 91.0, 82.0, and 77.0%, respectively, and the median and mean times from surgery to the last censoring date were 34.0 and 37.0 months, respectively. In the training cohort, the 1-, 3-, and 4-year CSS rates were 91.0, 83.0, and 79.0%, respectively, and in the validation cohorts, the 1-, 3-, and 4-year CSS rates were 90.0, 80.0, and 76.0% (SEER-B) and 92.0, 84.0, and 78.0% (SYSUCC), respectively. Clinical characteristics of patients in the SYSUCC are listed in **Table 2**.

In the training cohort, the number of patients who underwent lobectomy was 1,521 (87.1%). Of the remaining patients, 191 (10.9%) and 34 (0.2%) underwent sublobectomy and pneumonectomy, respectively (**Table 1**). The main histologic type was AC (N = 1,180, 67.6%) and SCC (N = 443, 25.4%). In this cohort, 839 (48.1%) patients had pleural invasion, with the remaining patients accounting for 51.9% (N = 907) of the study population. The majority of tumors were located in the upper lobe (N = 1,039, 59.5%), but some were in the lower lobe (N = 547, 31.3%), some were in the middle lobe (N = 117, 6.7%), and the remaining were in other locations (N = 38, 2.2%), including the main bronchi, multiple positions, etc. 906 (51.9%) patients had ≤ 8 LNs removed, while 840 (48.1%) had > 8 LNs removed. Regarding the degree of tumor differentiation, 321 (18.4%) were well-differentiated, 888 (50.6%) were moderately differentiated, 524 (30.0%) were poorly differentiated, and 13 (0.7%) were undifferentiated.

TABLE 1 | The associations of clinicopathological characteristics between training cohort (SEER-A) and validation cohort (SEER-B).

	All patients (N = 3,491)	Training Cohort (SEER-A, N = 1,746)	Validation Cohort (SEER-B, N = 1,745)	
Variables	**No. of patients (%)**			***P*-value**
Sex				0.446
Male	1,630 (46.7%)	804 (49.3%)	826 (50.7%)	
Female	1,861 (53.3%)	942 (50.6%)	919 (49.4%)	
Age at diagnosis (years)				0.397
≤65	1,417 (40.6%)	721 (50.9%)	696 (49.1%)	
>65	2,074 (59.4%)	1,025 (49.4%)	1,049 (50.6%)	
Race				0.745
White	2,878 (82.4%)	1,430 (49.7%)	1,148 (50.3%)	
Black	314 (9.0%)	166 (52.9%)	148 (47.1%)	
Other	288 (8.2%)	144 (50.0%)	144 (50.0%)	
Surgery Approach				0.460
Lobectomy	3,045 (87.2%)	1,521 (50.0%)	1,524 (50.0%)	
Sublobectomy	382 (10.9%)	191 (50.0%)	191 (50.0%)	
Pneumonectomy	62 (1.8%)	34 (54.8%)	28 (45.2%)	
LNs				0.285
≤8	1,843 (52.8%)	906 (49.2%)	937 (50.8%)	
>8	1,648 (47.2%)	840 (51.0%)	808 (49.0%)	
Tumor grade				0.402
Grade I	616 (17.6%)	321 (52.1%)	295 (47.9%)	
Grade II	1,768 (50.6%)	888 (50.2%)	880 (49.8%)	
Grade III	1,075 (30.8%)	524 (48.7%)	551 (51.3%)	
Grade IV	32 (0.9%)	13 (40.6%)	19 (59.4%)	
Histologic type				0.337*
Carcinoid	6 (0.2%)	5 (83.3%)	1 (16.7%)	
BAC	109 (3.1%)	50 (45.9%)	59 (54.1%)	
AC	2,187 (62.6)	1,108 (50.7%)	1,079 (49.3%)	
SCC	895 (25.6%)	443 (49.5%)	452 (50.5%)	
NT	294 (8.4%)	140 (47.6%)	154 (52.4%)	
Pleura invasion				0.412
Negative	1,524 (43.7%)	776 (50.9%)	748 (49.1%)	
Positive	1,696 (48.6%)	839 (49.5%)	857 (50.5%)	
Tumor Location				0.216
Upper lobe	2,114 (60.5%)	1,039 (49.1%)	1,075 (50.9%)	
Middle lobe	210 (6.0%)	117 (55.7%)	93 (44.3%)	
Lower lobe	1,068 (30.6%)	547 (51.2%)	521 (48.8%)	
Other location	84 (2.4%)	38 (45.2%)	46 (54.8%)	

P value was calculated by Fisher's exact test; P value was calculated by χ^2 test.*
SEER, the Surveillance, Epidemiology, and End Results; AC, adenocarcinoma; SCC, squamous cell carcinoma; BAC, bronchial alveolar carcinoma; NT, neuroendocrine tumor.

Univariate and Multivariate Analyses

Univariate and multivariate analyses were performed to investigate the correlations between the clinical characteristics and CSS. As shown in **Table 3**, univariate analyses identified the following clinical characteristics as significant CSS prognostic factors in patients with NSCLC: gender, age at diagnosis, lobectomy, sublobectomy, LNs, tumor differentiation, AC, SCC, and pleura invasion. Further multivariate analysis based on those characteristics confirmed gender (HR 0.700, 95% CI, 0.542–0.904, P = 0.006), age at diagnosis (HR 1.039, 95% CI, 1.023–1.056, P < 0.001), LNs (HR 0.974, 95% CI, 0.954–0.994, P = 0.012), tumor differentiation (HR 1.496, 95% CI, 1.235–1.813, P < 0.001), and pleura invasion (HR 1.459, 95% CI, 1.123–1.894, P = 0.005) as independent prognostic factors (**Table 3**). Our study revealed that these factors were significantly associated

TABLE 2 | The clinicopathological characteristics in Sun Yat-sen University Cancer Center.

Variables	No. of patients (%) *N* = 247
Sex	
Male	153 (61.9%)
Female	94 (38.1%)
Race/ethnicities	
Chinese	247 (100.0%)
Age (years)	
≤65	170 (68.8%)
>65	77 (31.2%)
Differentiation	
Grade I	32 (13.0%)
Grade II	127 (51.4%)
Grade III	88 (35.6%)
Chemotherapy	
No	247 (100.0%)
Yes	0 (0.0%)
Radiation	
No	247 (100.0%)
Yes	0 (0.0%)
Pleura invasion	
No	54 (21.9%)
Yes	193 (78.1%)
Tumor location	
Upper	133 (53.8%)
Middle	26 (10.5%)
Lower	77 (31.2%)
Other	7 (2.8%)
Surgery approach	
Sublobectomy	0 (0.0%)
Lobectomy	242 (98.0%)
Pneumonectomy	5 (2.0%)

with prognosis in stage IB NSCLC patients. Therefore, the five factors mentioned above were useful predictors of postoperative outcome in the training cohort.

Construction of a PM

Based on the results of the training cohort information analyses, we constructed the PM system and tested the covariates listed in **Table 4** for their association with CSS. The PM system was based on weighting (derived from the β-coefficient of the respective log[HRs]) of the five significant covariates in the training cohort (**Table 4**) that yielded a C index of 0.66 (95% CI, 0.64–0.68) for CSS. This model allowed us to define a high-risk subgroup presenting a significantly reduced likelihood of survival (HR 2.724, 95% CI, 2.074–3.577; $P < 0.001$, **Figure 3A**). The PM cutoff value was determined in order to distinguish the high-risk group from the low-risk group, using the X-tile software. The cutoff value was 291.5. Our five-factor PM predicted that the 12-month, 36-month, and 48-month CSS in the low-risk subgroup vs. that in the high-risk subgroup was 95.0 vs. 87.0%, 90.0 vs. 74.0%, and 87.0 vs. 69.0%, respectively, in the training cohort, 94.0 vs. 87.0%, 85.0 vs. 75.0%, and 80.0 vs. 73.0% (SEER-B) and 96.0 vs. 81.0%, 91.0 vs. 64.0% and 84.0 vs. 60.0% (SYSUCC), respectively, in the validation cohort.

Validation of the PM

In order to validate the predictive accuracy of the PM for CSS in IB NSCLC, we tested the PM independently in the validation cohort: an internal cohort of 1,745 patients and an external cohort of 247 patients. The same PM cutoff value of 291.5 allowed us to stratify patients in the validation cohort into the high-risk subgroup with a significantly inferior CSS or the low-risk subgroup (SEER-B: HR 1.679, 95% CI, 1.310–2.151, $P < 0.001$; SYSUCC: HR 3.649, 95% CI 2.203–6.043, $P < 0.001$, **Figures 3B,C**). The PM in the validation cohorts yielded a C index of 0.61 [95% CI, 0.60–0.63, (SEER-B)] and 0.67 [95% CI, 0.64–0.71, (SYSUCC)] for CSS.

In the SYSUCC, the median survival time of the high-risk subgroup was 76.0 months. However, there was no median survival time in SEER-A, SEER-B, and low-risk subgroup of SYSUCC.

Effect of Stratification

To observe the effect of stratification, we screened patients with stage IA and IIA who were translated into the 8th edition AJCC of the SEER database, which included 9,259 and 1,031 patients, respectively. The high-risk and low-risk group stage IB patients were compared with the stage IIA, and IA. We found that stage IA NSCLC patients had the highest CSS in the observation period ($P < 0.001$, **Figure 4A**). We found that there was no significant difference between stage IA and low risk stage IB in cancer-specific survival ($P = 0.029$, **Figure 4B**). High-risk stage IB patients did not have a significantly lower CSS than stage IIA patients ($P = 0.87$, **Figure 4C**).

Impact of PM on Different Races/Ethnicities

We hoped to further explore the impact of PM on different races/ethnicities. Accordingly, univariate analysis was used to estimate the association between PM and CSS. Our results showed that unadjusted HR exceeded 1 or, in other words, PM could be a risk indictor among different races/ethnicities (**Figure 5**). In addition, there were significant differences in white men ($P < 0.001$), other races ($P = 0.012$), and Chinese people ($P < 0.001$), while no significant differences were observed for black race ($P = 0.45$).

DISCUSSION

The occurrence and development of NSCLC is complex, and decisions regarding the administration of adjuvant therapy for stage IB NSCLC patients remains controversial. Some research studies have suggested that patients with stage IB NSCLC could benefit from adjuvant therapy (16–18), while other studies have reported no effects of adjuvant chemotherapy on patients (9, 11, 12, 19–21). Studies that have shown the benefit of adjuvant chemotherapy in stage

TABLE 3 | Univariate and multivariate Cox regression analysis for cancer-specific survival in patients with stage IB NSCLC (Cox regression's method is Forward: LR).

	Univariate analysis			Multivariate analysis		
	HR	**95% CI**	***P*-value**	**HR**	**95% CI**	***P*-value**
Gender						
Male/Female	0.641	0.502–0.818	**<0.001**	0.700	0.542–0.904	**0.006**
Age at diagnosis (years)						
Continuous	1.037	1.022–1.053	**<0.001**	1.039	1.023–1.056	**<0.001**
Surgery approach						
Lobectomy	0.655	0.474–0.906	**0.011**	NA	NA	0.084
Sublobectomy	1.457	1.024–2.072	**0.036**	NA	NA	0.371
Pneumonectomy	NA	NA	0.137			
LNs						
Continuous	0.973	0.954–0.992	**0.005**	0.974	0.954–0.994	**0.012**
Tumor differentiation						
Grade I vs. II vs. III vs. IV	1.486	1.246–1.771	**<0.001**	1.496	1.235–1.813	**<0.001**
Histologic type						
Carcinoid	NA	NA	0.067			
BAC	NA	NA	0.464			
AC	0.753	0.589–0.963	**0.024**	NA	NA	0.341
SCC	1.416	1.090–1.840	**0.009**	NA	NA	0.921
Neuroendocrine	NA	NA	0.965			
Tumor size						
Continuous	NA	NA	0.062			
Pleura invasion						
No/yes	1.547	1.192–2.006	**0.001**	1.459	1.123–1.894	**0.005**
Tumor location						
Upper	NA	NA	0.663			
Middle	NA	NA	0.564			
Lower	NA	NA	0.796			
Other	NA	NA	0.891			

NSCLC, non-small cell lung cancer; AC, adenocarcinoma; SCC, squamous cell carcinoma; BAC, bronchial alveolar carcinoma; NT, neuroendocrine tumor. The meaning of bold values is two-sided P < 0.05.

TABLE 4 | Constructed prognostic score to predict cancer-specific survival in stage IB NSCLC patients.

Covarite	**β [HR = exp (β)]**	**Score**
Gender	−0.356	−0.356 * (1/2; male = 1, female = 2)
Age	0.039	0.039 * Age at diagnosis
Nodes examined	−0.026	−0.026 * number of nodes examined
Grade	0.403	0.403* (1/2/3/4; Grade I = 1, Grade II = 2, Grade III = 3, Grade IV = 4)
Pleura invasion	0.378	0.378 * (0/1; no = 0, yes = 1)
		Total computed score *100
Risk stratification		
Low risk		≤291.5
High risk		>291.5

NSCLCL, non-small cell lung cancer.

IB patients tend to recommend adjuvant therapy for patients with tumor size ≥4 cm (7, 8, 12, 22, 23). However, stage IB (7th AJCC) disease with a tumor diameter >4 cm has been classified as stage IIA (8th AJCC) (4, 24). One retrospective study based on the 8th edition of the AJCC Staging Manual has shown that postoperative adjuvant treatment could benefit stage IB NSCLC patients (24). A recent meta-analysis, which included 9 randomized collected trials, suggested that patients with stage IB might not need adjuvant chemotherapy; however, the stage IB was based on the 7th AJCC in all trials (25).

Based on the above results, some researchers hoped to provide information regarding postoperative treatment decisions by studying the prognosis of early-stage patients. Factors such as age, pathological type, LINE-1 hypomethylation, individualized immune prognostic signature, quality measures, tumor size, preoperative platelet-to-lymphocyte ratio and lymphocyte-to-monocyte ratio, and visceral pleural invasion, were found to influence the prognosis of early-stage patients (7, 10, 11, 13, 14, 22, 26–30). However, the above mentioned studies were unable to individually predict the prognosis of patients. This study aimed to construct an individualized prognostic model and to provide useful information to support clinicians' decisions. We hope to

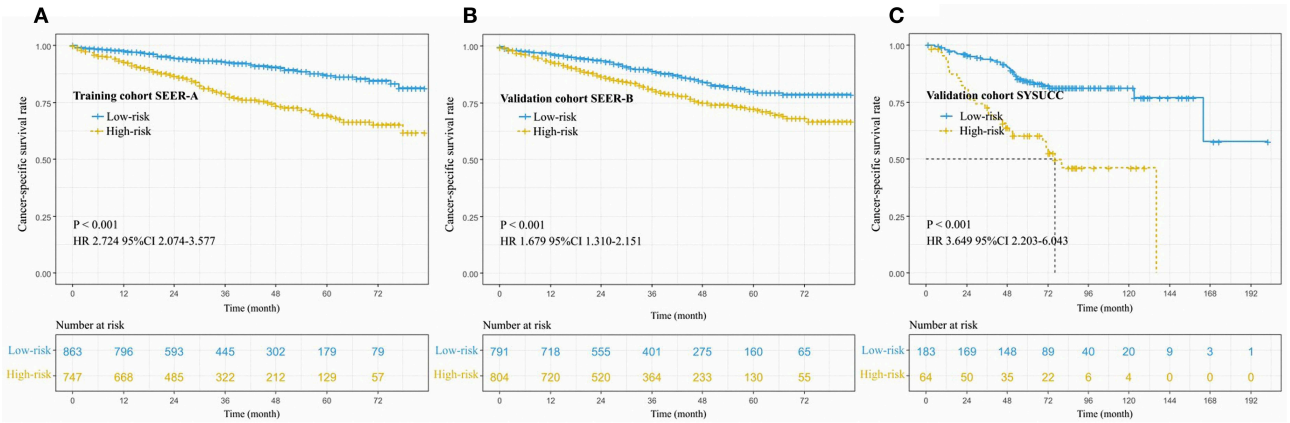


FIGURE 3 | Cancer-specific survival curve for patients with stage IB NSCLC according to the prognostic model in the training cohort **(A)**, internal validation cohort **(B)**, and external validation cohort **(C)**.

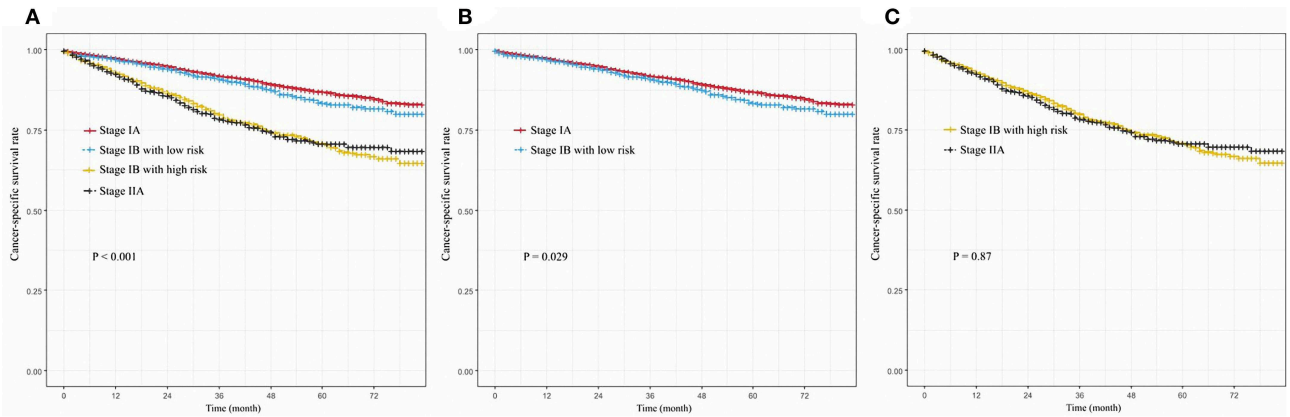


FIGURE 4 | (A) Cancer-specific survival curve for NSCLC patients with stage IA, low-risk group of IB, high-risk group of IB, and stage IIA**; (B)** Cancer-specific survival curve for NSCLC patients with stage IA, and low-risk group of IB; **(C)** Cancer-specific survival curve for NSCLC patients with stage IIA, and high-risk group of IB.

build a simple model by using some commonly obtained patient information. During the course of this research, we analyzed the patients' clinical information, including the indicators shown in **Table 1**. Eventually, five meaningful indicators were selected using univariate and multivariate analyses of the training cohort, including gender, age at diagnosis, white race, number of nodes removed, tumor differentiation, and pleura invasion. In this study, we considered information regarding pleural invasion and tumor size, based on which we defined patients with both tumor size >3 cm and pleural invasion (TSPI) as TSPI positive, and the rest of the patients as TSPI negative. We found that TSPI could be a risk prognostic factor (**Figure 6**). During data processing, the number of removed lymph nodes was considered to be a protective prognostic factor (**Table 3**).

We constructed a PM based on the above five indicators and successfully identified high-risk and low-risk populations in the training and validation cohorts. Our model had a significant impact on patient differentiation (**Figure 3**), because the C index for predicting CSS rates reached 0.66(SEER-A), 0.61(SEER-B), and 0.67(SYSUCC) in the training and validation cohorts,

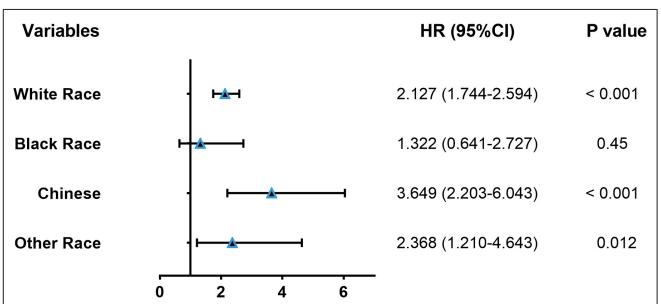


FIGURE 5 | Impact of prognostic model on survival in different races/ethnicities.

respectively. Even in comparison with stage IA and IIA, there was no significant difference in survival between the IB stage of the high-risk group and IIA stage (**Figure 4**). In terms of the clinical application, these indicators can be easily assessed. Information regarding sex and age can be obtained from the admission

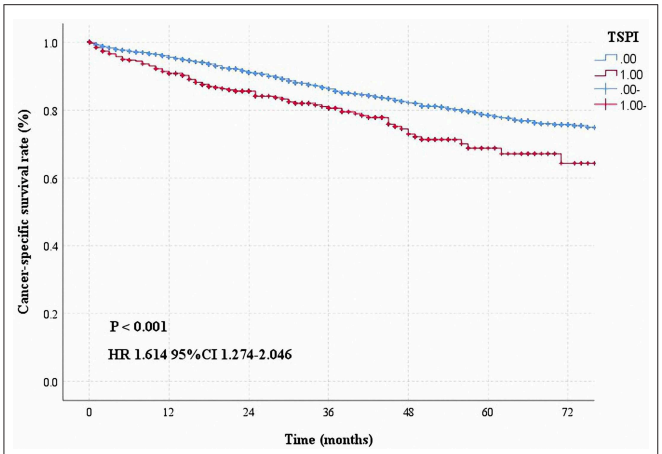


FIGURE 6 | Cancer-specific survival curve for stage IB NSCLC according to the status of TSPI (0: negative, 1: positive).

records, and data on the degree of tumor differentiation, status of pleural invasion, and number of dissected LNs can be obtained from postoperative pathology reports. Clinicians could use the above information and our PM to calculate scores of NSCLC patients with stage IB after surgery, and give patients advice on whether adjuvant therapy is necessary according to prediction of prognosis. In addition, this study included internal and external validation, thus promoting a wide range of applications of the model. According to results of validation of SYSUCC and SEER-B, we found that PM might be applied in different races/ethnicities (**Figure 5**). We noticed that the clinical popularization of gene test, such as EGFR, in some regions was inadequate (31–33). Therefore, to some extent, this PM in these patients who lack the results of molecular test may have a certain value of utility.

This study has certain limitations. First, the study used the SEER and SYSUCC database in which the distribution of ethnic groups is not balanced. It would be recommendable to include data from different regions in our study, which would balance the race/ethnicity distribution and make the results more generalized. Second, based on the limitations of the SEER database, information on chemotherapy was not comprehensive enough. We do not know whether neoadjuvant chemotherapy or adjuvant chemotherapy was administered, and therefore, when comparing with patients with high-risk stage IB to low-risk stage IB, it is not possible to conclude that patients with adjuvant chemotherapy have a better prognosis. In addition, the number of removed lymph nodes is quite different between SEER and SYSUCC, and the sample size for external validation is small so the number of high-risk patients in the SYSUCC is also relatively small. Thirdly, in the era of precision medicine, molecular detection plays an important role in judging the prognosis and treatment of patients. However, the information of driver genes is incomplete in the data of SEER and SYSUCC. Therefore, based on this research, information on molecular indicators such as EGFR, KRAS, TP53, and ALK can be collected (34–37). Information on these driver genes may increase the predictive ability of PM on CSS. In addition, we couldn't obtain complete information of pathological features such as vascular invasion, which may have an impact on prognosis, in the databases of SEER and SYSUCC. Further, only patients with stage IB NSCLC (8th AJCC) were enrolled; therefore, this model cannot predict or assess CSS in patients with a tumor size $\leq$4 cm and may only be applied to patients with stage IB NSCLC (8th AJCC). Eventually, further prospective and multicenter studies are necessary to confirm our findings.

DATA AVAILABILITY STATEMENT

Data from this study are available to any interested researchers upon reasonable request to the corresponding author.

ETHICS STATEMENT

The study was approved by the Clinical Research Ethic Committee of SYSUCC (IRB number: B2019-116-01), and informed consent of patients was waived.

AUTHOR CONTRIBUTIONS

L-LW designed the research and wrote the article. XL and W-MJ processed data. WH recorded the data of patients. PL, HL, and L-JZ reviewed and edited the article. G-WM analyzed the data and reviewed the article.

REFERENCES

1. 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. doi: 10.3322/caac.21492
2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin.* (2019) 69:7–34. doi: 10.3322/caac.21551
3. Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, et al. Cancer statistics in China 2015. *CA Cancer J Clin.* (2016) 66:115–32. doi: 10.3322/caac.21338
4. Rami-Porta R, Asamura H, Travis WD, Rusch VW. Lung cancer - major changes in the American joint committee on cancer eighth edition cancer staging manual. *CA Cancer J Clin.* (2017) 67:138–55. doi: 10.3322/caac.21390
5. Hung JJ, Jeng WJ, Hsu WH, Huang BS, Wu YC. Time trends of overall survival and survival after recurrence in completely resected stage I non-small cell lung cancer. *J Thorac Oncol.* (2012) 7:397–405. doi: 10.1097/JTO.0b013e31823b564a
6. Miller KD, Nogueira L, Mariotto AB, Rowland JH, Yabroff KR, Alfano CM, et al. Cancer treatment and survivorship statistics, 2019. *CA Cancer J Clin.* (2019) 69:363–85. doi: 10.3322/caac.21565
7. National Comprehensive Cancer Network. *Non-Small Cell Lung Cancer (Version 7.2019).* Availble online at: https://www.nccn.org/professionals/physician_gls/pdf/nscl.pdf (accessed August 30, 2019).
8. Postmus PE, Kerr KM, Oudkerk M, Senan S, Waller DA, Vansteenkiste J, et al. Early and locally advanced non-small-cell lung cancer (NSCLC): ESMO clinical practice guidelines for diagnosis, treatment and follow-up. *Ann Oncol.* (2017) 28: iv1–21. doi: 10.1093/annonc/mdx222

9. Kris MG, Gaspar LE, Chaft JE, Kennedy EB, Azzoli CG, Ellis PM, et al. Adjuvant systemic therapy and adjuvant radiation therapy for stage I to IIIA completely resected non–small-cell lung cancers: American society of clinical oncology/cancer care ontario clinical practice guideline update. *J Clin Oncol.* 35:2960–74. doi: 10.1200/JCO.2017.72.4401
10. Sakakura N, Mizuno T, Kuroda H, Arimura T, Yatabe Y, Yoshimura K, et al. The eighth TNM classification system for lung cancer: a consideration based on the degree of pleural invasion and involved neighboring structures. *Lung Cancer.* (2018) 118:134–8. doi: 10.1016/j.lungcan.2018.02.009
11. Park HJ, Park HS, Cha YJ, Lee S, Jeung HC, Cho JY, et al. Efficacy of adjuvant chemotherapy for completely resected stage IB non-small cell lung cancer: a retrospective study. *J Thorac Dis.* (2018) 10:2279–87. doi: 10.21037/jtd.2018.03.184
12. Li X, Zhang C, Sun Z, Yang F, Xiao R, Sui X, et al. Propensity-matched analysis of adjuvant chemotherapy for completely resected stage IB non-small-cell lung cancer patients. *Lung Cancer.* (2019) 133:75–82. doi: 10.1016/j.lungcan.2019.04.024
13. Iwano S, Umakoshi H, Kamiya S, Yokoi K, Kawaguchi K, Fukui T, et al. Postoperative recurrence of clinical early-stage non-small cell lung cancers: a comparison between solid and subsolid nodules. *Cancer Imag.* (2019) 19:33. doi: 10.1186/s40644-019-0219-3
14. Cho BC, DE Pas T, Kalofonos H, Wang Q, Ramlau R, Cheng Y, et al. Prognostic factors in early-stage NSCLC: analysis of the placebo group in the MAGRIT study. *Anticancer Res.* (2019) 39:1403–9. doi: 10.21873/anticanres.13255
15. Camp RL, Dolled-Filhart M, Rimm DL. Rimm. X-tile: a new bio-informatics tool for biomarker assessment and outcome-based cut-point optimization. *Clin Cancer Res.* (2004) 10:7252–9. doi: 10.1158/1078-0432.CCR-04-0713
16. Woodard GA, Wang SX, Kratz JR, Zoon-Besselink CT, Chiang CY, Gubens MA, et al. Adjuvant chemotherapy guided by molecular profiling and improved outcomes in early stage, non-small-cell lung cancer. *Clin Lung Cancer.* (2018) 19:58–64. doi: 10.1016/j.cllc.2017.05.015
17. Moon H, Zhao Y, Pluta D, Ahn H. Subgroup analysis based on prognostic and predictive gene signatures for adjuvant chemotherapy in early-stage non-small-cell lung cancer patients. *J Biopharm Stat.* (2018) 28:750–62. doi: 10.1080/10543406.2017.1397006
18. Williams CD, Gajra A, Ganti AK, Kelley MJ. Use and impact of adjuvant chemotherapy in patients with resected non-small cell lung cancer. *Cancer.* (2014) 120:1939–47. doi: 10.1002/cncr.28679
19. Hainsworth JD, Waterhouse DM, Shih KC, Boccia RV, Priego VM, McCleod MJ, et al. Phase II trial of preoperative pemetrexed plus carboplatin in patients with stage IB-III nonsquamous non-small cell lung cancer (NSCLC). *Lung Cancer.* (2018) 118:6–12. doi: 10.1016/j.lungcan.2018.01.009
20. Nagasaka M, Gadgeel SM. Role of chemotherapy and targeted therapy in early-stage non-small cell lung cancer. *Expert Rev Anticancer Ther.* (2018) 18:63–70. doi: 10.1080/14737140.2018.1409624
21. Scagliotti GV, Pastorino U, Vansteenkiste JF, Spaggiari L, Facciolo F, Orlowski TM, et al. Randomized phase III study of surgery alone or surgery plus preoperative cisplatin and gemcitabine in stages IB to IIIA non-small-cell lung cancer. *J Clin Oncol.* (2012) 30:172–8. doi: 10.1200/JCO.2010.33.7089
22. Naylor EC. Adjuvant therapy for stage I and II non-small cell lung cancer. *Surg Oncol Clin N Am.* (2016) 25:585–99. doi: 10.1016/j.soc.2016.03.003
23. Strauss GM, Herndon JE, Maddaus MA, Johnstone DW, Johnson EA, Harpole DH, et al. Adjuvant paclitaxel plus carboplatin compared with observation in stage IB non-small-cell lung cancer: CALGB 9633 with the cancer and leukemia group B, radiation therapy oncology group, and north central cancer treatment group study groups. *J Clin Oncol.* (2008) 26:5043–51. doi: 10.1200/JCO.2008.16.4855
24. Wang J, Wu N, Lv C, Yan S, Yang Y. Should patients with stage IB non-small cell lung cancer receive adjuvant chemotherapy? A comparison of survival between the 8th and 7th editions of the AJCC TNM staging system for stage IB patients. *J Cancer Res Clin Oncol.* (2019) 145:463–9. doi: 10.1007/s00432-018-2801-7
25. Li R, Yang G, Tian Y, Tian D. Comparing the benefits of postoperative adjuvant chemotherapy vs. observation for stage IB non-small cell lung cancer: a meta-analysis. *J Thorac Dis.* (2019) 11:3047–54. doi: 10.21037/jtd.2019.07.47
26. Park B, Lee G, Kim HK, Choi YS, Zo JI, Shim YM, et al. A retrospective comparative analysis of elderly and younger patients undergoing pulmonary resection for stage I non-small cell lung cancer. *World J Surg Oncol.* (2016) 14:13. doi: 10.1186/s12957-015-0762-8
27. Imperatori A, Sahnane N, Rotolo N, Franzi F, Nardecchia E, Libera L, et al. LINE-1 hypomethylation is associated to specific clinico-pathological features in stage I non-small cell lung cancer. *Lung Cancer.* (2017) 108:83–9. doi: 10.1016/j.lungcan.2017.03.003
28. Li B, Cui Y, Diehn M, Li R. Development and validation of an individualized immune prognostic signature in early-stage nonsquamous non-small cell lung cancer. *JAMA Oncol.* (2017) 3:1529–37. doi: 10.1001/jamaoncol.2017.1609
29. Samson P, Crabtree T, Broderick S, Kreisel D, Krupnick AS, Patterson GA, et al. Quality measures in clinical stage I non-small cell lung cancer: improved performance is associated with improved survival. *Ann Thorac Surg.* (2017) 103:303–11. doi: 10.1016/j.athoracsur.2016.07.003
30. Chen Y, Wang W, Zhang X, Yu X, Xi K, Wen Y, et al. Prognostic significance of combined preoperative platelet-to-lymphocyte ratio and lymphocyte-to-monocyte ratio in patients undergoing surgery with stage IB non-small-cell lung cancer. *Cancer Manag Res.* (2018) 10:5411–22. doi: 10.2147/CMAR.S177320
31. Shen C, Kehl KL, Zhao B, Simon GR, Zhou S, Giordano SH. Utilization patterns and trends in epidermal growth factor receptor (EGFR) mutation testing among patients with newly diagnosed metastatic lung cancer. *Clin Lung Cancer.* (2017) 18:e233–41. doi: 10.1016/j.cllc.2016.11.002
32. Cheema PK, Raphael S, El-Maraghi R, Li J, McClure R, Zibdawi L, et al. Rate of EGFR mutation testing for patients with nonsquamous non-small-cell lung cancer with implementation of reflex testing by pathologists. *Curr Oncol.* (2017) 24:16–22. doi: 10.3747/co.24.3266
33. Cheng Y, Wang Y, Zhao J, Liu Y, Gao H, Ma K, et al. Real-world EGFR testing in patients with stage IIIB/IV non-small-cell lung cancer in North China: a multicenter, non-interventional study. *Thorac Cancer.* (2018) 9:1461–9. doi: 10.1111/1759-7714.12859
34. Zhao J, Han Y, Li J, Chai R, Bai C. Prognostic value of KRAS/TP53/PIK3CA in non-small cell lung cancer. *Oncol Lett.* (2019) 17:3233–40. doi: 10.3892/ol.2019.10012
35. Sibon D, Nguyen DP, Schmitz N, Suzuki R, Feldman AL, Gressin R, et al. ALK-positive anaplastic large-cell lymphoma in adults: an individual patient data pooled analysis of 263 patients. *Haematologica.* (2019) 104:e562–5. doi: 10.3324/haematol.2018.213512
36. Shepherd FA, Domerg C, Hainaut P, Jänne PA, Pignon JP, Graziano S, et al. Pooled analysis of the prognostic and predictive effects of KRAS mutation status and KRAS mutation subtype in early-stage resected non-small-cell lung cancer in four trials of adjuvant chemotherapy. *J Clin Oncol.* (2013) 31:2173–81. doi: 10.1200/JCO.2012.48.1390
37. Shepherd FA, Lacas B, Le Teuff G, Hainaut P, Jänne PA, Pignon JP, et al. Pooled analysis of the prognostic and predictive effects of TP53 comutation status combined with KRAS or EGFR mutation in early-stage resected non–small-cell lung cancer in four trials of adjuvant chemotherapy. *J Clin Oncol.* (2017) 35:2018–27. doi: 10.1200/JCO.2016.71.2893

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

ORIGINAL RESEARCH
published: 08 May 2020
doi: 10.3389/fonc.2020.00691

Immunoscore Predicts Survival in Early-Stage Lung Adenocarcinoma Patients

Zihuan Zhao[1,2,3], *Dan Zhao*[4], *Ji Xia*[2], *Yi Wang*[3,5]* and *Buhai Wang*[1,2]*

[1] Department of Oncology, Subei People's Hospital of Jiangsu Province, Yangzhou, China, [2] Dalian Medical University, Dalian, China, [3] Department of Respiratory Disease, Nanjing Chest Hospital, Nanjing, China, [4] Department of Reproductive Center, Zhen Jiang Fourth People Hospital, Jiangsu, China, [5] Nanjing Medical University, Nanjing, China

Background: The lung cancer staging system is insufficient for a comprehensive evaluation of patient prognosis. We constructed a novel immunoscore model to predict patients with high risk and poor survival.

Method: Immunoscore was developed based on z-score transformed enrichment score of 11 immune-related gene sets of 109 immune risk genes. The immunoscore model was trained in lung adenocarcinoma cohort from The Cancer Genome Atlas (TCGA-LUAD) (n = 400), and validated in other two independent cohorts from Gene Expression Omnibus (GEO), GSE31210 (n = 219) and GSE68465 (n = 356). Meta-set (n = 975) was formed by combining all training and testing sets.

Result: High immunoscore conferred worse prognosis in all sets. It was an independent prognostic factors in multivariate Cox analysis in training, testing and meta-set [hazard ratio (HR) = 2.96 (2.24–3.9), $P < 0.001$ in training set; HR = 1.99 (1.21–3.26), $P = 0.006$ in testing set 1; HR = 1.48 (1.69–2.39), $P = 0.005$ in testing set 2; HR = 2.01 (1.69–2.39), $P < 0.001$ in meta-set]. Immunoscore-clinical prognostic signature (ICPS) was developed by integrating immunoscore and clinical characteristic, and had higher C-index than immunoscore or stage alone in all sets [0.72 (ICPS) vs. 0.7 (immunoscore) or 0.59 (stage) in training set; 0.75 vs. 0.72 or 0.7 in testing set 1; 0.65 vs. 0.61 or 0.62 in testing set 2; 0.7 vs. 0.66 or 0.64 in meta-set]. Genome analysis revealed that immunoscore was positively correlated with tumor mutation burden ($R = 0.22$, $P < 0.001$). Besides, high immunoscore was correlated with high proportion of carcinoma-associated fibroblasts ($R = 0.32$, $P < 0.001$) in tumor microenvironment but fewer CD8+ cells infiltration ($R = -0.28$, $P < 0.001$).

Conclusion: The immunoscore and ICPS are potential biomarkers for evaluating patient survival. Further investigations are required to validate and improve their prediction accuracy.

Keywords: immunoscore, lung adenocarcinoma, prognosis, immune gene set, ridge regression

OPEN ACCESS

Edited by:
Marius Tresor Chiasseu, Yale University, United States

Reviewed by:
Jai Narendra Patel, Levine Cancer Institute, United States
Sandip Patel, University of California, San Diego, United States

***Correspondence:**
Yi Wang
wangyi2104@njmu.edu.cn
Buhai Wang
wbhself@sina.com

Specialty section:
This article was submitted to Thoracic Oncology, a section of the journal Frontiers in Oncology

Received: *11 November 2019*
Accepted: *14 April 2020*
Published: *08 May 2020*

Citation:
Zhao Z, Zhao D, Xia J, Wang Y and Wang B (2020) Immunoscore Predicts Survival in Early-Stage Lung Adenocarcinoma Patients. Front. Oncol. 10:691. doi: 10.3389/fonc.2020.00691

INTRODUCTION

Lung cancer ranks the top of cancer-related death worldwide (1). Histologically, 15 percent of patients are categorized as small cell lung cancer (SCLC) while the other 85% as non-small cell lung cancer (NSCLC) (2). Among NSCLC, lung adenocarcinoma (LUAD) is the most common subtype (3). Surgical resection remains to be the standard clinical practice for patients with early-stage LUAD (4), and the 5-year survival rate is about 60% (5). Platin-based adjuvant chemotherapy has demonstrated the improvement of 5-year survival for stage II–IIIA patients for about 5%, at the price of chemotherapy-induced toxicity (6, 7). Adjuvant immunotherapy with immune checkpoint inhibitors has come into several clinical trials, but no definitive effectiveness made so far (8). Although using the American Joint Committee on Cancer (AJCC) TNM staging system improves prognostic prediction, it is still inconclusive due to other unknown factors. Thus, the development of new biomarkers is imperative for stratifying risk and optimizing treatment for lung cancer patients with early stage.

Tumor immune microenvironment (TIM) has long been recognized as a crucial factor in cancer progression and metastasis (9). Several studies have explored the TIM as a prognostic biomarker in lung cancer (10). For example, Brambilla et al. found higher CD4+/CD8+ ratio conferred better survival in patients with NSCLC (11). Also, for cancer cell itself, programmed cell death protein ligand 1 (PD-L1) expression and tumor mutation burden (TMB) have been used to predict outcome in NSCLC patients. Several investigations have indicated that patients with high TMB or high PD-L1 expression were associated with poor survival in resected NSCLC patients and might benefit from adjuvant chemotherapy (12, 13). However, substantial patients with low PD-L1 expression and low TMB still have poor outcomes. Therefore, exploring additional prognostic markers based on TIM could benefit larger population (14).

In our research, we developed novel prognostic early-stage lung cancer immunoscore model by integrating enrichment score of 11 immune gene sets using ssGSEA algorithm. ssGSEA algorithm was based on gene ranks in and out of the selected gene set (15). To date several signatures used for phenotype classification or survival prediction have been developed by leveraging this algorithm (16–18). After immunoscore model construction in the training set, we evaluated its prognostic abilities in training, testing and meta-set. Moreover, we built Immunoscore-clinical prognostic signature (ICPS) by incorporating both immunoscore and clinical factors.

MATERIALS AND METHODS

Clinical Data Processing

We used three largest publically available datasets, TCGA-LUAD, GSE31210, and GSE68465, deposited in Genomic Data Commons (GDC) portal (https://portal.gdc.cancer.gov) or Gene Expression Omnibus (GEO) website (https://www.ncbi.nlm.nih.gov/geo) (19–23). Clinical and pathological information regarding to TCGA-LUAD cohort were retrieved from cBioportal website (https://www.cbioportal.org) with "cdgsr" package (24–26), whereas information related to GSE31210 and GSE68465 were obtained through "GEOquery" package (27). Samples without overall survival (OS) information or with OS time of 0 were excluded. We also ruled out samples with documented neoadjuvant therapies to reduce potential confounding bias. TNM stage were used and transformed to AJCC staging groups. Samples with specific T subcategories (like T2a or T2b) were converted to staging groups according to AJCC 7th edition. T1N0, T2N0, T1N1, T2N1 or T3N0 were converted to stage 1A, 1B, 2A, 2B, respectively, conforming to AJCC 6th edition. For GSE31210 without TNM stage information, we used the pathological stage in the clinical file directly.

RNA-seq and Microarray Data Preprocessing

Raw ".CEL" files of microarray data were downloaded from GEO website and read by "affy" package with the latest brainarray CDF files (October 2019, version 24) (28, 29). Robust multi-array average (RMA) algorithm in "affy" package was then applied to normalize gene expression intensity (28, 30). RMA algorithm included background adjustment, quantile normalization, and measurement summation when multiple probes were used to quantify the same gene expression intensity. After normalization, "arrayQualityMetrics" package was utilized to detect and exclude possible outliers (31). For RNA-seq data, level 3 FPKM data were downloaded using TCGAbiolink *R* package (32). FPKM values were then transformed into TPM values, which allowed a more direct comparison between samples as the sum of all TPMs in each sample were the same. As a result, the inflated statistical significance was reduced (33). TPM values were subsequently log 2 transformed to fit a more normal distribution. Entrez IDs were used across all platforms. Only samples with clinical information were retained. Finally, TCGA-LUAD cohort was used as the training set for immunoscore model construction, which contained 400 patients with RNA-seq data and survival information. Two microarray datasets, GSE31210 ($n = 219$) from Affymetrix Human Genome U133 Plus 2.0 Array platform as testing set 1, and GSE68485 ($n = 356$) from Affymetrix Human Genome U133A Array platform as testing set 2, were used to assess the immunoscore performance in predicting survival of early-stage LUAD patients.

Immunoscore Construction

We searched Immport database (https://immport.niaid.nih.gov) and downloaded 1811 immune-related genes from 17 categories (18). Of 1,811 immune-related genes, 1,361 of them were contained in the training set. Univariate Cox proportional regression analysis was used to investigate their associations with patient survival using "survival" package (34). Only the genes with P-value < 0.05 and hazard ratio (HR) > 1 were screened out as immune risk genes for further study. We then implemented single sample gene set enrichment analysis (ssGSEA) algorithm to quantify the enrichment score of immune risk genes in various immune-related gene set using "GSVA" package (35). Difference of enrichment statistic of genes in the gene set and outside were computed, and normalized to fit a relatively uniform scale

as Barbie et al. described (15, 35). We then transformed the normalized enrichment score into Z-score to conform standard normal distribution using the following algorithm:

$$ZNES_{ij} = \frac{NES_{ij} - M_j}{SD_j}$$

The final Z-score transformed normalized enrichment score of sample i, immune gene set j was denoted by $ZNES_{ij}$. The normalized enrichment score of sample i, immune gene set j was denoted by NES_{ij}. The mean and standard deviation (SD) of enrichment score across all samples in immune gene set j were denoted by M_j and SD_j, respectively. This transformation obtained a uniform underlying distribution (mean = 0, standard deviation = 1) of each gene set across various platform; Immunoscore model was established by integrating all Z-score transformed normalized enrichment score using regularized Cox regression with the ridge penalty.

$$\text{Immunoscore}_\text{i} = \sum\nolimits_{\text{j}=1}^{n} \beta_\text{j} * \text{ZNES}_\text{ij}$$

Immunoscore of sample i was denoted by $Immunoscore_i$. Ridge Cox regression coefficient of gene set j was denoted by β_j and standard normal distribution transformed normalized enrichment score of sample i, immune gene set j was denoted by $SNES_{ij}$. Ridge regression was used to address the possible collinearity (i.e., the correlated immune gene sets) to prevent overfitting (36). It was conducted by "glmnet" package and the tuning parameter Lambda was chosen with minimum criteria (37). Thus, a new variable immunoscore was created to predict patient survival. It could also be served as the quantitative measurement of hazardous level of tumor immune microenvironment with its biological background.

Validation of Immunoscore

After immunoscore development, we applied the same formula to two independent testing sets, GSE31210 and GSE68485. Meta-set was formed by combining all training and testing sets. Univariate and Multivariate regression were used to evaluate the predictive power of the immunoscore model in all training, testing and meta-set. Age, stage, gender and smoking history were included in multivariable Cox analysis. Fraction of genome alteration in TCGA-LUAD clinical profile was also involved as a covariate in the TCGA-LUAD cohort. Patients were divided into high-immunoscore and low-immunoscore subgroups based on median value in the training set. Patients with immunoscore higher than cut-off value were assigned to high-immunoscore subgroup, while with immunoscore lower or equal to cut-off value were assigned to low-immunoscore subgroup. Kaplan-Meier analysis was performed to these two groups. Time-dependent receiver operator characteristic (ROC) curve analysis was utilized to assess the predictive accuracy for early-stage LUAD patients using "timeROC" package (38). The prognostic value of immunoscore in various treatment groups was evaluated in GSE68465, which contained detailed information of whether patients received adjuvant chemotherapy or radiotherapy with sufficient sample size in each category (75 patients with documented adjuvant therapy, 271 patients without documented adjucvant therapy).

Comparison With Other Gene Expression Signatures

The immunoscore was compared with other existing NSCLC prognostic signature to assess its clinical utility. To date, numerous gene expression signatures have been developed. We selected two immune-related signatures (39, 40), one glycolysis-based signature. In addition, another malignancy gene signature was included, which had the top-ranked prognostic capability when compared with random signature in lung adenocarcinoma patient (41, 42). Detailed information regarding these signatures was provided in **Supplementary Table 3**. Gene symbols in the signatures were transformed into Entrez IDs. Using coefficients provided in supplementary materials, continuous risk score of each signature was computed in TCGA-LUAD, GSE31210, and GSE68465 cohorts. For malignancy gene signature, risk score was generated in each set by first principal component of provided gene list. Hazard ratios of univariate and multivariate Cox regression were used to evaluate their survival associations. C-index derived from "coxph" function with default Efron method to handle ties was utilized to determine their prognostic classification capabilities.

Immunoscore-Clinical Prognostic Signature Construction

To make full use of both immunoscore and clinical variables in prognostic prediction, we constructed immunoscore-clinical prognostic signature (ICPS). Stage was converted to numeric variable. Stage IA, IB, IIA, IIB were assigned as 1, 2, 3, 4, respectively. Stage II with no subcategories were assigned as 3.5. Similarly, median value of ICPS in the training set was used as the cut-off value. C-index of ICPS was compared with stage or immunoscore alone using "compareC" package (43).

Genomic Analysis

Somatic mutation profile were downloaded from Genomic Data Common (GDC) website. Maftools was used to summarize the somatic mutation (44). Samples measured by Whole Genome Amplification (WGA) of Repli-G DNA (which could be identified by tumor barcode) were excluded to reduce possible bias. Tumor mutation burden (TMB) was calculated as previous study described:

$$\text{TMB}_\text{i} = 1.0 * \text{NTM}_\text{i} + 1.5 * \text{TM}_\text{i}$$

Tumor mutation burden of sample i was denoted by TMB_i. Total number of nontruncating mutation and total number of truncating mutation were denoted by NTMi and TMi, respectively.

The silent mutation was not included in the formula as it does not result in any change downstream. The truncating mutation was assigned a higher weight as it is considered more detrimental (45). Mutated genes between high-immunoscore and low-immunoscore groups were compared by fisher exact test using "mafcompare" function (44). Gene ontology and pathway

analyses were then performed using differentially mutated genes by "clusterProfiler" package (46).

Tumor Purity and Various Cell Composition Characterization

We established our immunoscore model based on the bulk gene expression data of the tumor. It could also be used as the measurement of hazardous level of tumor environment (TME) with its biological background. TME contained not only cancer cells, but surrounding non-cancerous immune and normal cells. To further delineate the correlation between immunoscore and TME, we need to first figure out the TME components. TCGA-LUAD cohort was used for TME evaluation. Tumor purity, the percentage of cancer cell inside the tumor, could be estimated in different ways. Aran et al. developed consensus measurement of purity estimations (CPE), which used the median value of several genomic algorithms and immunohistochemistry (IHC) after normalization by combined mean and standard deviation (47). As a result, the bias introduced by a single method or algorithm was minimized. We also utilized Estimating the Proportion of Immune and Cancer cells (EPIC) algorithm, a method to predict various cell types in tumor tissue using RNA-seq tumor gene expression profile (48). Non-log transformed TPM data of TCGA-LUAD samples were used and the Ensemble gene IDs were converted into the official gene symbols as the algorithm required. The EPIC algorithm was based on reference gene expression profiles to infer proportions of surrounding non-malignant cells with experimental measurements confirming its predictive robustness. Samples with convergence code other than 0 were excluded as these might be outliers.

Gene Set Enrichment Analysis and Association With Clinical or Molecular Variables

Gene set enrichment analysis was performed to assess the association of immunoscore to the functional immune pathways. Differential gene expression profile between high-immunoscore and low-immunoscore subgroups was derived by "eBayes" function using limma package (49). We run fgsea algorithm with top 12,000 genes using C5 gene set from MsigDB database (https://www.gsea-msigdb.org/gsea/msigdb/). Gene set related to immune system were extracted. The correlation of immunoscore with clinical factors and certain molecular markers were also evaluated.

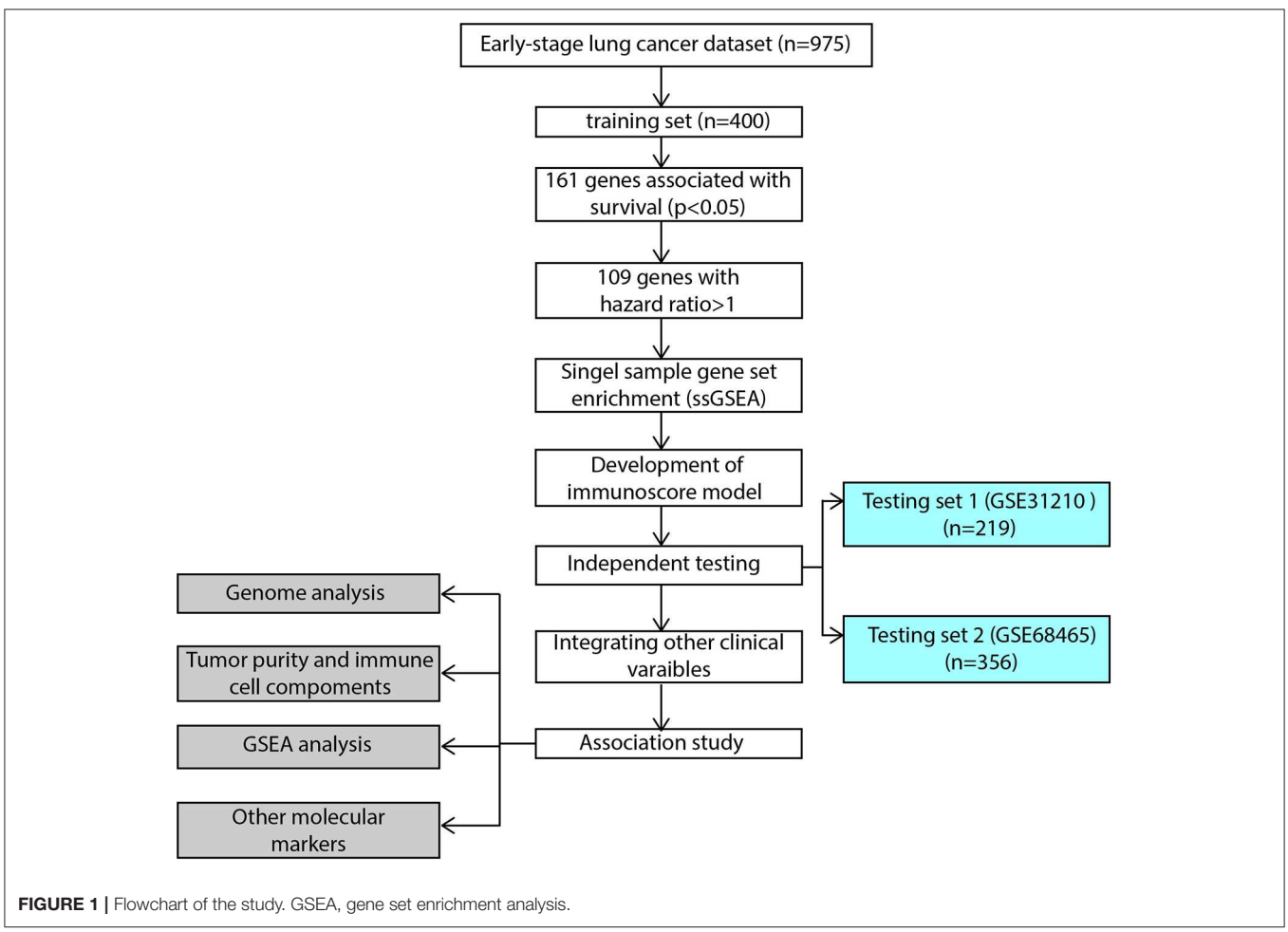


FIGURE 1 | Flowchart of the study. GSEA, gene set enrichment analysis.

Statistical Analysis

Group comparison between a continuous variable were conducted by *t*-test or ANOVA. All correlation analyses were performed with spearman method, and considered highly correlated when absolute value of correlation coefficient was >0.25. False discovery rate was calculated as the adjusted *P*-value. All statistical procedures were conducted by *R* software version 3.6.1 (50). All *p*-values were two-sided and considered statistically significant when <0.05. Gene set was with $P < 0.05$ and FDR < 0.25 was considered significantly enriched.

RESULT

Immunoscore Model Construction

The flowchart of our study procedures was illustrated in **Figure 1**. A total of 975 patients with early-stage lung adenocarcinoma were included in the study. Detailed clinical information was shown in **Table 1**. In the training set, 109 genes were correlated with worse prognosis (HR > 1, $P < 0.05$, **Supplementary Table 1**). Gene set "TGFb family members," "TGFb_Family_Member" and "Interferons" contained only 1 gene and were excluded from further analysis. Z-score transformed enrichment scores of the remaining 11 gene set were then calculated as the method described. All of them were correlated with poor survival in the training set (**Figure 2A**; **Supplementary Table 2**). Ridge Cox regression was then performed and immunoscore was derived by the sum of all Z-score transformed enrichment scores weighed by ridge regression coefficients (**Figures 2B,C**; **Supplementary Table 2**). The predictive accuracy of immunoscore to 2, 3, and 5-year survival were estimated by time-dependent receiver ROC analysis (**Figure 2D**).

TABLE 1 | Detailed patient clinical characteristics.

Characteristics	Training set	Testing set 1	Testing set 2
Source	TCGA	GSE31210	GSE68465
Sample size	400	219	356
Platform	RNA-seq	Affymetrix Human Genome U133 Plus 2.0 Array	Affymetrix Human Genome U133A Array
AJCC stage			
IA	130 (32.5)	112 (51.1)	112 (31.5)
IB	143 (35.8)	53 (24.2)	155 (43.5)
II	—	54 (24.7)	—
IIA	57 (14.2)	—	24 (6.7)
IIB	70 (17.5)	—	65 (18.3)
Age group			
≤65	190 (47.5)	170 (77.6)	186 (52.2)
> 65	201 (50.2)	49 (22.4)	170 (47.8)
Smoking history			
Non-smoker	57 (14.2)	112 (51.1)	40 (11.2)
Ever-smoker	333 (83.2)	107 (48.9)	243 (68.3)
Unknown	10 (2.5)	—	73 (20.5)
Gender			
Male	177 (45.6)	103 (46.4)	175 (49.2)
Female	211 (54.4)	119 (53.6)	181 (50.8)
Survival status			
Alive	273 (68.2)	186 (84.9)	189 (53.1)
Dead	127 (31.8)	33 (15.1)	167 (46.9)
Genome alteration			
≤ 0.2	170 (42.5)	—	—
> 0.2	229 (57.2)	—	—
Unknown	1 (0.2)	—	—

Values in parentheses are percentages.

Validation of Immuoscore

The immunoscore of the testing sets were calculated using the same formula. We also built a meta-set by combing all training and testing sets. Patients were stratified into high and low-immunoscore subgroups using median value of immunoscore in the training set as the cut-off value (−0.0126). Kaplan–Meier survival analysis and log-rank test was performed to compare the difference between these two subgroups. The result exhibited that patients from high-immunoscore subgroup were more likely to suffer worse overall survival ($P < 0.001$ in the training set, testing sets, and meta-set; **Figures 3A–D**). Similarly, patients with higher score were also linked to shorter disease-free survival (DFS) interval ($P < 0.001$ in training set, testing set 1 and meta-set, $P = 0.005$ in testing set 2, **Supplementary Figure 1**). Time-dependent ROC analyses were also performed to testing sets and meta-set (**Supplementary Figure 2**).

Cox regression was used to assess its survival association. Univariate Cox regression analysis revealed that immunoscore was a significant risk factor in all three training and testing sets (HR = 3.11, 95% confidence interval (CI) 2.4–4.04, $P < 0.001$ in training set; HR = 2.39, 95% CI 1.6–3.58, $P < 0.001$ in testing set 1; HR = 1.44, 95% CI 1.17–1.78, $P < 0.001$ in testing set 2; HR = 1.88, 95% CI 1.63–2.17 in meta-set; **Figure 4**). Multivariate Cox regression analysis indicated that immunoscore was an independent risk factor in training (HR = 2.96, 95% CI 2.24–3.9, $P < 0.001$), testing set 1 (HR = 1.99, 95% CI 1.21–3.26, $P = 0.006$), testing set 2 (HR= 1.48, 95% CI 1.13–1.93, $P = 0.005$), and meta-set (HR = 2.01, 95% CI 1.69–2.39, $P < 0.001$), as shown in **Figure 5**. Moreover, Immunoscore could identify patients with worse survival in all clinical subgroups in meta-set (**Supplementary Figure 3**).

Comparison of Immunoscore With Other Genomic Signatures

To assess the utility of immunoscore model, we compared prognostic association of immunoscore against other published genomic signatures (**Supplementary Table 3**). Besides Song et al. signature, most signatures had good performance in univariate and multivariate regression analyses (**Supplementary Figure 4**; **Figure 6A**). Immunoscore exhibited a generally higher C-index than other signatures in all three cohorts, except less than Chen2 et al. signature in GSE31210 (0.72 vs. 0.726, **Figure 6B**). Meanwhile, immunoscore achieved the highest mean C-index (0.68 vs. range from 0.58 to 0.64, **Figure 6B**).

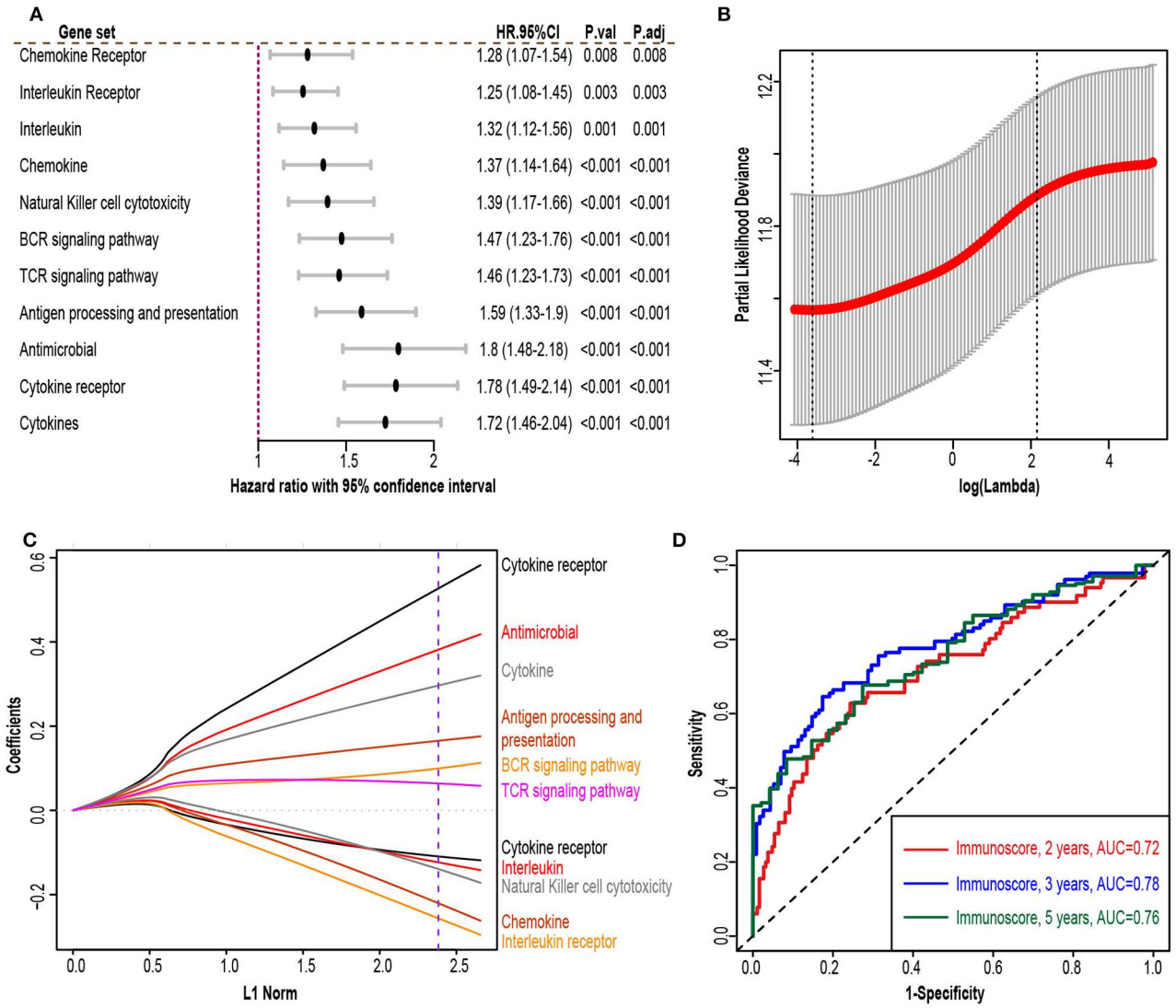


FIGURE 2 | Immunoscore model construction. **(A)** Forest plot exhibiting different immune gene sets and patient overall survival in the training set. **(B)** 10-fold cross-validation for tuning parameter selection in the ridge regression model. The partial likelihood deviance is plotted against log (λ), where λ is the tuning parameter. Partial likelihood deviance values are shown, with error bars representing standard error (SE). The dotted vertical line on the left was drawn by minimum criteria whereas the line on the right represented the 1-SE criteria. We chose the minimum criteria. **(C)** Ridge regression coefficients of the 13 immune gene sets. The dotted line indicated the value chosen by the minimum criteria of the 10-fold validation. **(D)** Time-dependent receiver operator analysis (ROC) of the immunoscore in the training set. HR.95%CI, hazard ratio with 95% confidence interval. P.adj, adjusted *P*-value by false discovery rate.

Immunoscore-Clinical Prognostic Signature Construction

Stage, age and immunoscore were all independent prognostic variables in multivariable Cox analysis in all 3 sets and meta-set. To explore whether combing these variables would improve prediction accuracy, coefficients of multivariate regression of these three factors in the training-set were used to introduce a new variable, immunoscore-clinical prognostic signature (ICPS).

$$ICPS = 1.06076428 * immunoscore + 0.19653598 * stage + 0.01961085 * age$$

Patients were stratified into high-ICPS and low-ICPS subgroups using median value of ICPS in training set as the cut-off (1.74). High-ICPS subgroup was significantly correlated with worse survival in each set ($P < 0.001$, **Figure 7**). **Figure 7** also exhibited C-index of ICPS was significantly higher than either immunoscore or stage, in training [0.72 (ICPS) vs. 0.7 (immunoscore) and 0.59 (stage), $P < 0.001$ when compared with stage], testing set 1 [0.75 (ICPS) vs. 0.72 (immunoscore) and 0.7 (stage), $P = 0.015$ when compared with stage], and testing set 2 [0.65 (ICPS) vs. 0.61 (immunoscore) and 0.62 (stage), $P < 0.001$ when compared with stage]. Moreover, C-index of ICPS was significantly higher than both of them in meta-set [0.7 (ICPS) vs.

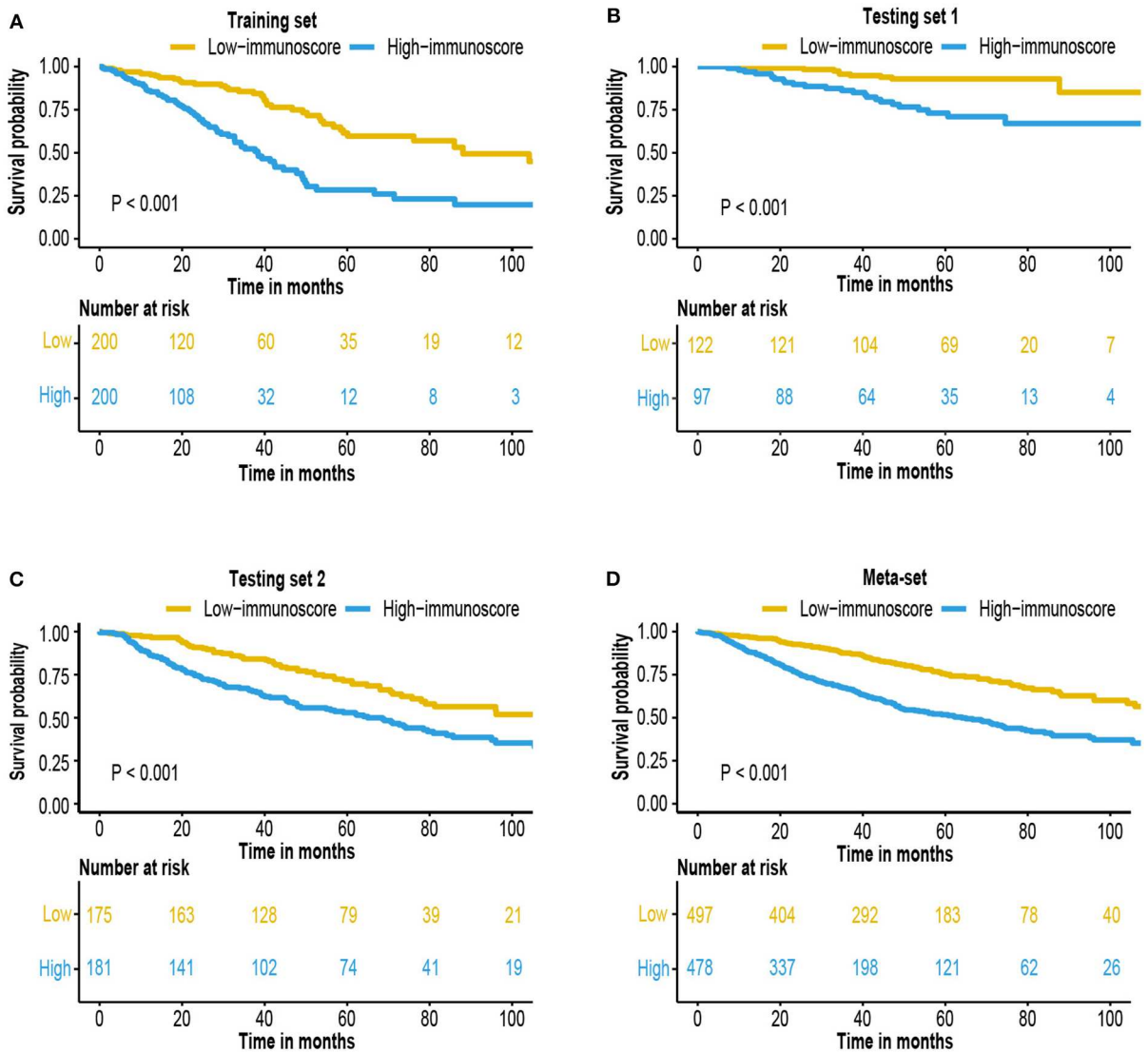


FIGURE 3 | Survival analysis of the immunoscore. Kaplan–Meier curves for patient overall survival by immunoscore group in the **(A)** training set, **(B)** testing set 1, **(C)** testing set 2, and **(D)** meta-set.

0.66 (immunoscore) and 0.64 (stage), $P < 0.001$ when compared with immunoscore or stage, **Figure 7**].

Immunoscore, ICPS, and Adjuvant Therapy

A small subset of early-stage LUAD patient received postoperative adjuvant chemotherapy or radiotherapy. To investigate whether various treatment strategies had an effect to immunoscore and ICPS model, we used GSE68465 cohort with comprehensive documentation of adjuvant therapy. Patients who received adjuvant therapy had a worse overall survival (**Figure 8A**). It might be due to clinical practice, as adjuvant therapy was more likely to be applied to patients with higher stage and worse condition. Survival analysis indicated that immunoscore and ICPS could still stratify patients with different prognosis in each treatment group (**Figures 8B,C**).

Genome Analysis

To explore the possible underlying causes of difference in immunoscore between patients, we searched GDC website and downloaded all available somatic mutation data of lung adenocarcinoma patients. Three hundred fifty-eight available mutation profiles in TCGA-LUAD cohort (174 in high-immunoscore subgroup, 148 in low-immunoscore subgroup) were summarized by maftools. The mutation profiles of high-immunoscore and low-immunoscore subgroups were illustrated in **Figures 9A,B**, respectively. Differentially mutated genes between low-immunocore and high-immunoscore subgroups were identified by Fisher exact test using "mafcompare" function. Twenty of them were shown in **Figure 9C**. TP53 was the most commonly mutated gene in high-immunoscore subgroup and had the smallest adjusted *P*-value. TP53 was a tumor suppressor

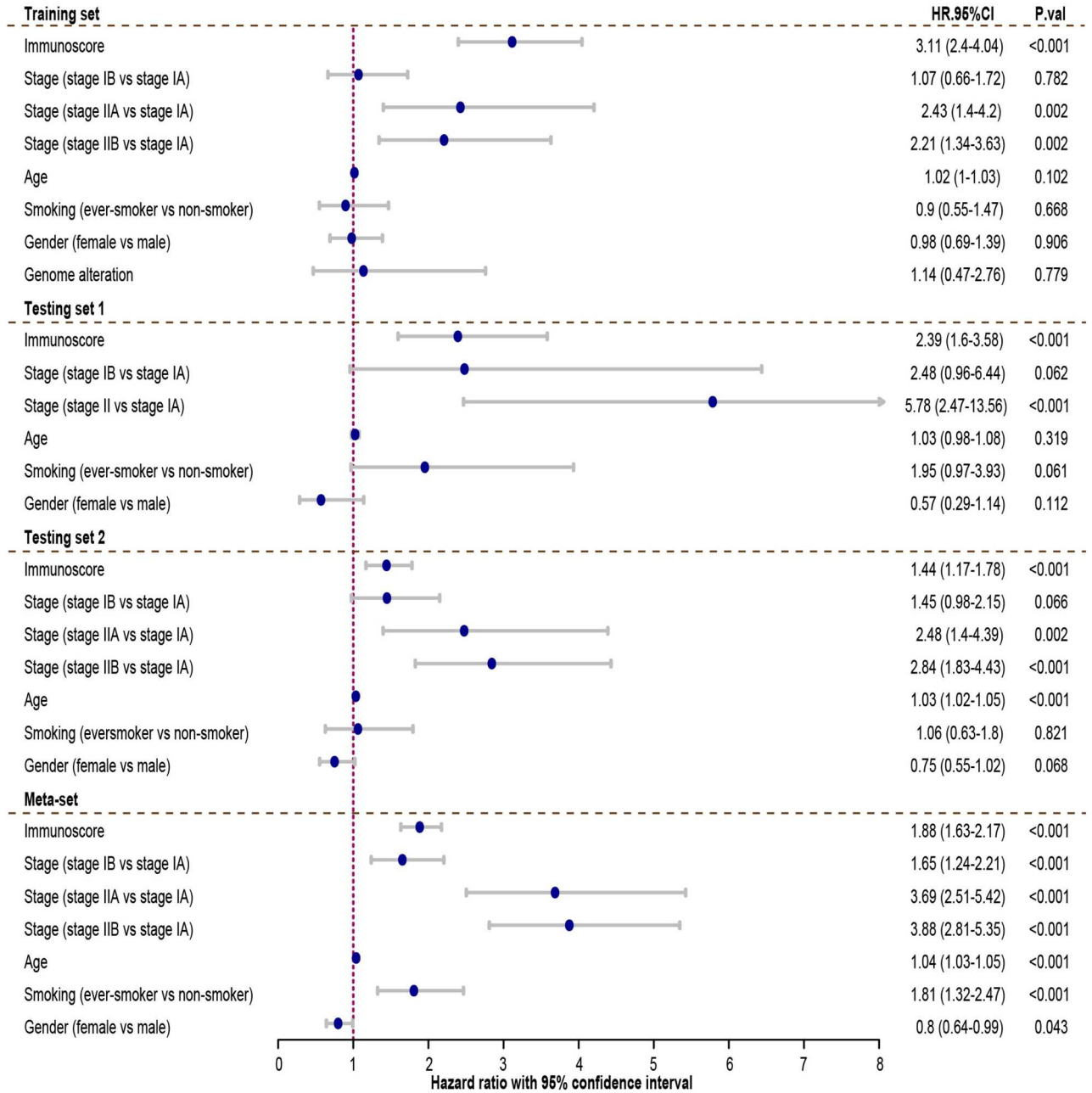


FIGURE 4 | The univariate Cox analysis of the immunoscore and clinicopathological factors. The HR in training cohort was 3.11, with 95% confidence interval (CI) from 2.44 to 4.04 ($P < 0.001$). The HR in testing set 1 was 2.39, with 95% CI from 1.6 to 3.58 ($P < 0.001$). The HR in testing set 2 was 1.44, with 95% CI from 1.17 to 1.78 ($P < 0.001$). The HR in meta-set was 1.88, with 95% CI from 1.63 to 2.17 ($P < 0.001$). HR.95%CI, hazard ratio with 95% confidence interval.

gene, encoding P53 transcriptional factor which responds to DNA damage repair. TP53 mutation has been recently reported to be associated with response to immunotherapy in certain subtype of NSCLC (51). We discovered that TP53 mutation was correlated with immunoscore. P53 mutation might induce genome instability, increasing neoantigen load, leading to a more dangerous tumor immune microenvironment, resulting in higher immunoscore. Another established immunotherapy biomarker, tumor mutation burden (TMB), was also positively correlated with immunoscore ($R = 0.22$, $P < 0.001$, **Figure 9D**). Gene ontology and KEGG pathway analyses of the differentially mutated genes were provided in **Supplementary Figure 5**.

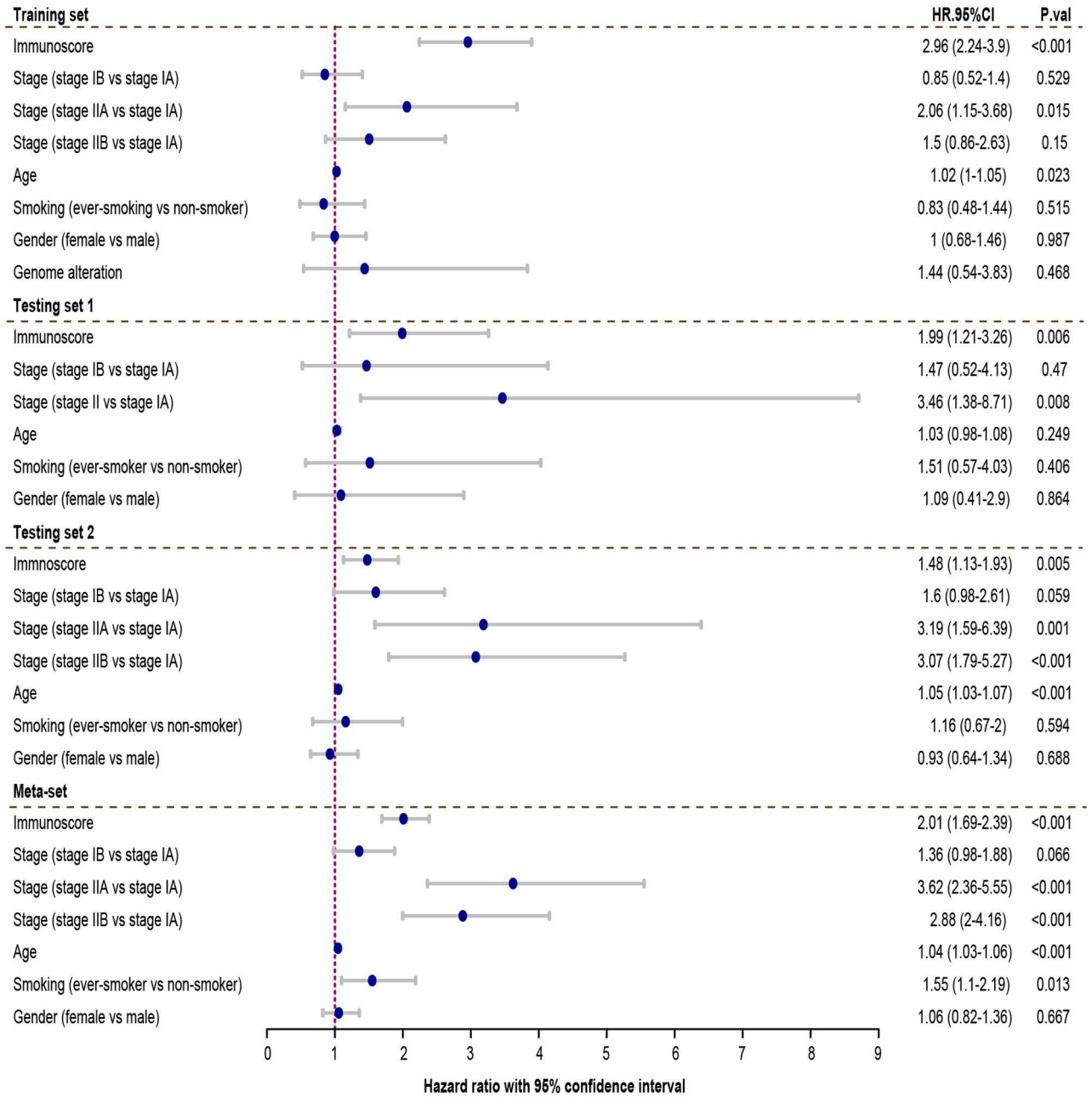


FIGURE 5 | Multivariate Cox analysis evaluating independently predictive ability of immunoscore for patient survival. The immunoscore was able to independently predict patient survival in training set (hazard ratio (HR) = 2.96, 95% confidence interval (CI) from 2.24 to 3.9, $P < 0.001$), testing set 1 (HR = 1.99, 95% CI from 1.05 to 3.81, $P = 0.006$), testing set 2 (HR = 1.48, 95% CI from 1.13 to 1.93, P=0.005), and meta-set (HR = 2.01, 95% CI from 1.69 to 2.39, P <0.001). HR.95%CI, hazard ratio with 95% confidence interval.

Immunoscore and Tumor Microenvironment

The relationship between immunoscore and tumor microenvironment was investigated using TCGA-LUAD cohort. Tumor purity, the percentage of cancer cells inside the tumor, was estimated by consensus measurement of purity estimations (CPE). Patients with high immunoscore tend to have low tumor purity ($R = -0.12$, $P = 0.015$, **Figure 10A**). Patients were divided into high-purity and low-purity subgroup using median value of tumor purity (0.637). Kaplan-Meier survival curves indicated high tumor purity tend to have generally worse survival, but did not reach statistical significance ($P = 0.3$, **Figure 10B**). We next investigated the cellular composition of TME. EPIC algorithm, which was designed specifically for RNA-seq data,

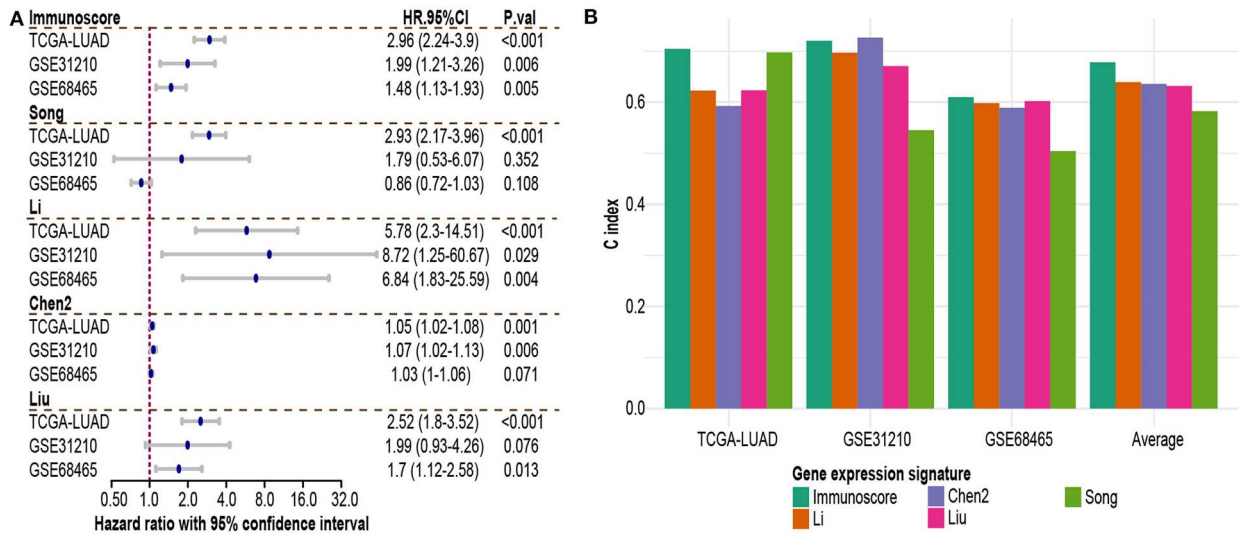


FIGURE 6 | Comparison of immunoscore and other existing NSCLC signatures. **(A)** Hazard ratio of each gene expression signature in multivariable Cox analysis. **(B)** C-index of each signature in each independent dataset and mean C-index. HR.95%CI, hazard ratio with 95% confidence interval.

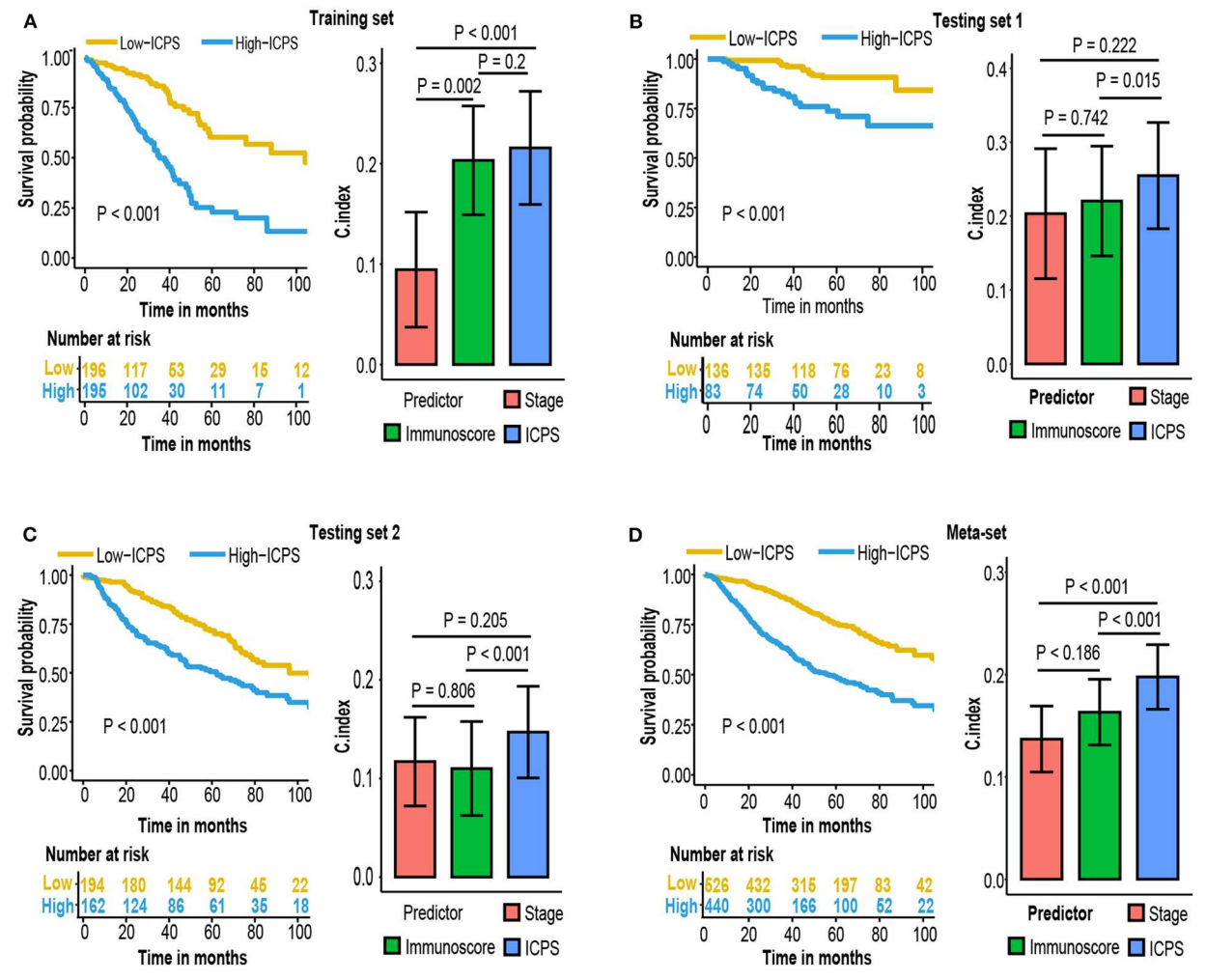


FIGURE 7 | Kaplan–Meier survival analysis and compare C-index of ICPS with immunoscore and stage in **(A)** training set, **(B)** testing set 1, **(C)** testing set 2, **(D)** meta-set.

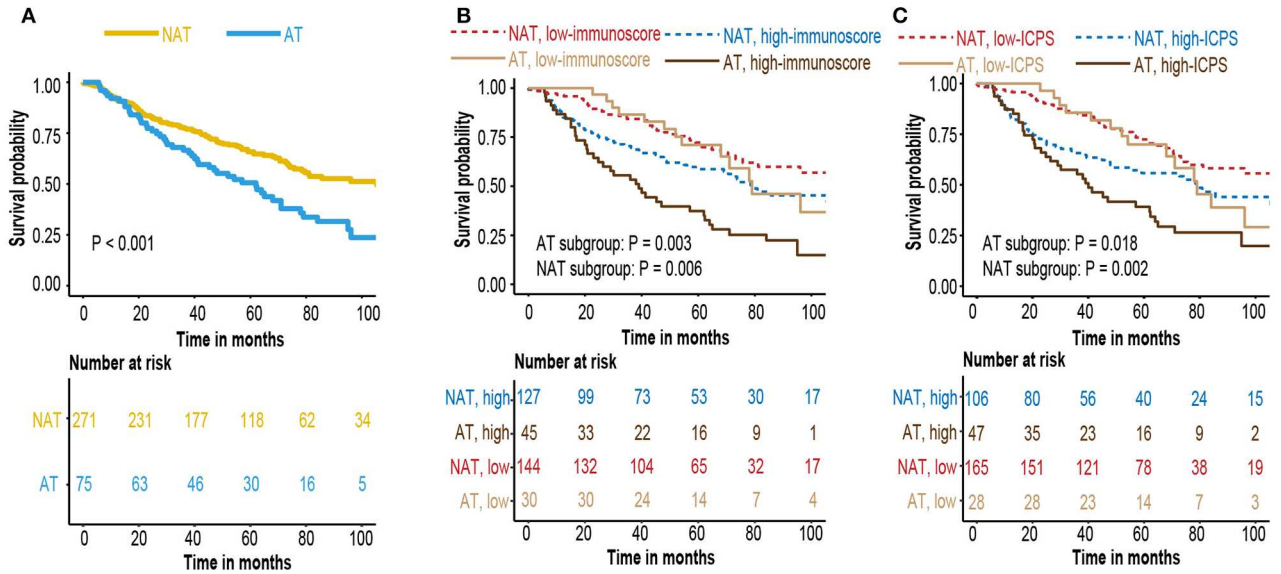


FIGURE 8 | (A) Kaplan–Meier curves for patients who received adjuvant therapy or not **(B)** Kaplan–Meier curves for survival prediction by the immunoscore in patients who received adjunctive therapy or not. **(C)** Kaplan–Meier curves for survival prediction by the ICPS in patients who received adjunctive therapy or not.

was used to infer the proportions of different infiltrating immune and stromal cells. Using absolute value of 0.25 as cut-off, cancer-associated fibroblast (CAF) ($R = 0.32$, $P < 0.001$) and CD8 T cell ($R = -0.27$, $P < 0.001$) were highly correlated to immunoscore (**Figure 10C**; **Supplementary Table 4**). In univariate Cox analysis, only CAF attained a borderline significant P value (HR = 2.42, 95% CI 0.9–6.55, $P = 0.081$, **Figure 10D**).

Immunoscore, Clinicopathological Characteristics, and Biological Phenotypes

Gene set enrichment analysis between high-immunoscore and low-immunoscore were conducted. Immune-related pathways were extracted and most of them were enriched to the low end (27 out of 29 Immune-related pathways). The relationship between immunoscore and other clinicopathological factors were assessed in TCGA-LUAD cohort. Higher T and *N* stage possessed greater immunoscore, whereas its distribution in age, gender and smoking status was not significantly different (**Figure 11B**). Of immune checkpoint molecules, immunoscore was only correlated to PD-L1 and LAG3 ($R = 0.16$, $P = 0.001$ for PD-L1; $R = 0.1$, $P = 0.04$ for LAG3, **Figure 11B**; **Supplementary Table 5**). Interestingly, Several hypoxia-inducible factor (HIF)-1 pathway markers, like HIF-1A ($R = 0.41$, $P < 0.001$), SLC2A1 ($R = 0.6$, $P < 0.001$), LOXL2 ($R = 0.55$, $P < 0.001$), PDK1 ($R = 0.27$, $P < 0.001$), and LDHA ($R = 0.53$, $P < 0.001$), were highly correlated with immunoscore (**Figure 11C**, **Supplementary Table 5**) (52).

DISCUSSION

Lung cancer treatment has been improved dramatically during the past decades, mainly owing to the constant discoveries of genomic alterations during lung cancer pathogenesis. However, the patient prognostic evaluation is still based on the AJCC staging system. Although it is a powerful prognostic prediction tool, it is inadequate to get a precise assessment of patient survival. In early-stage LUAD, the AJCC staging system is far from getting accurate prediction since about 30 percent of patients would develop recurrence, with 2-year survival at about 17% (53). To identify this subset of patients with high risk of recurrence and poor survival is critical since receiving adjuvant chemotherapy or newly developed adjuvant immunotherapy may of great benefit to them.

Up to now, Numerous gene expression signatures have been established for the prediction of lung cancer patient survival (41). Few of them, however, have been translated into real clinical practice. It might be caused by several defects in signature construction. First, some of them were trained from a small cohort with high variance and insufficient independent samples to test its robustness. Second, Gene expression data were measured by different experimental strategy with batch effect, which means that the signature constructed in one specific cohort cannot be generalized into other platforms. Third, most of the signatures were composed of several specific genes and ignore other possible causes, which severely decreased its stability and could potentially lead to overfitting.

In our research, we compared immunoscore with other gene expression signatures. Immunoscore achieved the highest mean C-index, indicating its superior prognostic classification capability. Of immune-related gene signatures, Li et al. (40) signature also had good performance. Li's signature used the binary variable, the pairwise comparison between immune-related gene groups, as features in model construction. Immunoscore and Li's signature had a lot in common, as both of them used some sort of gene ranks (ssGSEA in

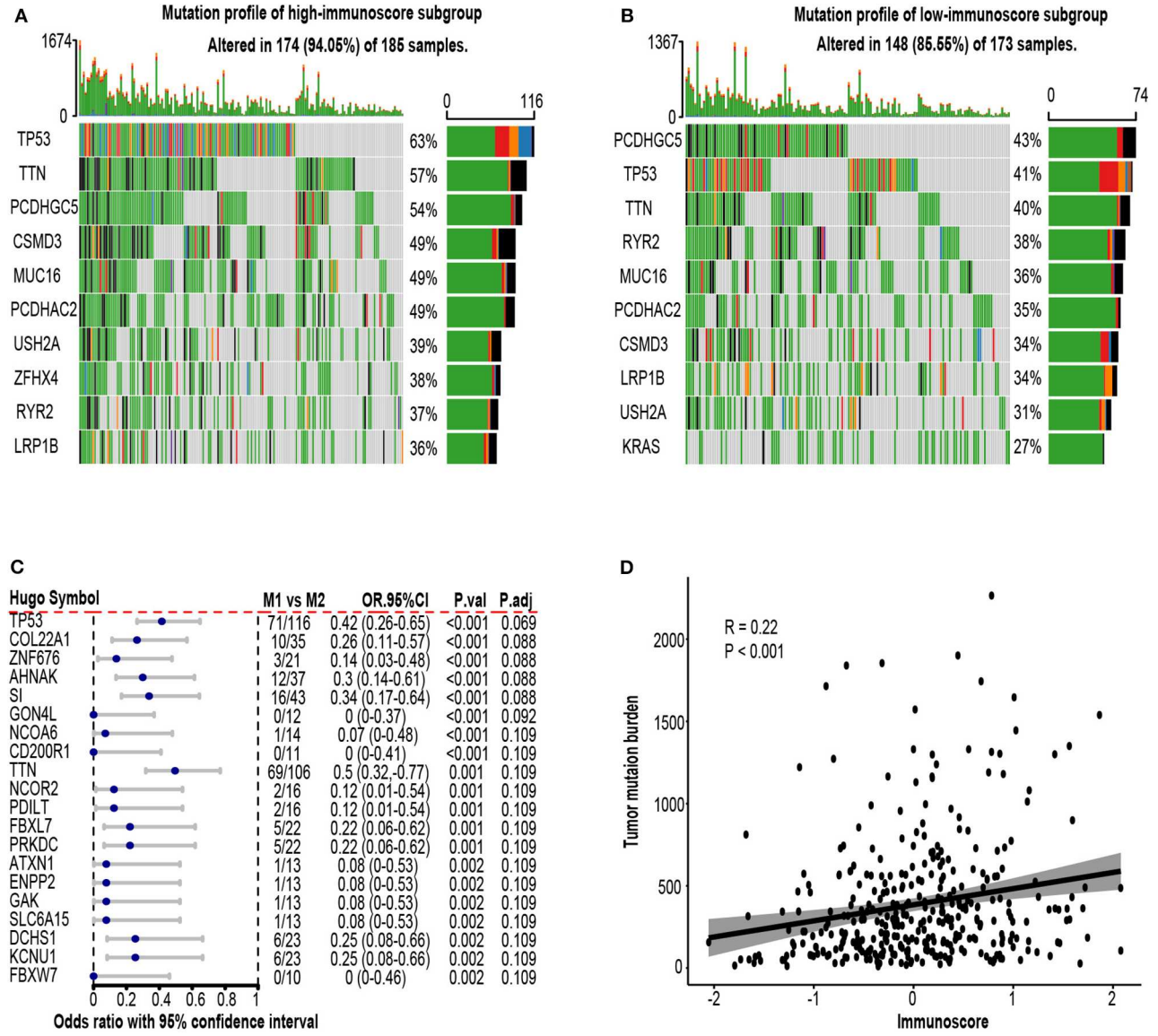


FIGURE 9 | Mutation profile in the TCGA-LUAD cohort. **(A)** Mutation profile of high-immunoscore subgroup. **(B)** Mutation profile of low-immunoscore subgroup. **(C)** Differentially mutated genes between high and low immunoscore patients. **(D)** Correlation between tumor mutation burden (TMB) and immunoscore. OR.95%CI, odds ratio with 95% confidence interval. P.adj, adjusted *P*-value by false discovery rate.

immunoscore; pairwise comparison in Li's signature) rather than gene expression intensity, making them not sensitive to preprocessing strategies and batch effect.

Our model also has its biomedical sense. It was constructed based on enrichment score of risk genes from multiple immune gene sets, and all selected immune gene sets were significantly correlated with worse patient survival. Thus, higher immunoscore indicated a more dangerous tumor microenvironment. The top three contributors to immunoscore were cytokine receptor, antimircrobial, and cytokine. Several cytokine-cytokine receptors signaling pathway have been identified to play a important role in cancer cell proliferation and survival. Most cytokine receptors were located at cell surface, and activated when contacting with specific cytokines. In GSEA analysis, innate immune response activating cell surface receptor signaling pathway ranked the top. Gene ontology analysis also indicated several gene sets related to cell membrane are enriched in differentially mutated genes. Overall, it implied that cell surface signaling pathways were tightly linked to immunoscore and disruption of these pathways might portend poor prognosis. In addition, drugs modifying cytokine-cytokine receptor signaling in combination with other immunotherapy might be a promising treatment strategy. Antimicrobial pathway has been linked to carcinogenesis, as infection by some microorganisms

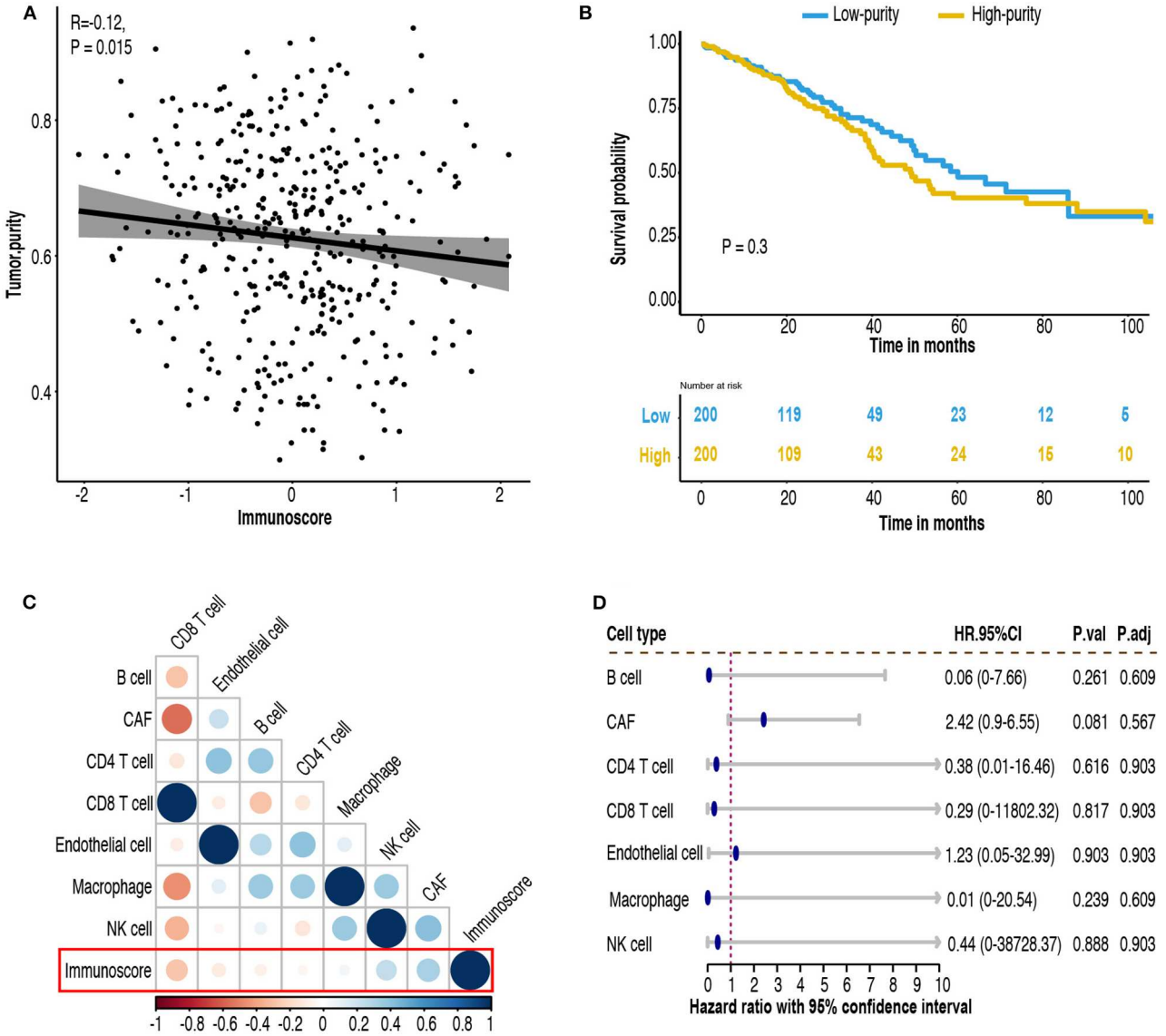


FIGURE 10 | Tumor microenvironment (TME) change associated with immunoscore. **(A)** Correlation between immunoscore and tumor purity. **(B)** Kaplan-Meier curves of patient survival according to tumor purity in the TCGA-LUAD cohort. **(C)** Correlation matrix of immunoscore and cell proportions. **(D)** Univariate Cox analysis of various cell type. HR.95%CI, hazard ratio with 95% confidence interval. P.adj, adjusted *P*-value by false discovery rate.

might lead to cell proliferation, and could be reversed by antimicrobials agents (54).

Tumor purity and cellular composition in tumor microenvironment were also investigated. Patients with high immunoscore tend to have low tumor purity. Furthermore, immunoscore was positively associated with CAFs and but inversely associated with CD8+ T cells. CD8+ T cell has direct cytolytic effect, whereas CAF, on the other hand, may suppress CD8+ cell function by upregulating PD-1, PD-L1, and FAS ligand on Treg cells (55). In addition, KEGG pathway analysis of differentially mutated genes also found ECM-interaction pathway abnormality. ECM stiffness might lead to activation of cancer cells and pro-tumor effect of CAF (56). Besides, cancer cell could induce CAF to remodel ECM, whereas CAF might sustain cancer growth by secreting aspartate (57). Further investigations are needed to figure out how fibroblast communicate with other cells or molecules inside TME and give insight to novel drug targets.

We next explored the phenotypical difference between samples of high and low immunoscore.

Most immune-related pathways were enriched in low-immunoscore subgroup, indicating high-immunoscore subgroup was a "immune cold" subtype. We also discovered multiple markers of HIF-hydroxylase oxygen-sensing pathway to be correlated with immunoscore. HIF could enhance tumor proliferation in TME by altering immune cell function

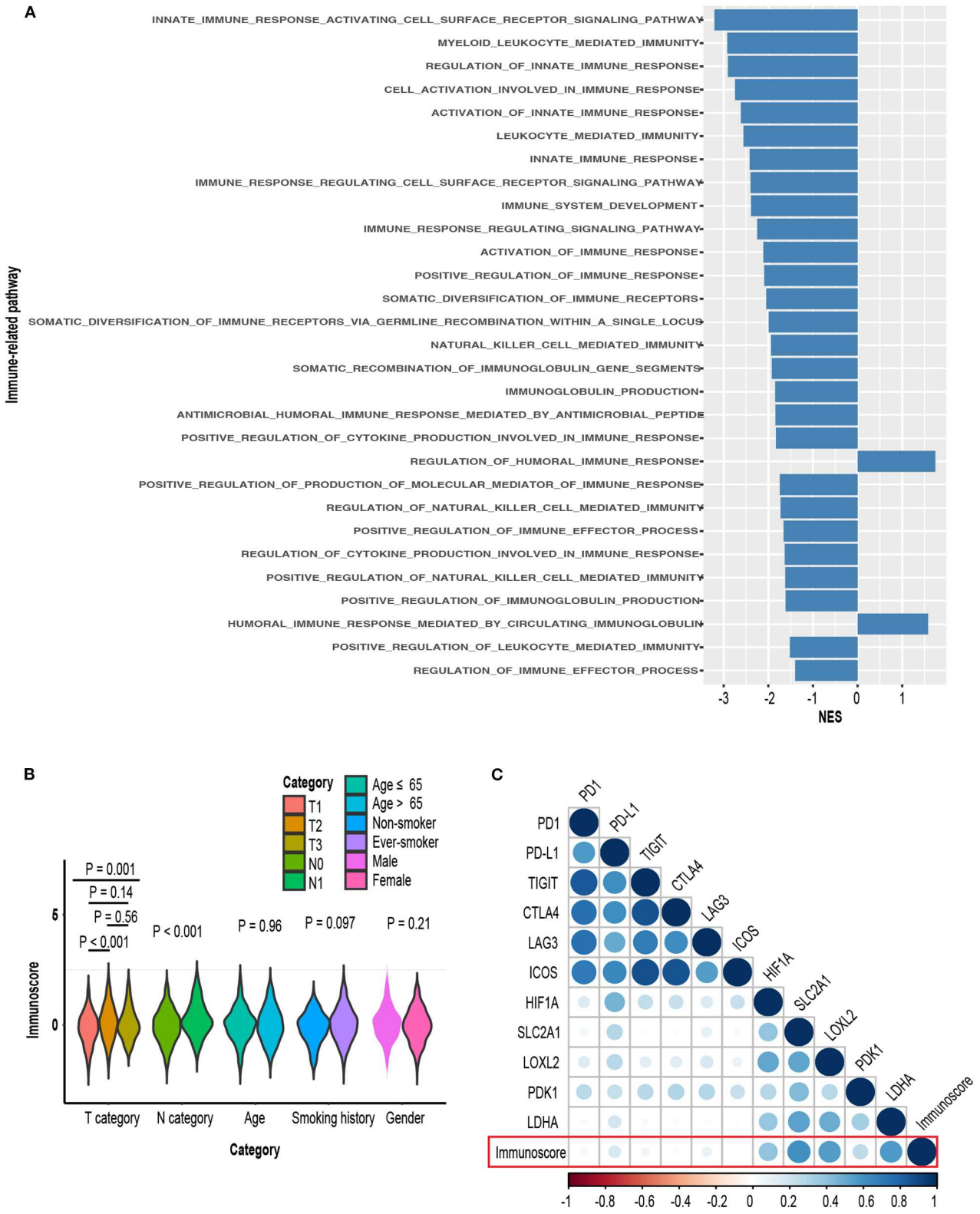


FIGURE 11 | (A) Comparison of enrichment levels of immune-related pathways between high-immunoscore and low-immunoscore subgroups. **(B)** Distribution of immunoscore in various clinical subgroups. **(C)** Correlation matrix of immunoscore and certain gene expression.

and recruiting pro-tumor immune cells (58). For example, expression of HIF1A in tumor-associated macrophage (TAM) might suppress T cell function (59). More experiments and analyses are required to elucidate how HIF pathway affect tumor immune microenvironment as HIF1A is an incredibly promising target for cancer therapy (60).

Our study has several advantages. First, we trained our model in a large cohort with sufficient samples used to test its robustness. Second, we built our immunoscore model to predict patient outcome based on the enrichment levels of different gene sets rather than several single genes, making it a more comprehensive evaluation of tumor immune microenvironment and prevent overfitting. Third, when integrating clinical factors and immunoscore to construct a new ICPS model, it outperformed either immunoscore or stage alone. Fourth, immunoscore itself could also be seen as a proxy variable, the measurement of tumor immune microenvironment, and we found that genome instability, several specific immune cell proportions and functional pathway activation were correlated to immunoscore.

We admit some limitations. First, we used publically available datasets in retrospective manner. We did not have all clinical information needed for the study. For example, patients with inherent immune disorder or taking drugs with impact on immune system should be ruled out. Second, gene expression signatures were developed in different platform with diverse preprocessing strategies and normalization procedure. Although immunoscore outperformed other signatures, it might be due to technical bias or batch effect. Third, the immunoscore model contained several genes with still unknown effects in LUAD, and this "black-box" impact severely undermined the model interpretability. More experiments are needed to elucidate their biological associations. Finally, we cannot estimate its predictive value to immune checkpoint inhibitors due to lack of response data to immunotherapy. Further studies are needed to validate and improve immunoscore model.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://portal.gdc.cancer.gov, https://www.ncbi.nlm.nih.gov/geo.

AUTHOR CONTRIBUTIONS

ZZ, YW, and BW were responsible for the study design. The analysis was performed by ZZ and DZ. ZZ and BW were involved in interpretation of the data. The manuscript was drafted by ZZ. JX and DZ have revised the manuscript. All authors read and approved the final manuscript.

FUNDING

This work was supported by scientific research project of maternal and child health of Jiangsu province (Grant Number F201865).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2020.00691/full#supplementary-material

Supplementary Figure 1 | Survival analysis of the immunoscore. Kaplan-Meier curves for patient disease-free survival by immunoscore group in the **(A)** training set, **(B)** testing set 1, **(C)** testing set 2, and **(D)** meta-set.

Supplementary Figure 2 | Time-dependent receiver operator analysis (ROC) of the immunoscore in the **(A)** testing set 1, **(B)** testing set 2, and **(C)** meta-set.

Supplementary Figure 3 | Subgroup analysis of immunoscore. Immunoscore was a significant risk factor in each clinical subgroup. HR.95%CI, hazard ratio with 95% confidence interval.

Supplementary Figure 4 | Hazard ratio of each gene expression signature in univariable Cox analysis. HR.95%CI, hazard ratio with 95% confidence interval.

Supplementary Figure 5 | Functional analysis of differentially mutated genes in genome analysis. Top 20 **(A)** biological process, **(B)** molecular function, **(C)** cellular component, and **(D)** KEGG pathway. KEGG, Kyoto Encyclopedia of Genes and Genomes.

REFERENCES

1. 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. doi: 10.3322/caac.21492
2. Gridelli C, Rossi A, Carbone DP, Guarize J, Karachaliou N, Mok T, et al. Non-small-cell lung cancer. *Nat Rev Dis Prim.* (2015) 1:15009. doi: 10.1038/nrdp.2015.9
3. Never-smoker NE-s. Comprehensive molecular profiling of lung adenocarcinoma. *Nature.* (2014) 511:543–50. doi: 10.1038/nature13385
4. Baltayiannis N, Chandrinos M, Anagnostopoulos D, Zarogoulidis P, Tsakiridis K, Mpakas A, et al. Lung cancer surgery: an up to date. *J Thorac Dis.* (2013) 5 (Suppl. 4):S425. doi: 10.3978/j.issn.2072-1439.2013.09.17
5. Goldstraw P, Chansky K, Crowley J, Rami-Porta R, Asamura H, Eberhardt WE, et al. The IASLC lung cancer staging project: proposals for revision of the TNM stage groupings in the forthcoming (eighth) edition of the TNM classification for lung cancer. *J Thorac Oncol.* (2016) 11:39–51. doi: 10.1016/j.jtho.2015.09.009
6. Pisters KM, Evans WK, Azzoli CG, Kris MG, Smith CA, Desch CE, et al. Cancer Care Ontario and American Society of Clinical Oncology adjuvant chemotherapy and adjuvant radiation therapy for stages I-IIIA resectable non–small-cell lung cancer guideline. *J Clin Oncol.* (2007) 25:5506–18. doi: 10.1200/JCO.2007.14.1226
7. Group IALCTC. Cisplatin-based adjuvant chemotherapy in patients with completely resected non–small-cell lung cancer. *New Engl J Med.* (2004) 350:351–60. doi: 10.1056/NEJMoa031644
8. Vansteenkiste J, Wauters E, Reymen B, Ackermann CJ, Peters S, De Ruysscher D. Current status of immune checkpoint inhibition in early-stage NSCLC. *Ann Oncol.* (2019) 30:1244–53. doi: 10.1093/annonc/mdz175
9. Gajewski TF, Schreiber H, Fu Y-X. Innate and adaptive immune cells in the tumor microenvironment. *Nat Immunol.* (2013) 14:1014. doi: 10.1038/ni.2703
10. Fridman WH, Zitvogel L, Sautès–Fridman C, Kroemer G. The immune contexture in cancer prognosis and treatment. *Nat Rev Clin Oncol.* (2017) 14:717. doi: 10.1038/nrclinonc.2017.101
11. Brambilla E, Le Teuff G, Marguet S, Lantuejoul S, Dunant A, Graziano S, et al. Prognostic effect of tumor lymphocytic infiltration in resectable non–small-cell lung cancer. *J Clin Oncol.* (2016) 34:1223. doi: 10.1200/JCO.2015.63.0970

12. Devarakonda S, Rotolo F, Tsao M-S, Lanc I, Brambilla E, Masood A, et al. Tumor mutation burden as a biomarker in resected non-small-cell lung cancer. *J Clin Oncol.* (2018) 36:2995–3006. doi: 10.1200/JCO.2018.78.1963
13. Tsao M-S, Le Teuff G, Shepherd F, Landais C, Hainaut P, Filipits M, et al. PD-L1 protein expression assessed by immunohistochemistry is neither prognostic nor predictive of benefit from adjuvant chemotherapy in resected non-small cell lung cancer. *Ann Oncol.* (2017) 28:882–9. doi: 10.1093/annonc/mdx003
14. Altorki NK, Markowitz GJ, Gao D, Port JL, Saxena A, Stiles B, et al. The lung microenvironment: an important regulator of tumour growth and metastasis. *Nat Rev Cancer.* (2019) 19:9–31. doi: 10.1038/s41568-018-0081-9
15. Barbie DA, Tamayo P, Boehm JS, Kim SY, Moody SE, Dunn IF, et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature.* (2009) 462:108–12. doi: 10.1038/nature08460
16. Senbabaoglu Y, Gejman RS, Winer AG, Liu M, Van Allen EM, de Velasco G, et al. Tumor immune microenvironment characterization in clear cell renal cell carcinoma identifies prognostic and immunotherapeutically relevant messenger RNA signatures. *Genome Biol.* (2016) 17:231. doi: 10.1186/s13059-016-1092-z
17. Shen S, Wang G, Zhang R, Zhao Y, Yu H, Wei Y, et al. Development and validation of an immune gene-set based Prognostic signature in ovarian cancer. *EBioMedicine.* (2019) 40:318–26. doi: 10.1016/j.ebiom.2018.12.054
18. Verhaak RG, Tamayo P, Yang J-Y, Hubbard D, Zhang H, Creighton CJ, et al. Prognostically relevant gene signatures of high-grade serous ovarian carcinoma. *J Clin Investig.* (2012) 123:517–25. doi: 10.1172/JCI65833
19. Grossman RL, Heath AP, Ferretti V, Varmus HE, Lowy DR, Kibbe WA, et al. Toward a shared vision for cancer genomic data. *New Engl J Med.* (2016) 375:1109–12. doi: 10.1056/NEJMp1607591
20. Yamauchi M, Yamaguchi R, Nakata A, Kohno T, Nagasaki M, Shimamura T, et al. Epidermal growth factor receptor tyrosine kinase defines critical prognostic genes of stage I lung adenocarcinoma. *PLoS ONE.* (2012) 7:e43923. doi: 10.1371/journal.pone.0043923
21. Okayama H, Kohno T, Ishii Y, Shimada Y, Shiraishi K, Iwakawa R, et al. Identification of genes upregulated in ALK-positive and EGFR/KRAS/ALK-negative lung adenocarcinomas. *Cancer Res.* (2012) 72:100–11. doi: 10.1158/0008-5472.CAN-11-1403
22. Shedden K, Taylor JM, Enkemann SA, Tsao M-S, Yeatman TJ, Gerald WL, et al. Gene expression–based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study. *Nat Med.* (2008) 14:822. doi: 10.1038/nm.1790
23. Barrett T, Troup DB, Wilhite SE, Ledoux P, Rudnev D, Evangelista C, et al. NCBI GEO: archive for high-throughput functional genomic data. *Nucleic Acids Res.* (2009) 37(Suppl_1):D885–90. doi: 10.1093/nar/gkn764
24. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* (2013) 6:pl1. doi: 10.1126/scisignal.2004088
25. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* (2012) 2:401–4. doi: 10.1158/2159-8290.CD-12-0095
26. Jacobsen A, Luna A. *cgdsr: R-Based API for Accessing the MSKCC Cancer Genomics Data Server (CGDS).* R Package Version 1.3.0. (2015). Available online at: https://CRAN.R-project.org/package=cgdsr
27. Davis S, Meltzer PS. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and bioconductor. *Bioinformatics.* (2007) 23:1846–7. doi: 10.1093/bioinformatics/btm254
28. Gautier L, Cope L, Bolstad BM, Irizarry RA. affy—analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics.* (2004) 20:307–15. doi: 10.1093/bioinformatics/btg405
29. Dai M, Wang P, Boyd AD, Kostov G, Athey B, Jones EG, et al. Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. *Nucleic Acids Res.* (2005) 33:e175. doi: 10.1093/nar/gni179
30. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics.* (2003) 4:249–64. doi: 10.1093/biostatistics/4.2.249
31. Kauffmann A, Gentleman R, Huber W. arrayQualityMetrics—a bioconductor package for quality assessment of microarray data. *Bioinformatics.* (2009) 25:415–6. doi: 10.1093/bioinformatics/btn647
32. Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, et al. TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res.* (2015) 44:e71. doi: 10.1093/nar/gkv1507
33. Wagner GP, Kin K, Lynch VJ. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory Biosci.* (2012) 131:281–5. doi: 10.1007/s12064-012-0162-3
34. Therneau T. *A Package for Survival Analysis in S. Version 2.38* (2015). Available online at: https://CRAN.R-project.org/package=survival
35. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics.* (2013) 14:7. doi: 10.1186/1471-2105-14-7
36. Hoerl AE, Kennard RW. Ridge regression: biased estimation for nonorthogonal problems. *Technometrics.* (1970) 12:55–67. doi: 10.1080/00401706.1970.10488634
37. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw.* (2010) 33:1. doi: 10.18637/jss.v033.i01
38. Blanche P, Dartigues JF, Jacqmin-Gadda H. Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Stat Med.* (2013) 32:5381–97. doi: 10.1002/sim.5958
39. Song Q, Shang J, Yang Z, Zhang L, Zhang C, Chen J, et al. Identification of an immune signature predicting prognosis risk of patients in lung adenocarcinoma. *J Transl Med.* (2019) 17:70. doi: 10.1186/s12967-019-1824-4
40. Li B, Cui Y, Diehn M, Li R. Development and validation of an individualized immune prognostic signature in early-stage nonsquamous non–small cell lung cancer. *JAMA Oncol.* (2017) 3:1529–37. doi: 10.1001/jamaoncol.2017.1609
41. Tang H, Wang S, Xiao G, Schiller J, Papadimitrakopoulou V, Minna J, et al. Comprehensive evaluation of published gene expression prognostic signatures for biomarker-based lung cancer clinical studies. *Ann Oncol.* (2017) 28:733–40. doi: 10.1093/annonc/mdw683
42. Chen D-T, Hsu Y-L, Fulp WJ, Coppola D, Haura EB, Yeatman TJ, et al. Prognostic and predictive value of a malignancy-risk gene signature in early-stage non–small cell lung cancer. *J Natl Cancer Inst.* (2011) 103:1859–70. doi: 10.1093/jnci/djr420
43. Kang L, Chen W, Petrick NA, Gallas BD. Comparing two correlated C indices with right-censored survival outcome: a one-shot nonparametric approach. *Stat Med.* (2015) 34:685–703. doi: 10.1002/sim.6370
44. Mayakonda A, Lin D-C, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res.* (2018) 28:1747–56. doi: 10.1101/gr.239244.118
45. Liu Z, Li M, Jiang Z, Wang X. A comprehensive immunologic portrait of triple-negative breast cancer. *Transl Oncol.* (2018) 11:311–29. doi: 10.1016/j.tranon.2018.01.011
46. Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics.* (2012) 16:284–7. doi: 10.1089/omi.2011.0118
47. Aran D, Sirota M, Butte AJ. Systematic pan-cancer analysis of tumour purity. *Nat Commun.* (2015) 6:8971. doi: 10.1038/ncomms9971
48. Racle J, de Jonge K, Baumgaertner P, Speiser DE, Gfeller D. Simultaneous enumeration of cancer and immune cell types from bulk tumor gene expression data. *Elife.* (2017) 6:e26476. doi: 10.7554/eLife.26476.049
49. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* (2015) 43:e47. doi: 10.1093/nar/gkv007
50. R Core Team. *R: A language and environment for statistical computing.* Vienna: R Foundation for Statistical Computing (2019). Available online at: https://www.R-project.org/
51. Chen Y, Chen G, Li J, Huang Y-Y, Li Y, Lin J, et al. Association of tumor protein p53 and ataxia-telangiectasia mutated comutation with response to immune checkpoint inhibitors and mortality in patients with non–small cell lung cancer. *JAMA Netw Open.* (2019) 2:e1911895. doi: 10.1001/jamanetworkopen.2019.11895

52. Luo W, Wang Y. Epigenetic regulators: multifunctional proteins modulating hypoxia-inducible factor-α protein stability and activity. *Cell Mol Life Sci.* (2018) 75:1043–56. doi: 10.1007/s00018-017-2684-9
53. Fedor D, Johnson WR, Singhal S. Local recurrence following lung cancer surgery: incidence, risk factors, and outcomes. *Surg Oncol.* (2013) 22:156–61. doi: 10.1016/j.suronc.2013.04.002
54. Alibek K, Bekmurzayeva A, Mussabekova A, Sultankulov B. Using antimicrobial adjuvant therapy in cancer treatment: a review. *Infect Agents Cancer.* (2012) 7:33. doi: 10.1186/1750-9378-7-33
55. Nazareth MR, Broderick L, Simpson-Abelson MR, Kelleher RJ, Yokota SJ, Bankert RB. Characterization of human lung tumor-associated fibroblasts and their ability to modulate the activation of tumor-associated T cells. *J Immunol.* (2007) 178:5552–62. doi: 10.4049/jimmunol.178.9.5552
56. Pickup MW, Mouw JK, Weaver VM. The extracellular matrix modulates the hallmarks of cancer. *EMBO Rep.* (2014) 15:1243–53. doi: 10.15252/embr.201439246
57. Bertero T, Oldham WM, Grasset EM, Bourget I, Boulter E, Pisano S, et al. Tumor-stroma mechanics coordinate amino acid availability to sustain tumor growth and malignancy. *Cell Metab.* (2019) 29:124–40.e110. doi: 10.1016/j.cmet.2018.09.012
58. Triner D, Shah YM. Hypoxia-inducible factors: a central link between inflammation and cancer. *J Clin Investig.* (2016) 126:3689–98. doi: 10.1172/JCI84430
59. Doedens AL, Stockmann C, Rubinstein MP, Liao D, Zhang N, DeNardo DG, et al. Macrophage expression of hypoxia-inducible factor-1α suppresses T-cell function and promotes tumor progression. *Cancer Res.* (2010) 70:7465–75. doi: 10.1158/0008-5472.CAN-10-1439
60. Soni S, Padwad YS. HIF-1 in cancer therapy: two decade long story of a transcription factor. *Acta Oncol.* (2017) 56:503–15. doi: 10.1080/0284186X.2017.1301680

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

ORIGINAL RESEARCH
published: 13 May 2020
doi: 10.3389/fonc.2020.00743

Preoperative Prediction of Lymph Node Metastasis in Patients With Early-T-Stage Non-small Cell Lung Cancer by Machine Learning Algorithms

Yijun Wu[1,2†], Jianghao Liu[1,2†], Chang Han[1,2], Xinyu Liu[2,3], Yuming Chong[1,2], Zhile Wang[1,2], Liang Gong[1,2], Jiaqi Zhang[1], Xuehan Gao[1], Chao Guo[1], Naixin Liang[1] and Shanqing Li[1*]*

[1] Department of Thoracic Surgery, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China, [2] Peking Union Medical College, Eight-year MD Program, Chinese Academy of Medical Sciences, Beijing, China, [3] Department of Radiology, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

OPEN ACCESS

Edited by:
Umberto Malapelle, University of Naples Federico II, Italy

Reviewed by:
Francesco Pepe, University of Naples Federico II, Italy
Dario De Biase, University of Bologna, Italy

***Correspondence:**
Shanqing Li
lsq6768@163.com
Naixin Liang
pumchnelson@163.com

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Thoracic Oncology, a section of the journal Frontiers in Oncology

Received: *27 November 2019*
Accepted: *20 April 2020*
Published: *13 May 2020*

Citation:
Wu Y, Liu J, Han C, Liu X, Chong Y, Wang Z, Gong L, Zhang J, Gao X, Guo C, Liang N and Li S (2020) Preoperative Prediction of Lymph Node Metastasis in Patients With Early-T-Stage Non-small Cell Lung Cancer by Machine Learning Algorithms. Front. Oncol. 10:743. doi: 10.3389/fonc.2020.00743

Background: Lymph node metastasis (LNM) is difficult to precisely predict before surgery in patients with early-T-stage non-small cell lung cancer (NSCLC). This study aimed to develop machine learning (ML)-based predictive models for LNM.

Methods: Clinical characteristics and imaging features were retrospectively collected from 1,102 NSCLC $\leq$ 2 cm patients. A total of 23 variables were included to develop predictive models for LNM by multiple ML algorithms. The models were evaluated by the receiver operating characteristic (ROC) curve for predictive performance and decision curve analysis (DCA) for clinical values. A feature selection approach was used to identify optimal predictive factors.

Results: The areas under the ROC curve (AUCs) of the 8 models ranged from 0.784 to 0.899. Some ML-based models performed better than models using conventional statistical methods in both ROC curves and decision curves. The random forest classifier (RFC) model with 9 variables introduced was identified as the best predictive model. The feature selection indicated the top five predictors were tumor size, imaging density, carcinoembryonic antigen (CEA), maximal standardized uptake value (SUV_{max}), and age.

Conclusions: By incorporating clinical characteristics and radiographical features, it is feasible to develop ML-based models for the preoperative prediction of LNM in early-T-stage NSCLC, and the RFC model performed best.

Keywords: non-small cell lung cancer, machine learning, lymph node metastasis, predictive model, cross-validation

INTRODUCTION

Lung cancer remains the leading cause of global cancer death (1). Early-T-stage non-small cell lung cancer (NSCLC) has been detected more frequently following the rapid development and employment of radiographical technology (2). An accurate nodal stage is critical for treatment decision-making (3). Currently, there are several evaluation methods, such as computed tomography (CT), positron emission tomography/CT (PET/CT), mediastinoscopy

and endobronchial ultrasound transbronchial needle aspiration (EBUS-TBNA), that can be used to classify the nodal stage before operation. However, performing mediastinoscopy or EBUS-TBNA is not cost-effective for patients with early-stage NSCLC. Furthermore, although CT and PET/CT have been widely used for the preoperative evaluation of lung cancer, the incidence of occult lymph node metastasis (LNM) in early-T-stage NSCLC remains high and cannot be ignored (4, 5). Therefore, new reliable methods for the preoperative prediction of LNM are highly required.

Machine learning (ML) is an emerging computer-based method that has been widely used for data analysis in medicine during the past decade (6, 7). It learns from data and finds the dataset pattern to identify the outcome (7, 8). Supervised ML is a process in which the model is trained with fully labeled and classified data. Compared with conventional statistical methods such as logistic regression (LR), which relies on predetermined models, ML can deeply detect the interactions among variations and iteratively learn from data to update algorithms (9).

A number of predictive models have been made based on ML algorithms. Several studies have reported effective ML-based models for the prediction of LNM in other carcinomas, such as breast cancer (10, 11). It was reported that radiomics could be used to predict LNM by analyzing radiological images in NSCLC (12). However, few reports have incorporated clinical characteristics and radiographical features as in our study. This study aimed to develop and validate effective ML-based models for the prediction of LNM in patients with early-T-stage NSCLC.

MATERIALS AND METHODS

Study Population

Between January 2013 and June 2019, 1,102 patients who underwent surgical resection for NSCLC at Peking Union Medical College Hospital were included in this study. The inclusion criteria were as follows: (1) single NSCLC lesion; (2) tumor maximum diameter $\leq$ 2 cm on CT; and (3) receiving lung resection with systematic lymph node dissection. The exclusion criteria were as follows: (1) small cell lung cancer (SCLC); (2) multiple lung cancer; (3) receiving radiotherapy or chemotherapy before surgery; (4) distant metastasis; and (5) incomplete clinical records. The pathological classification of carcinomas was based on the 2015 World Health Organization (WHO) classification (13). The clinical and pathological staging was performed according to the 8th edition of the TNM staging system (14). This study was approved by the Ethics Committee of Peking Union Medical College Hospital. All patients signed informed consent before operation.

Clinical Characteristics and Radiographical Features

A total of 23 variables were analyzed in this study. The patients' clinical characteristics included age, sex, smoking status and serum tumor biomarkers. All preoperative serum tumor biomarkers were measured within 3 months before surgery, including carbohydrate antigen 24-2 (CA242), squamous cell carcinoma antigen (SCCAg), carcinoembryonic antigen (CEA), carbohydrate antigen 19-9 (CA199), carbohydrate antigen 12-5 (CA125), carbohydrate antigen 72-4 (CA724), carbohydrate antigen 15-3 (CA153), neuron-specific enolase (NSE), tissue polypeptide-specific antigen (TPS), cytokeratin 19-fragments (Cyfra211) and pro-gastrin-releasing peptide (proGRP). CT features were reviewed by one radiologist and two thoracic surgeons independently, including tumor location side, tumor maximum size, spiculation, vessel convergence, lobulation, pleural indentation, calcification, and imaging density. If disagreement occurred, the final result was reached by consensus. Based on imaging density on CT, the cancer lesions were divided into pure ground-glass opacity (pGGO), mixed GGO (mGGO) and solid nodules. The mGGO was further divided into two groups according to different percentages of solid components, whose cut-off value was 50% (the ratio between the maximal diameter of the solid component at the mediastinal window and the maximal tumor diameter at the lung window). In addition, the maximal standardized uptake value (SUV_{max}) on PET/CT was also included. However, PET scan was not routinely performed in early-T-stage NSCLC. All patients underwent CT or PET scan within 60 days at our hospital before the operation.

Construction of ML-Based Models

All patients were randomly divided into training and testing groups at a ratio of 8:2, keeping the distribution of node-positive and node-negative data in both groups consistent. To construct more reliable ML-based predictive models, all continuous variables were preprocessed by z-score normalization except for multinomial naïve Bayes (MNB) in which min-max normalization is preferred (15). Some continuous variables with missing data (**Table S1**), such as SUV_{max} and tumor biomarkers, were imputed by median value (16, 17).

Eight algorithms were applied to predict LNM, including adaptive boosting (AdaBoost), artificial neural network (ANN), decision tree (DT), gradient boosting decision tree (GBDT), logistic regression (LR), MNB, random forest classifier (RFC), and extreme gradient boosting (XGBoost) (18–23). Among all 8 algorithms, LR and MNB are considered conventional methods, and the others are representative supervised ML-based algorithms. Only DT, LR, and MNB were interpretable, in which users were able to recognize function between variable and predictive outcome.

The prediction ability of the 8 models was first evaluated by the receiver operating characteristic (ROC) curve, which is a conventional diagnostic test method that only pays attention to the sensitivity and specificity but ignores the clinical utility of predictive information. Decision curve analysis (DCA) was performed to calculate the clinical values of these models, which is a novel method to assess the information value between diagnostic models by considering the possible range of a patient's risk and benefit preferences without actually measuring these preferences for one particular patient (24).

Validation Strategy and Feature Selection

Overfitting is a common problem in ML, especially with high dimensions (number of variables). To minimize the negative influence of overfitting, some strategies, such as the preselection of variables and cross-validation, were feasible (25,

TABLE 1 | Univariate analysis of patients' clinical characteristics and image features.

	Total	Lymph node status		*P*-value
		pN_+	pN_0	
All patients	1102	116 (10.5)	986 (89.3)	
Age, years	58 [51–65]	59 [53–66]	58 [50–64]	0.382
Sex				
Male	403 (36.6)	52 (44.8)	351 (35.6)	0.051
Female	699 (63.4)	64 (55.2)	635 (64.4)	
Smoking status				
Yes	218 (19.8)	32 (27.6)	186 (18.9)	0.026
No	884 (80.2)	84 (72.4)	800 (81.1)	
Tumor side				
Left	461 (41.8)	49 (42.2)	412 (41.8)	0.925
Right	641 (58.2)	67 (57.8)	574 (58.2)	
Tumor size, cm	1.3 [1.0–1.7]	1.7 [1.5–2.0]	1.2 [1.0–1.6]	<0.001
Imaging density				
pGGO	431 (39.1)	0 (0.0)	431 (43.7)	<0.001
mGGO (solid < 50%)	330 (30.0)	51 (44.0)	279 (28.3)	
mGGO (solid ≥ 50%)	146 (13.2)	27 (23.3)	119 (12.1)	
Solid nodule	195 (17.7)	38 (32.7)	157 (15.9)	
Spiculation				
Yes	587 (53.3)	70 (60.3)	517 (52.4)	0.106
No	515 (46.7)	46 (39.7)	469 (47.6)	
Vessel convergence				
Yes	234 (21.2)	17 (14.7)	217 (22.0)	0.067
No	868 (78.8)	99 (85.3)	769 (78.0)	
Lobulation				
Yes	403 (36.6)	52 (44.8)	351 (35.6)	0.071
No	699 (63.5)	64 (55.2)	635 (64.4)	
Pleural indentation				
Yes	294 (26.7)	43 (37.1)	251 (25.5)	0.007
No	808 (73.3)	73 (62.9)	735 (74.5)	
Calcification				
Yes	21 (1.9)	4 (3.4)	17 (1.8)	0.414
No	1081 (98.1)	112 (96.6)	969 (98.2)	
Tumor SUV_{max}	1.3 [0.7–2.9]	5.9 [3.2–8.7]	1.2 [0.7–2.3]	<0.001
CA242	6.4 [3.4–12.7]	7.5 [4.5–16.5]	6.1 [3.3–12.5]	0.131
SCCAg	0.8 [0.6–1.0]	0.8 [0.6–1.0]	0.8 [0.6–1.0]	0.473
CEA	1.89 [1.20–2.83]	3.63 [2.08–6.69]	1.79 [1.15–2.60]	<0.001
CA199	10 [6.8–16.9]	12.1 [7.4–22.0]	9.9 [6.7–16.8]	0.072
CA125	10.7 [8.0–15.0]	13.3 [9.0–30.1]	10.5 [7.9–14.1]	0.001
CA724	1.9 [1.2–4.3]	2.5 [1.4–5.6]	1.9 [1.2–4.2]	0.128
CA153	9.6 [7.3–13.1]	10.6 [8.0–14.4]	9.5 [7.2–12.9]	0.030
NSE	13.6 [11.6–15.6]	13.5 [11.8–15.8]	13.6 [11.5–15.6]	0.577
TPS	46.68 [29.41–83.10]	54.22 [28.77–110.40]	46.68 [29.30–79.80]	0.492
Cyfra211	1.92 [1.42–2.68]	2.01 [1.63–2.97]	1.90 [1.40–2.62]	0.013
ProGRP	32.1 [26.0–40.5]	33.6 [26.5–45.4]	32.1 [26.0–40.1]	0.115

pGGO, pure ground glass opacity; mGGO, mixed ground glass opacity; Solid < 50%/Solid > 50%: the ratio between the maximal diameter of the solid component at the mediastinal window and the maximal tumor diameter at the lung window < 50%/ > 50% in mGGO; SUV_{max}, maximal standardized uptake value; CA242, carbohydrate antigen 24-2; SCCAg, squamous cell carcinoma antigen; CEA, carcinoembryonic antigen; CA199, carbohydrate antigen 19-9; CA125, carbohydrate antigen 12-5; CA724, carbohydrate antigen 72-4; CA153, carbohydrate antigen 15-3; NSE, neuron-specific enolase; TPS, tissue polypeptide-specific antigen; Cyfra211, cytokeratin 19-fragments; proGRP, pro-gastrin-releasing peptide.

26). Therefore, 5-fold cross-validation and feature selection were performed in this study. The 5-fold cross-validation randomly split the dataset into 5 subsets. For each repeated time, four subsets were used as the training group and the remaining subset was used as the testing data. This procedure was repeated 5 times, and each subset should be used exactly once as the testing group. To rank and select meaningful variables, a classifier-specific evaluator was used, returning a ranked list of variables for each algorithm. The ranks of each variable in different algorithms were compared, and the variables with high ranks were identified.

Statistical Analysis

Univariate analysis was performed using IBM SPSS 25.0 (SPSS Inc; Chicago, IL, USA). Quantitative data were first tested for normality by the Shapiro-Wilk test. Normal data are expressed as the mean ± standard deviation (SD), while non-normal data are expressed as the median with interquartile range (IQR). Student's *t*-test was used to compare normal quantitative parameters, while the Mann-Whitney U test was used to compare non-normal quantitative parameters. For categorical data, Pearson's chi square test or Fisher's exact test was applied. Python programming language (version 3.7, Python Software Foundation) was used for the construction of ML models and DCA. Student's *t*-test was also used for the comparison of different ML models (AUCs). A P-value < 0.05 was considered statistically significant.

RESULTS

Patient Characteristics

All 1,102 patients' clinical characteristics and radiographical features are listed in **Table 1**. Univariate analysis was performed for data without a median value imputed. LNM occurred in 10.5% (116/1102) of patients with NSCLC $\leq$ 2 cm. In total, 699 (63.4%) patients were female, and LNM occurred more frequently in smokers ($P = 0.026$). The maximum tumor size on CT in patients with positive nodes was significantly larger than that in patients with negative nodes ($P < 0.001$). All patients had a maximal diameter no smaller than 4 mm. Tumor imaging density ($P < 0.001$) and pleural indentation ($P = 0.006$) also presented significant differences between node-positive and node-negative patients. None of the patients with positive nodes in this study had a pGGO cancer nodule. Moreover, patients with LNM were significantly different from those without LNM in 4 serum tumor biomarkers: CEA ($P < 0.001$), CA125 ($P = 0.001$), CA153 ($P = 0.030$), and Cyfra211 ($P = 0.013$).

Predictive Performance and Clinical Utility of ML-Based Models

A total of 23 preoperative variables were used to develop predictive models for LNM based on 8 algorithms. The predictive performance of all models is shown in **Figure 1** and **Table 2**. The best performance was observed in the GBDT model (AUC = 0.899, $SD = 0.048$), which performed similarly to RFC (AUC = 0.890, $SD = 0.045$, $P = 0.773$), XGBoost (AUC = 0.883, $SD = 0.047$, $P = 0.627$), AdaBoost (AUC = 0.873, $SD = 0.048$, $P = 0.432$), and ANN (AUC = 0.868, $SD = 0.049$, $P = 0.341$). All

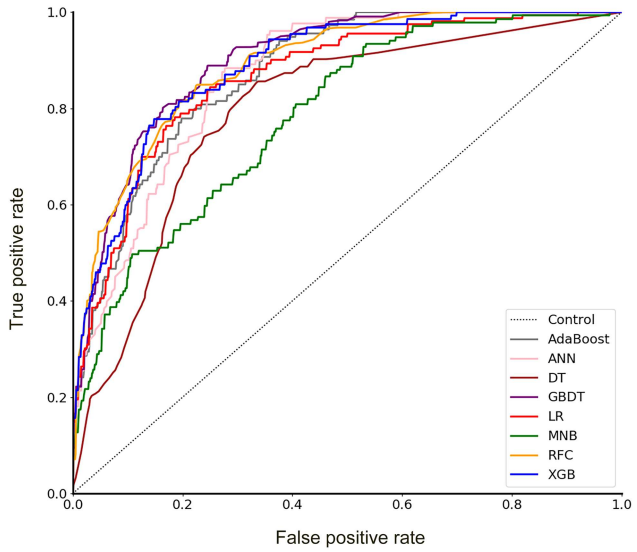


FIGURE 1 | Receiver operating characteristic (ROC) curve for 8 models. AdaBoost, adaptive boosting; ANN, artificial neural network; DT, decision tree; GBDT, gradient boosting decision tree; LR, logistic regression; MNB, multinomial naïve Bayes; RFC, random forest classifier; XGBoost, extreme gradient boosting.

ML-based models except DT (AUC = 0.802, *SD* = 0.057) were better than the two conventional methods, LR (AUC = 0.867, *SD* = 0.049, *P* = 0.338) and MNB (AUC = 0.784, *SD* = 0.058, *P* = 0.002). Moreover, all models performed significantly better than using only tumor size (AUC = 0.753, *SD* = 0.023, *P* < 0.001; the cut-off value was 1.5 cm), SUV_{max} (AUC = 0.734, *SD* = 0.024, *P* < 0.001; the cut-off value was 2.8) or CEA (AUC = 0.720, *SD* = 0.026, *p* < 0.001; the cut-off value was 2.98 ng/ml).

Furthermore, the decision curve showed the clinical values of these models (**Figure 2**). The net benefits of 8 models at each threshold probability are shown in **Table S2**. Most of these models presented better net benefits than two control models that were represented by positive and negative line, respectively. The negative line represents the net benefit is zero when none of patients receive lobectomy with systematic lymph node dissection (SND), assuming that all patients have no positive nodes. On the contrary, the positive line represents the net benefits at the time when all patients have positive nodes and receive lobectomy with SND. Four models (RFC, XGBoost, GBDT, and LR) performed significantly better than the others at most of threshold points. At the range of 0.2–0.5, the LR model was less beneficial than RFC, XGBoost and GBDT on most occasions. The RFC model with 9 variables introduced, which achieved a very high AUC (0.890) and had the highest net benefits almost across the entire range of threshold probabilities, was regarded as the best predictive model in this study, although its AUC value was slightly lower than that of GBDT (*P* = 0.773).

Variable Importance

By feature selection, the 23 variables for each algorithm were ranked by their predictive importance (**Table S3**). The top 10 variables are shown in **Figure 3**. The five top-ranked predictors were tumor size, imaging density, CEA, SUV_{max}, and age. The relationship between the AUCs of models and the number of variables were evaluated in **Figure 4**. The AUCs of most models reached a plateau when 7 variables were introduced, while those of ANN, DT, and MNB started to drop down when they reached the highest points. The AUCs of RFC for each number of variables are shown in **Figure 5**. Its AUC value reached a plateau when 9 variables were introduced and reached the highest value when 13 variables were introduced, but it did not increase significantly with the change from 9 variables (AUC = 0.886) to 13 variables (AUC = 0.890) introduced. Considering the clinical utility, the 9 top-ranked variables were identified to construct the optimal predictive model, which included tumor size, SUV_{max}, imaging density, vessel convergence sign, CEA, CA125, sex, age, and spiculation sign.

TABLE 2 | Predictive performance (AUC) of 8 models and using several variables alone.

Model	AUC			No. of optimal dimensions
	Mean	SD	95% CI	
AdaBoost	0.873	0.048	0.779–0.968	7
ANN	0.868	0.049	0.772–0.964	7
DT	0.802	0.057	0.691–0.913	2
GBDT	0.899	0.044	0.813–0.985	11
LR	0.867	0.049	0.771–0.963	13
MNB	0.784	0.058	0.670–0.898	11
RFC	0.890	0.045	0.801–0.979	13
XGBoost	0.883	0.047	0.792–0.975	7
Tumor size	0.753	0.023	0.707–0.798	1
SUV_{max}	0.734	0.024	0.688–0.780	1
CEA	0.720	0.026	0.669–0.770	1

AUC, area under the receiver operating characteristic curve; AdaBoost, adaptive boosting; ANN, artificial neural network; DT, decision tree; GBDT, gradient boosting decision tree; LR, logistic regression; MNB, multinomial naïve Bayes; RFC, random forest classifier; XGBoost, extreme gradient boosting; SUV_{max}, maximal standardized uptake value; CEA, carcinoembryonic antigen; CA125, carbohydrate antigen 12-5; Cyfra211, cytokeratin 19-fragments; CA153, carbohydrate antigen 15-3.

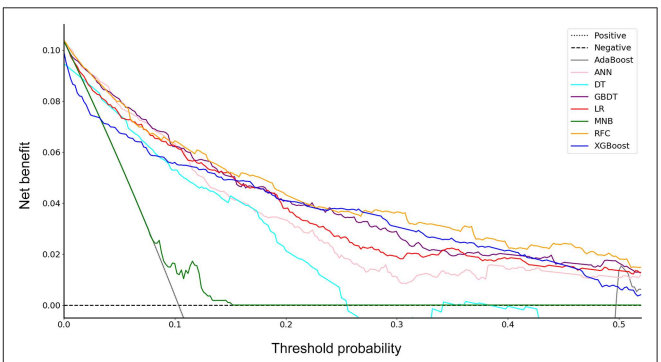


FIGURE 2 | Decision curve for 8 models. AdaBoost, adaptive boosting; ANN, artificial neural network; DT, decision tree; GBDT, gradient boosting decision tree; LR, logistic regression; MNB, multinomial naïve Bayes; RFC, random forest classifier; XGBoost, extreme gradient boosting.

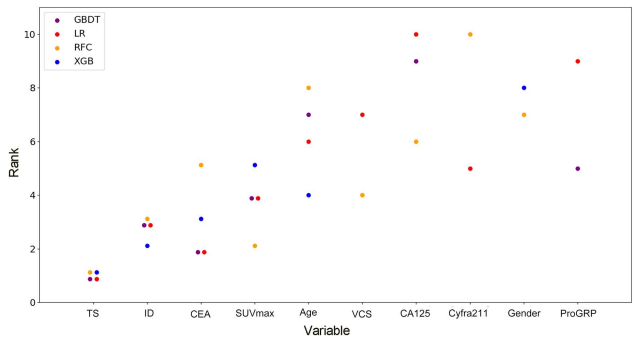


FIGURE 3 | Ranks of the top 10 variables for the prediction of lymph node metastasis. Variables were ranked using a classifier-specific evaluator based on machine learning algorithms. Each variable was ordered according to their mean ranks. The lower rank represents more contributions to the prediction of lymph node metastasis. For example, SUV_{max} was ranked 2nd, 3rd, 3rd, and 5th in RFC, GBDT, LR, and XGB, respectively. TS, tumor size; ID, imaging density; CEA, carcinoembryonic antigen; SUVmax, maximal standardized uptake value; VCS, vessel convergence sign on CT; CA125, carbohydrate antigen 12-5; Cyfra211, cytokeratin 19-fragments; proGRP, pro-gastrin-releasing peptide.

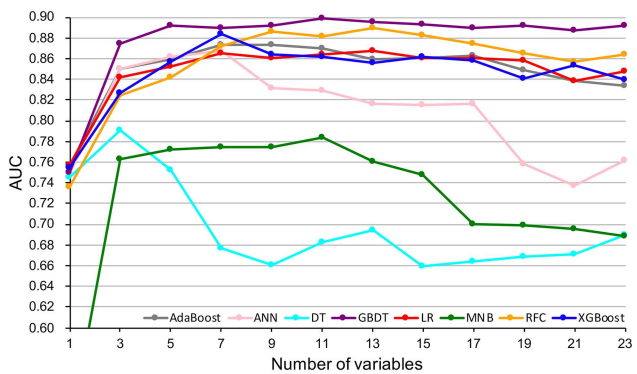


FIGURE 4 | Predictive performance (AUCs) of 8 models as number of variables increases. AdaBoost, adaptive boosting; ANN, artificial neural network; DT, decision tree; GBDT, gradient boosting decision tree; LR, logistic regression; MNB, multinomial naïve Bayes; RFC, random forest classifier; XGBoost, extreme gradient boosting.

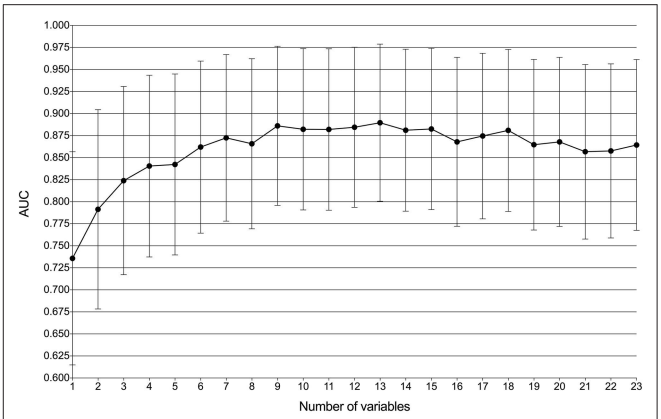


FIGURE 5 | Predictive performance (AUCs) of the random forest classifier (RFC) model at each number of variables.

DISCUSSION

Lobectomy with systematic lymph node dissection remains the standard treatment for patients with early-T-stage NSCLC ($\leq$ 2 cm) (27). However, sublobar resection, including segmentectomy and wedge resection, has been proposed to achieve more precise intervention with the advancement of imaging techniques in recent years. In addition, the reasonable extent of lymph node dissection remains controversial. An exact nodal status is critical for treatment selection and prognosis.

In this study, using ML algorithms, we developed 8 models to predict LNM in 1,102 patients with NSCLC $\leq$ 2 cm, incorporating their clinical characteristics and radiographical features. ROC analysis and DCA were used to evaluate the predictive performance and clinical values of the models, respectively. Most of 8 models maintained high AUCs and All ML-based models (with AUCs ranging from 0.868 to 0.899) except DT performed better than two models using conventional statistical methods (LR and MNB) in the prediction of LNM (**Figure 1** and **Table 2**).

DCA has been used for many medical studies and has shown great clinical utility (28, 29). In the decision curve, most of these models performed better than positive line and negative line, indicating that the overall net benefit of giving lobectomy with SND to patients identified by the models to have high risk of LNM was higher than that of giving the same surgical procedures to all patients or no patient. Four models (RFC, XGBoost, GBDT, and LR) performed better than the others at most of threshold points (**Figure 2**). Thus, these four potential models were used to identify variable importance by feature selection (**Figure 3**). The other four models, AdaBoost, MNB, DT, and ANN, had lower net benefits in the decision curve (**Figure 2**), although they possessed high AUCs in the ROC curve. This indicated that models with high predictive accuracy might not be clinically practical and require further evaluation by other methods, such as DCA.

Using conventional univariate analysis, previous studies reported the risk factors associated with LNM in NSCLC $\leq$ 2 cm, including tumor size, serum CEA and imaging density (30, 31). In addition, SUV_{max} was also thought to be a risk factor in patients with cT1 NSCLC (32). Thus, the AUCs when using tumor size (AUC = 0.753), SUV_{max} (AUC = 0.734), or CEA (AUC = 0.720) alone were also calculated, which were significantly lower than those of ML-based models (**Table 2**). Thus, previous studies might not provide precise predictive information for LNM. Reliable predictive models for LNM in patients with NSCLC are needed. To our knowledge, our study was the first to provide potential models for the prediction of LNM in patients with NSCLC by incorporating clinical characteristics and radiographical features.

Although most of the ML-based models in our study cannot demonstrate the connection between the predictive variables and the outcomes, the contribution of each variable to the models could be inferred by feature selection. Tumor size, imaging density, serum CEA, SUV_{max}, and age were indicated to be the most contributive risk factors of LNM (**Figure 3**), which was similar to the results of univariate analysis (**Table 1**). Since none of the patients with pGGO NSCLC had positive nodes in our and previous studies (30, 31), it could be inferred that pGGO might be predictive of node-negative status in early-T-stage NSCLC. It was also reported that a higher serum CEA level was significantly associated with a higher incidence of LNM (31, 33). Although only 611 patients' SUV_{max} values (pN+: n = 62, pN0: n = 549; $p > 0.05$) were available because some patients did not undergo PET scans, SUV_{max} was ranked at 4 among the four potential models (**Figure 3**) and was ranked at 2 in the RFC model (**Figure 4**). Meanwhile, a high AUC (0.734) for SUV_{max} was also obtained. Above all, SUV_{max} might be one of the most important predictive factors, which was consistent with previous studies (32, 34). Surprisingly, age showed no significance in univariate analysis ($p = 0.382$) but was ranked at the top 5 (**Figure 3**). This might be attributed to the surprising superiority of ML-based models in data mining, which could find more relations between the variables and the outcomes than conventional methods.

According to the ROC curve (**Figure 1**) and decision curve (**Figure 2**), the RFC model with 9 variables introduced (AUC = 0.890) was identified as the optimal model. By considering the clinical utility, an application based on the RFC algorithm with 9 variables (AUC = 0.886) should be developed in the future. These 9 variables were tumor size, SUV_{max}, imaging density, vessel convergence sign, CEA, CA125, sex, age, and spiculation sign. Thus, clinicians from other hospitals could benefit from our study.

In addition to the clinical values, there were several methodological indications in our study. First, although there were several studies of machine learning involving NSCLC, few of them have reported predictive models for LNM using ML algorithms by incorporating clinical characteristics and radiographical features. Most of them performed image analysis by radiographical data (12) or histological slides (35). This is the first study to predict LNM in NSCLC $\leq$ 2 cm, indicating the feasibility and potential of ML algorithms applied in NSCLC. More predictive models of NSCLC may be developed using ML algorithms to solve clinical problems in the future. Second, based on ROC analysis and DCA, multiple supervised ML algorithms performed better than conventional methods. Thus, the ML algorithms would play an important role in the analysis of large medical datasets. Third, in addition to the ROC curve, a decision curve was used to evaluate the clinical utility of these models. Some models performed worse in the decision curve, although they had very high AUCs. This provides a method to further evaluate the clinical values of ML-based models.

There were also some limitations in our study. First, there were some patients who received sublobar resection (wedge resection or segmentectomy), and thus, the incidence of LNM in this population might have been underestimated. Second, missing data were inevitable. This is because not all patients with early-T-stage NSCLC receive PET scans or tumor biomarker tests. Except for SUV_{max} and serum biomarkers, the clinical records of other variables were complete. The median value was imputed to solve this problem (16, 17). Third, this is a retrospective study that could not completely avoid data selection and measurement biases. More prospective studies or multicenter studies may be needed to develop predictive models in the future.

CONCLUSIONS

ML-based models are effective in the prediction of LNM in NSCLC $\leq$ 2 cm by incorporating clinical and radiographical characteristics. Based on ROC analysis and DCA, some ML-based models performed better than models using conventional methods, and the RFC model performed best. The feature selection approach identified that tumor size, imaging density, CEA, SUV_{max}, and age were the most important predictive risk factors for LNM.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Committee of Peking Union Medical College Hospital. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

SL, NL, YW, and JL: conceptualization. YW and JL: methodology. YW, JL, CH, XL, and YC: formal analysis. YW, ZW, LG, JZ, XG, and CG: investigation. YW and JL: writing—original draft preparation. YW, SL, and NL: writing—review and editing. SL: supervision.

FUNDING

This research was funded by (1) Foundation for Key Program of Ministry of Education, China (Grant No. 311037); (2) CAMS Innovation Fund for Medical Sciences (CIFMS), (2017-12M-1-009; 2019-I2M-1-001); (3) Beijing Natural Science Foundation (7182132); (4) Special Data Service for Oncology, The National Population and Health Scientific Data Sharing Platform (NCMI-ABD02-201809; NCMI-YF02N-201906), supported by Ministry of Science and

Technology of the People's Republic of China (MOST); 5) CSCO-CSCO Y-2019GENECAST-051.

ACKNOWLEDGMENTS

We would like to give our sincere thanks to Professor Hongsheng Liu, Yushang Cui, Zhijun Han and Zhili Cao for their contributions to the clinical works and we also would like to thank American Journal Experts (www.aje.com) for its linguistic assistance during the preparation of this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2020.00743/full#supplementary-material

REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin.* (2019) 69:7–34. doi: 10.3322/caac.21551
2. Aberle DR, DeMello S, Berg CD, Black WC, Brewer B, Church TR, et al. Results of the two incidence screenings in the National Lung Screening Trial. *N Engl J Med.* (2013) 369:920–31. doi: 10.1056/NEJMoa1208962
3. Krantz SB, Lutfi W, Kuchta K, Wang CH, Kim KW, Howington JA. Improved lymph node staging in early-stage lung cancer in the national cancer database. *Ann Thorac Surg.* (2017) 104:1805–14. doi: 10.1016/j.athoracsur.2017.06.066
4. Smeltzer MP, Faris N, Yu X, Ramirez RA, Ramirez LE, Wang CG, et al. Missed intrapulmonary lymph node metastasis and survival after resection of non-small cell lung cancer. *Ann Thorac Surg.* (2016) 102:448–53. doi: 10.1016/j.athoracsur.2016.03.096
5. Hung JJ, Yeh YC, Jeng WJ, Wu YC, Chou TY, Hsu WH. Factors predicting occult lymph node metastasis in completely resected lung adenocarcinoma of 3 cm or smaller. *Eur J Cardiothorac Surg.* (2016) 50:329–36. doi: 10.1093/ejcts/ezv485
6. Thrall JH, Li X, Li Q, Cruz C, Do S, Dreyer K, et al. Artificial intelligence and machine learning in radiology: opportunities, challenges, pitfalls, and criteria for success. *J Am Coll Radiol.* (2018) 15(3 Pt B):504–8. doi: 10.1016/j.jacr.2017.12.026
7. Deo RC. Machine learning in medicine. *Circulation.* (2015) 132:1920–30. doi: 10.1161/CIRCULATIONAHA.115.001593
8. Shouval R, Bondi O, Mishan H, Shimoni A, Unger RA. Application of machine learning algorithms for clinical predictive modeling: a data-mining approach in SCT. *Bone Marrow Transplant.* (2014) 49:332–7. doi: 10.1038/bmt.2013.146
9. Waljee AK, Higgins PD. Machine learning in medicine: a primer for physicians. *Am J Gastroenterol.* (2010) 105:1224–6. doi: 10.1038/ajg.2010.173
10. Dihge L, Vallon-Christersson J, Hegardt C, Saal LH, Hakkinen J, Larsson C, et al. Prediction of lymph node metastasis in breast cancer by gene expression and clinicopathological models: development and validation within a population-based cohort. *Clin Cancer Res.* (2019) 25:clincanres.0075.2019. doi: 10.1158/1078-0432.CCR-19-0075
11. Ehteshami Bejnordi B, Veta M, Johannes van Diest P, van Ginneken B, Karssemeijer N, Litjens GJ, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA.* (2017) 318:2199–210. doi: 10.1001/jama.2017.14585
12. Zhong Y, Yuan M, Zhang T, Zhang YD, Li H, Yu TF. Radiomics approach to prediction of occult mediastinal lymph node metastasis of lung adenocarcinoma. *AJR Am J Roentgenol.* (2018) 211:109–13. doi: 10.2214/AJR.17.19074
13. Travis WD, Brambilla E, Nicholson AG, Yatabe Y, Austin JHM, Beasley MB, et al. The 2015 World Health Organization classification of lung tumors: impact of genetic, clinical and radiologic advances since the 2004 classification. *J Thorac Oncol.* (2015) 10:1243–60. doi: 10.1097/JTO.0000000000000630
14. Goldstraw P, Chansky K, Crowley J, Rami-Porta R, Asamura H, Eberhardt WE, et al. The IASLC lung cancer staging project: proposals for revision of the TNM stage groupings in the forthcoming (Eighth) edition of the TNM classification for lung cancer. *J Thorac Oncol.* (2016) 11:39–51. doi: 10.1016/j.jtho.2015.09.009
15. Shalabi LA, Shaaban Z, Kasasbeh B. Data mining: a preprocessing engine. *J Comput Sci.* (2006) 2:735–9. doi: 10.3844/jcssp.2006.735.739
16. Wei R, Wang J, Su M, Jia E, Chen S, Chen T, et al. Missing value imputation approach for mass spectrometry-based metabolomics data. *Sci Rep.* (2018) 8:663. doi: 10.1038/s41598-017-19120-0
17. Zhou XH, Eckert GJ, Tierney WM. Multiple imputation in public health research. *Stat Med.* (2001) 20:1541–9. doi: 10.1002/sim.689
18. Ngiam KY, Khor IW. Big data and machine learning algorithms for health-care delivery. *Lancet Oncol.* (2019) 20:e262–73. doi: 10.1016/S1470-2045(19)30149-4
19. Gonzalez GH, Tahsin T, Goodale BC, Greene AC, Greene CS. Recent advances and emerging applications in text and data mining for biomedical discovery. *Brief Bioinform.* (2016) 17:33–42. doi: 10.1093/bib/bbv087
20. Breiman L, Cutler A. Random forests. *Mach Learn.* (2001) 45:5–32. Available online at: http://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm (accessed June 12, 2011).
21. Freund Y, Schapire RE. A short introduction to boosting. *Jinko Chino Gakkaishi.* (1999) 14:771–80. doi: 10.1109/CICC.1996.510579
22. Freund Y, Mason L. The alternating decision tree learning algorithm. *ICML.* (1999) 99:124–33.
23. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. In: *Acm Sigkdd International Conference on Knowledge Discovery & Data Mining.* (2016). doi: 10.1145/2939672.2939785
24. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making.* (2006) 26:565–74. doi: 10.1177/0272989X06295361
25. Jung Y. Multiple predicting K-fold cross-validation for model selection. *J Nonparametr Stat.* (2018) 30:197–215. doi: 10.1080/10485252.2017.1404598
26. Cook JA, Ranstam J. Overfitting. *BJS.* (2016) 103:1804–14. doi: 10.1002/bjs.10244
27. Ginsberg RJ, Rubinstein LV. Randomized trial of lobectomy versus limited resection for T1 N0 non-small cell lung cancer. Lung Cancer Study Group. *Ann Thorac Surg.* (1995) 60:615–22; discussion 622–3. doi: 10.1016/0003-4975(95)00537-U
28. Cadrin-Tourigny J, Bosman LP, Nozza A, Wang W, Tadros R, Bhonsale A, et al. A new prediction model for ventricular arrhythmias in arrhythmogenic right ventricular cardiomyopathy. *Eur Heart J.* (2019) 40:1850–8. doi: 10.1093/eurheartj/ehz103
29. Hijazi Z, Oldgren J, Lindback J, Alexander JH, Connolly SJ, Eikelboom JW, et al. The novel biomarker-based ABC (age, biomarkers, clinical history)-bleeding risk score for patients with atrial fibrillation: a derivation and validation study. *Lancet.* (2016) 387:2302–11. doi: 10.1016/S0140-6736(16)00741-8
30. Pani E, Kennedy G, Zheng X, Ukert B, Jarrar D, Gaughan C, et al. Factors associated with nodal metastasis in 2-centimeter or less non-small cell lung cancer. *J Thorac Cardiovasc Surg.* (2020) 159:1088–96.e1. doi: 10.1016/j.jtcvs.2019.07.089
31. Yu X, Li Y, Shi C, Han B. Risk factors of lymph node metastasis in patients with non-small cell lung cancer $\leq$ 2 cm in size: A monocentric population-based analysis. *Thoracic Cancer.* (2018) 9:3–9. doi: 10.1111/1759-7714.12490
32. Park HK, Jeon K, Koh WJ, Suh GY, Kim H, Kwon OJ, et al. Occult nodal metastasis in patients with non-small cell lung cancer at clinical stage IA by PET/CT. *Respirology.* (2010) 15:1179–84. doi: 10.1111/j.1440-1843.2010.01793.x
33. Song CY, Kimura D, Sakai T, Tsushima T, Fukuda I. Novel approach for predicting occult lymph node metastasis in peripheral

clinical stage I lung adenocarcinoma. *J Thorac Dis.* (2019) 11:1410–20. doi: 10.21037/jtd.2019.03.57

34. Li L, Ren S, Zhang Y, Guan Y, Zhao J, Liu J, et al. Risk factors for predicting the occult nodal metastasis in T1-2N0M0 NSCLC patients staged by PET/CT: potential value in the clinic. *Lung Cancer.* (2013) 81:213–7. doi: 10.1016/j.lungcan.2013.04.012
35. Takamatsu M, Yamamoto N, Kawachi H, Chino A, Saito S, Ueno M, et al. Prediction of early colorectal cancer metastasis by machine learning using digital slide images. *Comput Methods Programs Biomed.* (2019) 178:155–61. doi: 10.1016/j.cmpb.2019.06.022

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

CASE REPORT
published: 31 July 2020
doi: 10.3389/fonc.2020.01259

Durable Complete Response to Alectinib in a Lung Adenocarcinoma Patient With Brain Metastases and Low-Abundance *EML4-ALK* Variant in Liquid Biopsy: A Case Report

Yingying Zhu[1], *Ran Jia*[2], *Yang W. Shao*[3], *Liuqing Zhu*[3], *Qiuxiang Ou*[4], *Man Yu*[4], *Xue Wu*[4] *and Yanbei Zhang*[1]*

[1] Department of Geriatric Respiratory and Critical Care, Institute of Respiratory Diseases, The First Affiliated Hospital of Anhui Medical University, Hefei, China, [2] Department of Hepatobiliary Surgery, The First Affiliated Hospital of Anhui Medical University, Hefei, China, [3] Medical Department, Nanjing Geneseeq Technology Inc., Nanjing, China, [4] Translational Medicine Research Institute, Geneseeq Technology Inc., Toronto, ON, Canada

OPEN ACCESS

Edited by:
Umberto Malapelle, University of Naples Federico II, Italy

Reviewed by:
Danilo Rocco, Azienda Ospedaliera dei Colli, Italy
Xabier Mielgo Rubio, Hospital Universitario Fundación Alcorcón, Spain
Francesco Passiglia, Paolo Giaccone University Hospital in Palermo, Italy

***Correspondence:**
Yanbei Zhang
zhangyanbei1963@126.com

Specialty section:
This article was submitted to Thoracic Oncology, a section of the journal Frontiers in Oncology

Received: *28 November 2019*
Accepted: *18 June 2020*
Published: *31 July 2020*

Citation:
Zhu Y, Jia R, Shao YW, Zhu L, Ou Q, Yu M, Wu X and Zhang Y (2020) Durable Complete Response to Alectinib in a Lung Adenocarcinoma Patient With Brain Metastases and Low-Abundance EML4-ALK Variant in Liquid Biopsy: A Case Report. Front. Oncol. 10:1259. doi: 10.3389/fonc.2020.01259

EML4-ALK fusions are targetable oncogenic drivers in a subset of advanced non-small cell lung cancer (NSCLC) patients that can benefit from selected ALK inhibitors. Precise detection of *ALK* fusions may yield critical information for selection of appropriate therapy and hence improve patient survival. Analysis of circulating tumor DNA (ctDNA) in liquid biopsies using next generation sequencing (NGS) prior to or during treatment hold great promise for disease monitoring and treatment guidance of various cancers including NSCLC. Herein, we report a case of a 21-year-old advanced lung adenocarcinoma patient with a low abundance (0.03%) of *EML4-ALK* rearrangement identified in plasma ctDNA upon progression on two lines of chemotherapy that demonstrated long-term complete response to alectinib (>13 months) including metastatic brain tumors. Patient's clinical and pathologic characteristics, computerized tomography (CT) scans and brain magnetic resonance imaging (MRI) were reviewed retrospectively. Taken together, our report not only reinforces the translational utility of NGS-based genomic sequencing of liquid biopsy in guiding clinical practice, but also highlights the superior efficacy of alectinib than chemotherapy in *ALK*+ NSCLC with brain metastases, albeit at a low variant allele abundance.

Keywords: lung adenocarcinoma, *ALK* rearrangement, alectinib, liquid biopsy, brain metastases

INTRODUCTION

Aberrant *ALK* rearrangements have been recognized as central oncogenic drivers for many solid malignancies. *EML4-ALK* fusions occur in ∼2–7% of advanced non-small cell lung cancer (NSCLC) patients, and are more frequently detected in lung adenocarcinoma as well as in never- or light- smokers or young adults (1). Despite of high overall response rates (ORR) with the first-generation ALK inhibitor crizotinib (2), drug resistance inevitably develops with the central nervous system (CNS) as the most common site of progressive disease in nearly 70% of *ALK*+ patients undergoing crizotinib treatment. Approximately 15–35% of *ALK*+ NSCLC patients manifest with CNS metastases at initial diagnosis (3), dramatically impacting patient prognosis and

quality of life. Previous studies have provided convincing evidence for the superior potency and improved tolerability of alectinib over crizotinib in *ALK*+ patients with baseline brain metastases or leptomeningeal disease (4, 5), which underlines the potential efficacy of alectinib in treating *ALK*-driven NSCLC, particularly in the management of those harboring CNS metastases. Analysis of circulating tumor DNA (ctDNA) in liquid biopsies using next generation sequencing (NGS) provide a non-invasive approach to tumor molecular profiling and is increasingly utilized to screen presence of disease, guide therapy selection, and evaluate treatment response (6). However, adequate assessment of low-abundance ctDNA alterations and their translational significance may be challenging under some circumstances. In this case study, we report a lung adenocarcinoma patient with brain metastases whose disease progressed upon chemotherapy but responded completely to alectinib with detection of a low-abundance *EML4-ALK* fusion in plasma ctDNA.

CASE PRESENTATION

A 21-year-old Chinese male with neither personal smoking history nor family medical history was diagnosed with stage IV lung adenocarcinoma with multiple metastases in cervical, hilar and mediastinal lymph nodes, and pericardial effusion in February 2016 (**Figures 1A,B**). Cervical lymph node biopsy was performed but not the primary lung tumor due to a number of reasons including the tumor size (1.5 cm by 1.5 cm), its proximity to the heart, and a large accumulation of pericardial effusion present at diagnosis. The baseline lymph node biopsy specimen, pericardial effusion ctDNA, and plasma ctDNA samples were immediately subject to comprehensive genomic profiling using next generation sequencing (NGS) by targeting 382 cancer-relevant genes, but no actionable driver mutations were detected in any sample (**Table 1**). FISH or IHC against biomarkers including ALK was not adopted for routine clinical diagnosis at our institution in 2016. The patient soon received the first-line

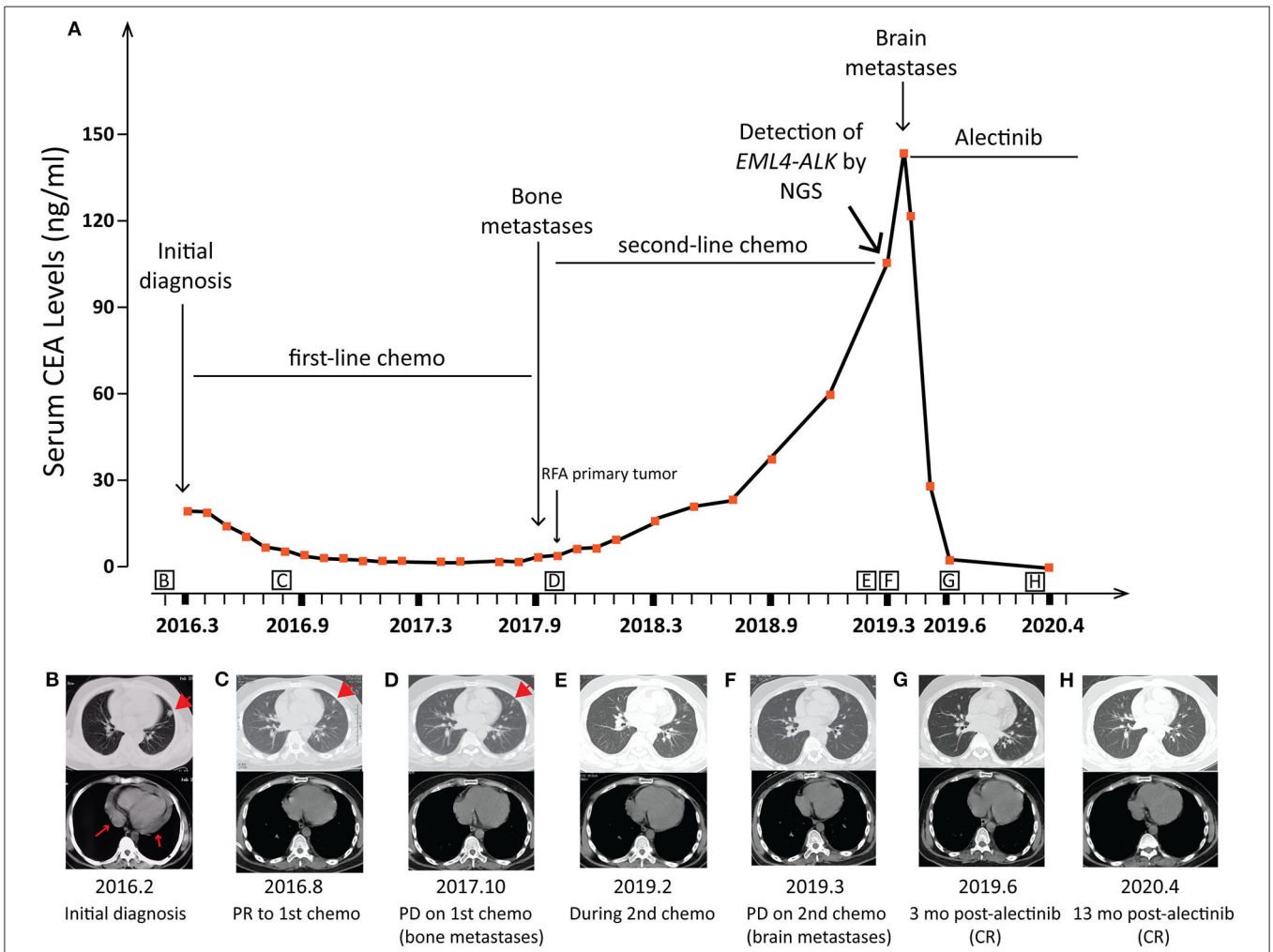


FIGURE 1 | Schematics showing the patient's treatment history. **(A)** A diagram showing the chronological changes of therapeutic regimens and serum CEA levels during the entire course of treatment. RFA, radiofrequency ablation. **(B–H)** The patient's chest computed tomography (CT) scans at different clinical time points as shown. Arrowheads: location of the primary lung lesion in different scans. Arrows: paricardial effusion at initial diagnosis. PR, partial response; PD, progressed disease.

TABLE 1 | Genetic alterations identified by NGS in the patient's FFPE and liquid biopsy samples.

Time points (in Figure 1A)	Sample type	Gene	AA change	MAF%
At diagnosis (2016.2)	FFPE (cervical lymph node)	*ARID1A*	p.G87X	37.00
		STAT3	p.E166Q	2.00
		TP53	p.T155N	61.00
	Plasma ctDNA	-		
PR to 1st-line chemotherapy (2016.3)	Pericardial effusion ctDNA	*BRCA1*	p.Q1240X	2.00
		ARID1A	p.Q2100X	4.00
		CTNNB1	p.D32H	4.00
		TP53	p.T155N	7.00
PR to 1st-line chemotherapy (2016.8)	Plasma ctDNA	-		
PD on 2nd-line chemotherapy (2019.3)	Plasma ctDNA	*ARID1A*	p.Q2100X	0.20
		TP53	p.T23N	0.40
		EML4-ALK (E6:A20)		0.03

PR, partial response; PD, progressive disease; ctDNA, circulating tumor DNA; AA, amino acid; MAF, mutant allele frequency; "-," no somatic alterations detected.

chemotherapy of six cycles of pemetrexed (500 mg/m^2) and cisplatin (75 mg/m^2) and achieved a partial response (PR) according to the RECIST guideline version 1.1(7) (**Figure 1C**). No mutation was detected in the post-chemo plasma ctDNA sample (August 2016) using NGS by the same targeted gene panel (**Table 1**). A maintenance chemotherapy continued with ten cycles of pemetrexed (500 mg/m^2). The patient demonstrated a progression-free survival (PFS) of about 20 months in total during the course of first-line chemotherapy until the disease progressed with the occurrence of bone metastases in October 2017, although the primary lung lesion remained stable (**Figure 1D**).

The patient was then treated with radiofrequency (RF) ablation for the primary lung lesion in October 2017, followed by two cycles of second-line pemetrexed (500 mg/m^2) and carboplatin (6 mg/ml/min). In January 2018, the patient was switched to single-agent pemetrexed (5 cycles, 500 mg/m^2) due to severe allergic reactions to carboplatin. The primary lung tumor demonstrated durable complete response to second-line chemotherapy (**Figure 1E**). However, serum CEA levels steadily increased and reached 105.4 μg/L by February 2019 (**Figure 1A**). Remarkably, an *EML4-ALK* fusion variant (*E6:A20*) was detected at a low allele frequency (AF) of 0.03% in plasma ctDNA using the same targeted NGS panel (**Table 1**). However, it could not be validated by IHC or FISH due to an insufficient quantity of primary lesion for biopsy. Considering that there were few or no other treatment options, despite a low abundance of *EML4-ALK* fusion, we still determined to treat this patient with alectinib (600 mg twice daily) in March 2019 upon the diagnosis of multiple brain metastases by brain magnetic resonance imaging (MRI; **Figures 1F**, **2**). CEA levels declined markedly (**Figure 1A**) following the treatment and a significant reduction in the size of brain lesions was also observed (**Figure 2**). After 3 months, metastatic brain tumors disappeared completely, while the primary lung lesion remained under control (**Figure 1G**) and CEA levels dropped to 2.2 μg/L (**Figure 1A**). The patient remained relapse-free during the entire follow-up period of about 13 months up till April 2020 (**Figures 1H**, **2**).

DISCUSSION

Here, we describe an advanced lung adenocarcinoma case with a low-abundance *EML4-ALK* fusion detected in plasma ctDNA. This patient had a progression of disease after two lines of chemotherapy but displayed a durable complete response to alectinib including brain metastases. In accordance with the results of the ALEX phase three trial (5), alectinib showed potent systemic and CNS efficacy and low toxicity in the patients carrying a low frequency of *EML4-ALK* fusion, further consolidating alectinib as the standard of care for untreated, advanced *ALK*+ NSCLC, irrespective from the presence or absence of baseline CNS metastases.

We acknowledge that the sensitivity of fusion detection in cell-free DNA (cfDNA) was reported to be lower than that for mutations or indels (8), and differences in fusion detection were also noted between different cfDNA NGS assays (9). However, cfDNA should be considered as a rule in vs. a rule out test even if the AF is lower than the reportable threshold of the cfDNA assay. Furthermore, it has been largely debated on TKI efficacy irrespective of mutation AF in advanced NSCLC including *EGFR*-mutant tumors. Given that a high abundance of *EGFR* activating mutation was reported to be significantly associated with better objective response to EGFR TKIs and greater PFS benefits (10, 11), a treatment regimen of chemotherapy in combination with TKIs may be considered for advanced NSCLC with *EGFR* activating mutations of low AF. It is also worth noting that this *EML4-ALK* aberration was not detected at diagnosis or during first-line chemotherapy by the same NGS-based mutation panel under the same detection threshold. This tumor-plasma discordance may be partly explained by a high degree of tumor heterogeneity of the patient, highlighting the importance of intra-patient tumor heterogeneity as previously reported in *EGFR*-mutant NSCLC (12–14). Together, these data underscored the translational utility of NGS-based genomic sequencing of liquid biopsy in guiding clinical practice, which allows a more comprehensive analysis of tumor heterogeneity.

Although alectinib has demonstrated potent antitumor activity against *ALK*-rearranged NSCLC, the disease inevitably relapses in the clinic mainly owing to acquired therapy resistance mediated by multiple mechanisms. A number of ALK inhibitors including brigatinib (15) and lorlatinib (16) have been documented to have highly selective activity against *ALK* mutants resistant to first- and second- generation ALK-TKIs. In particular, lorlatinib was reported to be very active against almost all *ALK* mutants, including the G1202R variant. More importantly, lorlatinib displayed a strong brain-penetrant property and anticancer potency toward intracranial metastatic tumors in a phase 1, dose-escalation trial of advanced *ALK*- or *ROS1*-positive NSCLC patients, most of whom had CNS metastases (17). Late phases of clinical trials of third generation ALK inhibitors

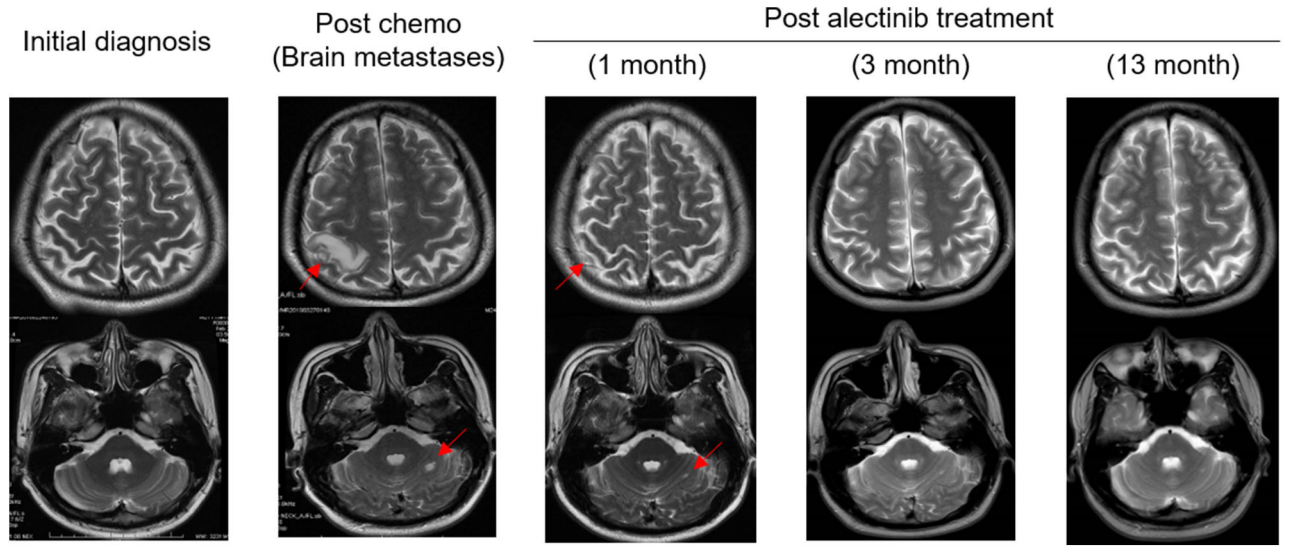


FIGURE 2 | Serial MRI scans showing patient's metastatic brain tumors responded completely to alectinib therapy. The patient achieved partial response (PR) 1 month after the initiation of alectinib therapy, and the lesions disappeared (CR) completely at the 3- and 13- month post-alectinib treatment time points. Arrows: metastatic brain tumors. CR, complete response.

are currently undergoing, which may open up a new avenue for patients who develop brain metastases after the acquisition of resistance mechanisms to currently available ALK-TKIs.

In conclusion, this case report emphasizes the importance of NGS-based genomic sequencing of liquid biopsy in disease monitoring and therapy guidance and highlights the superior efficacy of alectinib than chemotherapy in primary treatment of *ALK*+ NSCLC patients with CNS metastases, including those with low-abundance *ALK* rearrangements.

LABORATORY INVESTIGATIONS AND DIAGNOSTIC TESTS

Comprehensive genomic profiling was performed using next generation sequencing by targeting 382 cancer-relevant genes in a Clinical Laboratory Improvement Amendments-certified, College of American Pathologists-accredited laboratory (Nanjing Geneseeq Technology, Nanjing, Jiangsu Province, China). In brief, genomic DNA were extracted from cervical lymph node biopsy specimen using the DNeasy Blood & Tissue kit (Qiagen) according to the manufacturer's protocols. Cell-free (cfDNA) from pericardial effusion or plasma samples was extracted using the QIAamp Circulating Nucleic Acid kit (Qiagen). Approximately 200 ng of cfDNA was used for subsequent library preparation using the KAPA Hyper Prep kit (KAPA Biosystems) according to manufacturer's suggestions for different sample types. Sequencing library preparation, targeted gene enrichment, and sequencing data processing were carried out following the methods as previously described (18).

CLINICAL PRACTICE POINTS

- An advanced lung adenocarcinoma patient with a low abundance of *EML4-ALK* fusion demonstrated a durable complete response to alectinib.
- Analysis of plasma ctDNA changes using NGS-based liquid biopsy assays holds great promise for tracing disease progression or recurrence and guiding treatment decision-making.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the First Affiliated Hospital of Anhui Medical University, Hefei, Anhui, China. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the patient for publication of this case report and any accompanying images.

AUTHOR CONTRIBUTIONS

YZhu and RJ conducted data curation and project management. YS, LZ, QO, and XW reviewed and analyzed data. YZha designed the concept and methodology and supervised the entire study. QO, MY, and YZha wrote the manuscript and all authors read and approved the final manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Natural Science Foundation of Anhui Province (Grant No. KJ2018A0208) and the Foundation of the First Affiliated Hospital of Anhui Medical University (Grant No. 2019KJ05).

ACKNOWLEDGMENTS

We would like to thank the patient and his family for participating in this study and giving permission for collection of clinical samples for NGS testing. We also thank all the investigators and research staff involved in this case study.

REFERENCES

1. Kwak EL, Bang YJ, Camidge DR, Shaw AT, Solomon B, Maki RG, et al. Anaplastic lymphoma kinase inhibition in non-small-cell lung cancer. *N Engl J Med*. (2010) 363:1693–703. doi: 10.1056/NEJMoa1006448
2. Costa DB, Shaw AT, Ou SH, Solomon BJ, Riely GJ, Ahn MJ, et al. Clinical experience with crizotinib in patients with advanced ALK-rearranged non-small-cell lung cancer and brain metastases. *J Clin Oncol*. (2015) 33:1881–8. doi: 10.1200/JCO.2014.59.0539
3. Guerin A, Sasane M, Zhang J, Culver KW, Dea K, Nitulescu R, et al. Brain metastases in patients with ALK+ non-small cell lung cancer: clinical symptoms, treatment patterns and economic burden. *J Med Econ*. (2015) 18:312–22. doi: 10.3111/13696998.2014.1003644
4. Gadgeel SM, Shaw AT, Govindan R, Gandhi L, Socinski MA, Camidge DR, et al. Pooled analysis of CNS response to alectinib in two studies of pretreated patients with ALK-positive non-small-cell lung cancer. *J Clin Oncol*. (2016) 34:4079–85. doi: 10.1200/JCO.2016.68.4639
5. Peters S, Camidge DR, Shaw AT, Gadgeel S, Ahn JS, Kim DW, et al. Alectinib versus crizotinib in untreated ALK-positive non-small-cell lung cancer. *N Engl J Med*. (2017) 377:829–38. doi: 10.1056/NEJMoa1704795
6. Aggarwal C, Thompson JC, Black TA, Katz SI, Fan R, Yee SS, et al. Clinical implications of plasma-based genotyping with the delivery of personalized therapy in metastatic non-small cell lung cancer. *JAMA Oncol*. (2019) 5:173–80. doi: 10.1001/jamaoncol.2018.4305
7. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer*. (2009) 45:228–47. doi: 10.1016/j.ejca.2008.10.026
8. Mccoach CE, Blakely CM, Banks KC, Levy B, Chue BM, Raymond VM, et al. Clinical utility of cell-free DNA for the detection of ALK fusions and genomic mechanisms of ALK inhibitor resistance in non-small cell lung cancer. *Clin Cancer Res*. (2018) 24:2758–70. doi: 10.1158/1078-0432.CCR-17-2588
9. Supplee JG, Milan MSD, Lim LP, Potts KT, Sholl LM, Oxnard GR, et al. Sensitivity of next-generation sequencing assays detecting oncogenic fusions in plasma cell-free DNA. *Lung Cancer*. (2019) 134:96–9. doi: 10.1016/j.lungcan.2019.06.004
10. Li X, Cai W, Yang G, Su C, Ren S, Zhao C, et al. Comprehensive analysis of EGFR-mutant abundance and its effect on efficacy of EGFR TKIs in advanced NSCLC with EGFR mutations. *J Thorac Oncol*. (2017) 12:1388–97. doi: 10.1016/j.jtho.2017.06.006
11. Wang H, Zhang M, Tang W, Ma J, Wei B, Niu Y, et al. Mutation abundance affects the therapeutic efficacy of EGFR-TKI in patients with advanced lung adenocarcinoma: a retrospective analysis. *Cancer Biol Ther*. (2018) 19:687–94. doi: 10.1080/15384047.2018.1450115
12. Thress KS, Brant R, Carr TH, Dearden S, Jenkins S, Brown H, et al. EGFR mutation detection in ctDNA from NSCLC patient plasma: a cross-platform comparison of leading technologies to support the clinical development of AZD9291. *Lung Cancer*. (2015) 90:509–15. doi: 10.1016/j.lungcan.2015.10.004
13. Chabon JJ, Simmons AD, Lovejoy AF, Esfahani MS, Newman AM, Haringsma HJ, et al. Circulating tumour DNA profiling reveals heterogeneity of EGFR inhibitor resistance mechanisms in lung cancer patients. *Nat Commun*. (2016) 7:11815. doi: 10.1038/ncomms11815
14. Singh AP, Li S, Cheng H. Circulating DNA in EGFR-mutated lung cancer. *Ann Transl Med*. (2017) 5:379. doi: 10.21037/atm.2017.07.10
15. Katayama R, Khan TM, Benes C, Lifshits E, Ebi H, Rivera VM, et al. Therapeutic strategies to overcome crizotinib resistance in non-small cell lung cancers harboring the fusion oncogene EML4-ALK. *Proc Natl Acad Sci USA*. (2011) 108:7535–40. doi: 10.1073/pnas.1019559108
16. Zou HY, Friboulet L, Kodack DP, Engstrom LD, Li Q, West M, et al. PF-06463922, an ALK/ROS1 inhibitor, overcomes resistance to first and second generation ALK inhibitors in preclinical models. *Cancer Cell*. (2015) 28:70–81. doi: 10.1016/j.ccell.2015.05.010
17. Shaw AT, Felip E, Bauer TM, Besse B, Navarro A, Postel-Vinay S, et al. Lorlatinib in non-small-cell lung cancer with ALK or ROS1 rearrangement: an international, multicentre, open-label, single-arm first-in-man phase 1 trial. *Lancet Oncol*. (2017) 18:1590–9. doi: 10.1016/S1470-2045(17)30680-0
18. Yang Z, Yang N, Ou Q, Xiang Y, Jiang T, Wu X, et al. Investigating novel resistance mechanisms to third-generation EGFR tyrosine kinase inhibitor osimertinib in non-small cell lung cancer patients. *Clin Cancer Res*. (2018) 24:3097–107. doi: 10.1158/1078-0432.CCR-17-2310

Conflict of Interest: YS and LZ are the employees of Nanjing Geneseeq Technology Inc., China. QO, MY, and XW are the employees of Geneseeq Technology Inc., Canada.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership