



A novel NGS-based diagnostic algorithm for classifying multifocal lung adenocarcinomas in pNOMO patients

Xin Zhang¹, Xiaoxi Fan¹, Changbo Sun^{1,2}, Liang Wang³, Yuan Miao³ , Liming Wang¹, Peng Yang⁴, Yang Xu⁴ , Xue Ren⁴, Xue Wu⁴ and Shun Xu^{1*}

¹Department of Thoracic Surgery, The First Hospital of China Medical University, Shenyang, Liaoning, PR China

²Department of Thoracic Surgery, The University of Tokyo Graduate School of Medicine, Tokyo, Japan

³Department of Pathology, The First Hospital of China Medical University, Shenyang, Liaoning, PR China

⁴Geneseeq Research Institute, Nanjing Geneseeq Technology Inc., Nanjing, Jiangsu, PR China

*Correspondence to: Shun Xu, Department of Thoracic Surgery, The First Hospital of China Medical University, 155 Nanjing North Street, Shenyang 110001, Liaoning, PR China. E-mail: xushun610539@sina.com

Abstract

The classification of multifocal lung adenocarcinomas (MLAs), including multiple primary lung adenocarcinomas (MPLAs) and intrapulmonary metastases (IPMs), has great clinical significance in staging and treatment determination. However, the application of molecular approaches in pNOMO MLA diagnosis has not been well investigated. Here, we performed next-generation sequencing (NGS) analysis in 45 pNOMO MLA patients (101 lesion pairs) who were initially diagnosed as having MPLA by comprehensive histologic assessment (CHA). Five additional patients with intrathoracic metastases were used as positive controls, while 197 patients with unifocal lung adenocarcinomas (425 random lesion pairs) were used as negative controls. By utilizing a predefined NGS criterion, all IPMs in the positive control group could be accurately classified, whereas 13 lesion pairs (3.1%) in the negative control cohort were misdiagnosed as IPMs. Additionally, 14 IPM lesion pairs were diagnosed in the study group, with at least 7 misdiagnoses. We thus developed a refined algorithm, incorporating both NGS and histologic results, that could correctly diagnose all the known MPLAs and IPMs. In particular, all IPMs identified by the refined algorithm were diagnosed to be IPMs or suspected IPMs by CHA reassessment. The refined algorithm-diagnosed MPLAs patients also had significantly better progression-free survival than the refined algorithm-diagnosed IPMs ($p < 0.0001$), which is superior to conventional NGS or CHA diagnoses. Overall, we developed an NGS-based algorithm that could accurately distinguish IPMs from MPLAs in MLA patients. Our results demonstrate a promising clinical utility of NGS to complement traditional CHA-based MLA diagnosis and help determine patient staging and treatment.

Keywords: multifocal lung adenocarcinoma; multiple primary lung adenocarcinoma; intrapulmonary metastasis; comprehensive histologic assessment; next-generation sequencing

Received 13 September 2022; Revised 26 October 2022; Accepted 23 November 2022

Conflict of interest statement: PY, YX, XR, and XW are employees of Nanjing Geneseeq Technology Inc. All other authors declare no conflict of interest.

Introduction

Due to the wide use of high-resolution imaging technologies (e.g. low-dose computed tomography [CT]) in regular clinical evaluation, multifocal lung adenocarcinomas (MLAs) are increasingly observed in clinical practice worldwide [1]. MLAs can be classified into multiple primary lung adenocarcinomas (MPLAs) and intrapulmonary metastases (IPMs), and the distinction of the two types of MLA is important to stage classification and subsequent treatment determination.

According to the American Joint Committee on Cancer staging system, MPLAs are staged individually based on the highest-T-stage lesion, while IPMs can be staged as pT3, pT4, or even pM1a based on the location of the additional lesions [2]. The MPLA patients without lymph node and distant metastases will be staged as invasive adenocarcinoma (IA) if the diameter of the largest lesion is <3 cm, and these patients would be likely to undergo wait and watch after surgical resection. On the other hand, patients with the same conditions will be staged as IIB or

above if they have IPMs, and these patients should instead receive adjuvant therapy after surgical resection [3]. Nevertheless, despite its clinical importance, the distinction between the two MLA types is usually challenging to pathologists and clinicians.

The first criteria for the distinction between MPLAs and IPMs were proposed by Martini and Melamed in 1975, which relied on the histologic pattern of each lesion [4]. The Martini and Melamed criteria were further modified to become the comprehensive histologic assessment (CHA), which was the most commonly recommended approach to separate MPLAs from IPMs [5–7]. Moreover, CHA combined with a nonmucinous lepidic component with mild atypia was confirmed as a better option for MPLA diagnosis [8,9]. However, given that the morphologic assessment mainly depends on the judgment of the pathologist, wide interobserver variability is inevitable [10,11] and a more objective and accurate approach is thus imperative.

Molecular methods are promising unbiased approaches to distinguish MPLAs from IPMs. In recent years, several molecular methods for MLA analysis were explored and evolved, including array comparative genomic hybridization (aCGH) [12], the genetic assessment of a single or a subset of driver mutations (e.g. *TP53*, *EGFR*, *KRAS*, and *BRAF*) [13,14], microsatellite analysis [15,16], and next-generation sequencing (NGS) [17–21]. However, most of these studies were based on either limited genomic information or limited sample size, which compromise the ability to accurately elucidate the clonal relationship among different lesions in MLAs. A recent study utilized broad-panel NGS (468 cancer-related genes) on 60 non-small cell lung cancer (NSCLC) patients with multifocal tumors, which included a total of 128 lesions, and showed that NGS was superior to CHA for distinguishing multiple primary lung cancer and IPMs [22]. Nevertheless, large-scale NGS studies in MLA patients with pN0M0 disease, especially in the Asian population, are still lacking. In this study, we developed an NGS-based diagnostic algorithm to distinguish MPLAs from IPMs, aiming to provide a more accurate diagnosis and more appropriate treatment regimens for MLA patients.

Materials and methods

Patient selection and data collection

From November 2018 to November 2019, a total of 45 MLA patients without lymph node (N0) and distant

metastases (M0) were diagnosed in The First Hospital of China Medical University. All 45 MLA patients were initially classified as MPLA by the CHA approach and they were assigned to the study cohort. These 45 patients underwent pulmonary resection of at least two lesions, including adenocarcinoma *in situ* (AIS), minimally invasive adenocarcinoma (MIA), and IA. An additional five IA patients (a total of 15 lesions) with parietal pleura metastases were assigned as a positive control cohort, given that pleura metastases could be easily diagnosed as compared to IPMs. The clinical, radiologic, pathologic, and molecular data of patients from both the study cohort and the positive control cohort were collected. Furthermore, NGS data from a total of 197 patients with unifocal lung adenocarcinomas were obtained from previously tested patients in The First Hospital of China Medical University (supplementary material, Table S1), and these patients were randomly grouped to form 425 patient pairs (supplementary material, Table S2), which resemble the MPLAs and were used as a negative control cohort.

Comprehensive histologic assessment

CHA was performed by two experienced thoracic pathologists (LW and YM) independently based on the proportion of components, including lepidic, acinar, papillary, micropapillary, and solid. The criteria for distinguishing MPLAs and IPMs were reported by Detterbeck *et al* [23].

DNA extraction

For each patient, archived formalin-fixed paraffin-embedded (FFPE) blocks of tumor tissues that were collected through surgical resection were used for the DNA extraction. The tumor content of all samples was confirmed to be at least 10% by the pathologists (supplementary material, Tables S1 and S3). FFPE samples were deparaffinized with xylene, and genomic DNA was extracted using QIAamp DNA FFPE Tissue Kit (Qiagen, Germantown, MD, USA) with matched blood normal control in all cases according to the manufacturer's protocols. Purified genomic DNA was qualified by Nanodrop 2000 for A260/280 and A260/A230 ratios (Thermo Fisher Scientific, Waltham, MA, USA). All DNA samples were quantified by Qubit 3.0 using the dsDNA HS Assay Kit (Life Technologies, Carlsbad, CA, USA) according to the manufacturer's recommendation.

Targeted NGS

Customized xGen lockdown probes (Integrated DNA Technologies, San Diego, CA, USA) targeting 425 cancer-relevant genes were used for hybridization enrichment. The capture reaction was performed with Dynabeads M-270 (Life Technologies) and xGen Lockdown Hybridization and Wash Kit (Integrated DNA Technologies) according to manufacturers' protocols. Captured libraries were on-beads PCR amplified with Illumina p5 (5'-AAT GAT ACG GCG ACC ACC GA-3') and p7 168 primers (5'-CAA GCA GAA GAC GGC ATA CGA GAT-3') in KAPA HiFi HotStart 169 ReadyMix (KAPA Biosystems, Wilmington, MA, USA), followed by purification using Agencourt AMPure XP beads. Libraries were quantified by quantitative polymerase chain reaction using KAPA Library Quantification Kit (KAPA Biosystems). Library fragment size was determined by Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA). The target-enriched library was then sequenced on Illumina HiSeq4000 platform (San Diego, CA, USA), following the manufacturer's instructions.

Sequencing data processing

Trimmomatic was used for FASTQ file quality control. Leading/trailing low quality (quality reading below 20) or N bases were removed. Paired-end reads were then aligned to the reference h19 genome (Human Genome version 19) using the Burrows–Wheeler Aligner with the default parameters. PCR duplicates were removed using Picard and local realignments around indels and base quality score recalibration was performed using genome analysis toolkit. VarScan2 was employed for the detection of single-nucleotide variations and insertion/deletion mutations with the following parameters: minimum read depth = 20, minimum base quality = 25, minimum variant allele frequency = 0.03, minimum variant supporting reads = 3, variant supporting reads mapped to both strands, and strand bias no greater than 10%. Gene fusions were identified by FACTERA and copy number variations (CNVs) were analyzed with ADTEX. The log₂ ratio cut-off for copy number gain was defined as 2.0 for tissue samples. A log₂ ratio cut-off of 0.67 was used for copy number loss detection in all sample types. The thresholds were determined from previous assay validation using the absolute CNVs detected by droplet digital PCR. Tumor mutational burden (TMB) in this study was defined as the number of somatic synonymous mutations per megabase in each sample, with hot-spot/fusion mutations excluded.

Evolutionary analysis methods

To estimate the clonal relationship between lesions, we compared the nonsynonymous somatic mutations and structural variants, which include both pathogenic variants and variants of unknown significance. The pairs with completely unique mutations were classified as MPLAs and the pairs with entirely overlapping mutations were classified as IPMs. For lesions only sharing driver gene mutations, we used Treeomics analysis, a new tool to reconstruct the tumor phylogeny using commonly available sequencing technologies [24], for the extended review. Treeomics employs a uniquely designed Bayesian inference model to account for error-prone sequencing and varying low neoplastic cell content to calculate the probability that a specific variant is present or absent in each lesion. For each case, Treeomics was used to calculate the posterior probabilities of a variant being present based on total read depth and the number of reads covering the alternative allele. We used two indices to measure genetic heterogeneity: Jaccard similarity coefficients and genetic distances ('divergence'). The Jaccard similarity coefficient is defined as the ratio of shared variants over all variants (shared plus discordant) between two samples. We calculated the Jaccard similarity coefficients of the various pairs of metastases on the basis of their validated mutations. We chose the Jaccard similarity coefficient of 0.4 to separate MPLAs from IPMs, which is derived from previous studies [25]. Genetic distance is defined as the total number of nonshared genetic variants present between two samples.

Statistical analysis

Statistical analyses were performed via SPSS 19 software (SPSS Inc, Chicago, IL, USA). Kaplan–Meier survival curve was used to analyze the progression-free survival (PFS) data of patients, and the statistical difference was analyzed using the log-rank test. The Chi-square test was used to compare two factors, with a *P* value of less than 0.05 considered statistically significant.

Results

Patient and lesion characteristics

Forty-five MLA patients without lymph node and distant metastases were initially diagnosed as MPLAs by the CHA approach, and we grouped these 45 patients as the study cohort and performed broad-panel NGS to

assess their MLA types. Of the 45 MLA patients, there were 32 females and 13 males, and the majority of them (73.3%) did not have a smoking history (Table 1). More than half of these patients (62.2%) did not have a family history of cancer, while the rest of the patients had a family history of lung cancer (20.0%) or other cancer types (17.8%) (Table 1). According to the radiologic pattern of the 45 MLA patients, they can be classified into three groups: 31 patients (68.9%) had multi-ground-glass nodules (multi-GGN), that is, their lesions showed pure GGN or mixed GGN with consolidation/tumor ratio (CTR) ≤ 0.5 ; 9 patients were grouped as solid-GGN as their lesions included one solid nodule/subsolid (CTR > 0.5) with one or more GGNs; 5 patients had multiple solid lesions, including both solid and subsolid lesions.

A total of 113 lesions were collected from the 45 MLA patients, including 31 patients with 2 lesions, 6 patients with 3 lesions, 7 patients with 4 lesions, and 1 patient with 5 lesions, and these 113 lesions can be

Table 1. Clinicopathological characteristics of the MLA patients in the study group ($n = 45$)

Patient features	Number	%
Age		
≤ 60	24	53.3
> 60	21	46.7
Sex		
Male	13	28.9
Female	32	71.1
Family history		
Lung cancer	9	20.0
Other cancer	8	17.8
None	28	62.2
Smoking history		
Current/former	12	26.7
Never	33	73.3
Number of lesions		
2	31	68.9
3	6	13.3
4	7	15.6
5	1	2.2
Imaging pattern		
Multi-GGN	31	68.9
Solid-GGN	9	20.0
Multisolid	5	11.1
Location of lesions		
Ipsilateral same lobe	17	37.8
Ipsilateral other lobe	22	48.9
Contralateral lobe	6	13.3
Surgical procedure		
Sublobar resection (SLS)	17	37.8
Lobectomy + SLS	13	28.9
Lobectomy	14	31.1
Bilobectomy	1	2.2

grouped into a total of 101 lesion pairs. One hundred and seven out of 113 (94.7%) lesions were < 3 cm, and the right upper lobe was the most frequent location ($n = 35$, 31.0%) (supplementary material, Table S4). Thirteen (11.5%) lesions were diagnosed as AIS, 36 (31.9%) lesions were MIA, and 64 (56.6%) lesions were IA. As AIS and MIA lesions generally could neither be derived from nor lead to IPMs, any lesion pairs that contain AIS/MIA are very likely to be MPLAs. Consequently, the 56 AIS/MIA-containing lesion pairs from the study cohort were grouped as the internal negative control, whereas the remaining 45 lesion pairs from the study cohort were used as the testing group (supplementary material, Figure S1).

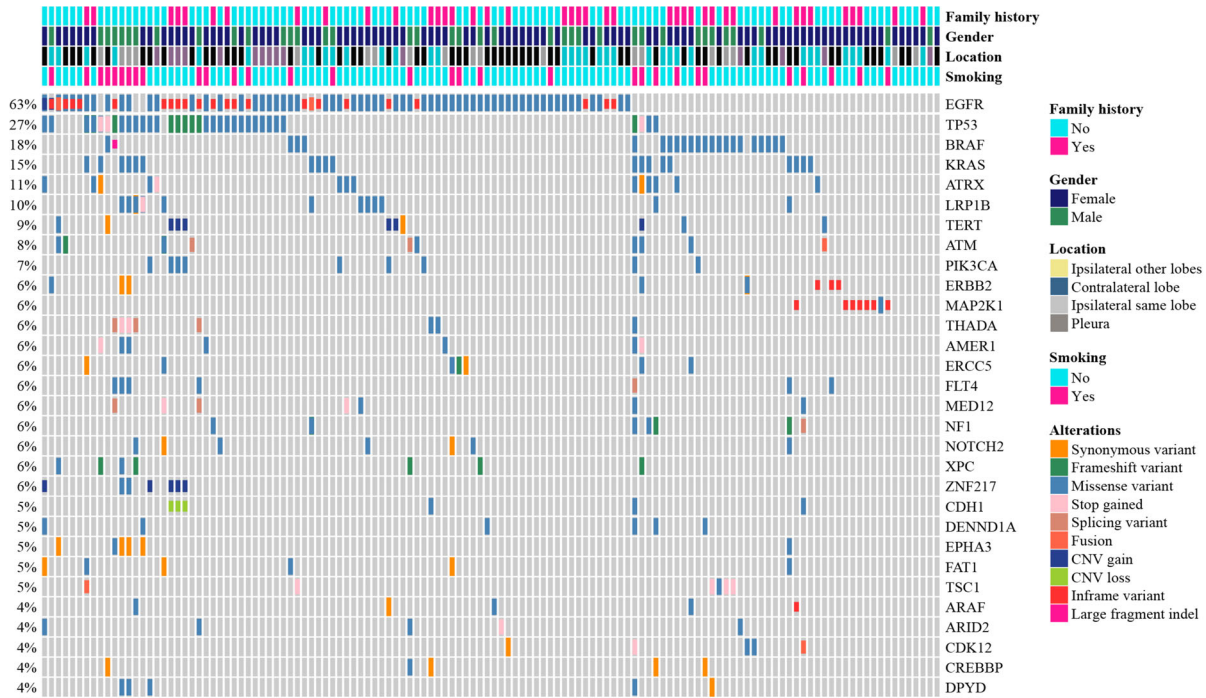
An additional five IA patients with parietal pleura metastases (15 lesions; 14 lesion pairs) were used as a positive control cohort for IPMs, given that pleura metastases could be easily and accurately diagnosed as compared to IPMs (supplementary material, Figure S1). Besides the study cohort and the positive control cohort, we obtained NGS data from 197 patients with unifocal lung adenocarcinomas (supplementary material, Table S1) from previously tested patients, and these patients were randomly grouped to form 425 patient pairs (supplementary material, Table S2), which resemble the MPLAs and were used as a negative control cohort (supplementary material, Figure S1).

The genetic features in the study cohort and the positive control cohort

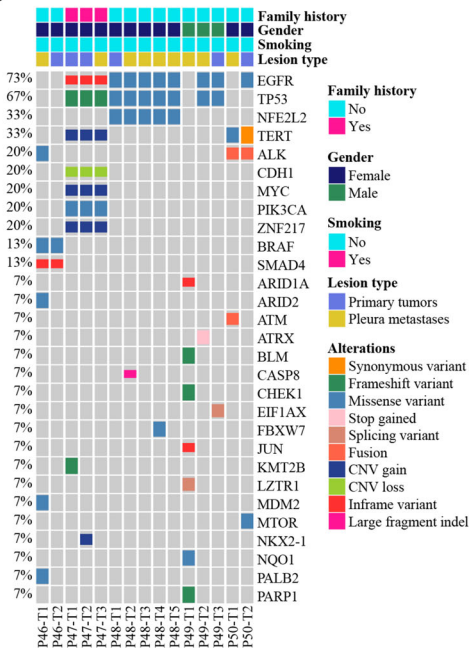
We performed NGS of 425 cancer-related genes to characterize the genetic profile of 113 lesions in the study cohort and 15 lesions in the positive control cohort. The most frequently mutated genes in the study cohort were *EGFR* (63.3%), *TP53* (26.6%), *BRAF* (18.0%), and *KRAS* (14.8%), while the mutation of other oncogenes was also detected, including *TERT* (8.6%), *PIK3CA* (7%), and *ERBB2* (6.2%) (Figure 1A). The positive control cohort demonstrates a similar genetic profile as compared to the study cohort, with *EGFR* (73.3%), *TP53* (66.7%), *NFE2L2* (33.3%), and *TERT* (26.7%) being the top altered genes (Figure 1B).

Next, we characterized the molecular features between the primary tumors and metastases of the positive control group. Among these five patients with pleura metastases, the genomic profiling showed high concordance between the primary lesion and the metastatic lesion(s) in nonsynonymous mutations (Figure 1B). For example, *EGFR* p.L858R (c.2573T>G) and *TP53* p.G105C (c.313G>T) mutations were shared among the primary

A



B



C

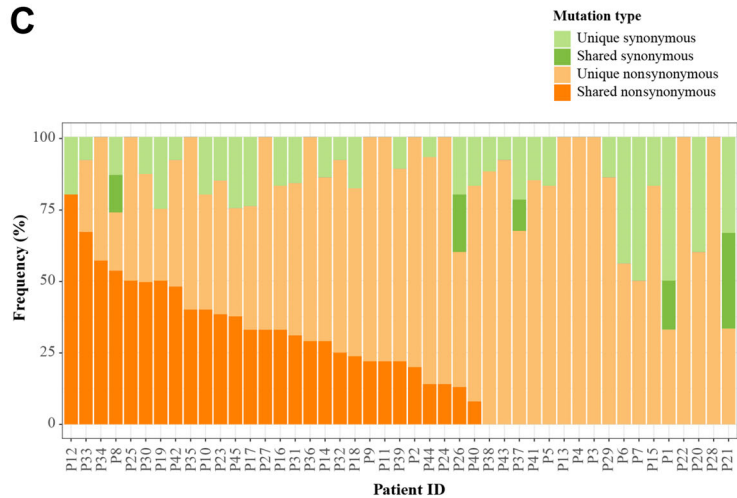


Figure 1. Analysis of the genetic features in the study group using broad-panel NGS. (A) The genetic landscape of high-frequency molecular alterations detected in 113 lesions of the study group ($n = 45$). The frequency of each mutation is shown on the left. The types of alteration are represented by the colors indicated. (B) The genetic landscape of high-frequency molecular alterations detected in 15 lesions of the positive control group ($n = 5$). The percentages of lesions identified with the indicated alterations are shown on the left. Alteration types are represented by the colors indicated. (C) Distribution of somatic mutations among the 113 lesions from 45 patients in the study group. Each column represents a patient.

lesion and five metastatic lesions in patient P48, with almost no unique mutations being detected (Figure 1B). Patient P50 shared the identical break points of the *ALK* fusion aberration (*EML4*: exon 13 ~ *ALK*: exon 20) between the primary lesion and metastatic lesion (Figure 1B). For patient P49, the primary lesion (P49-T3) and one of the metastatic lesions (P49-T2) shared the same driver mutations; however, the second metastatic lesion (P49-T1) had a completely different genetic profile compared with the other two lesions, implying that this metastatic lesion may be derived from another undetected primary tumor. Overall, the molecular analysis in the control group suggested that there was a high concordance of nonsynonymous mutations between primary and metastatic lesions in patients with pleura metastases.

To investigate the clonal relationship between the primary lesion and the metastatic lesion(s), we compared the tumors sharing multiple mutations in the study group. As shown in Figure 1C, for each lesion of the 45 MLA patients, we divided their genetic alterations into four types: shared nonsynonymous somatic mutations, unique nonsynonymous somatic mutations, shared synonymous somatic mutations, and unique synonymous somatic mutations. The majority of the genetic alterations were nonsynonymous, and more than half of the patients harbored lesions that had shared nonsynonymous changes while the rest of the patients harbored lesions that did not have shared nonsynonymous changes at all (Figure 1C).

Prediction of MLA subtypes in the positive and negative control lesion pairs

As discussed earlier, patient P49 harbored lesions from two different origins (Figure 1B), so there were 12 IPM lesion pairs and 2 MPLA lesion pairs (P49-T1 versus P49-T2 and P49-T1 versus P49-T3) in the positive control cohort. We thus calculated the gene similarity score (see [Materials and methods](#) section for more details) and the number of shared nonsynonymous mutations in the positive control cohort. As shown in Table 2, 9 out of 12 (75.0%) of IPM lesion pairs shared three nonsynonymous mutations. Additionally, using a gene similarity score of 0.4 as the cut-off point could efficiently separate all IPM lesion pairs from MPLA lesion pairs in the positive control cohort (Table 2). We, therefore, used 'gene similarity score ≥ 0.4 ' as the NGS-based criterion to distinguish IPMs from MPLAs.

We then tested the NGS-based criterion in the negative control cohort, which consisted of 197 unifocal lung adenocarcinomas and 425 randomly formed

tumor pairs to resemble MPLAs. The mean gene similarity score for the 425 tumor pairs was only 0.08, and 412 out of the 425 (96.9%) tumor pairs had Jaccard similarity coefficients smaller than 0.4 (supplementary material, Table S2). Intriguingly, although all 13 misdiagnosed tumor pairs had Jaccard similarity coefficients higher than or equal to 0.4, at least one tumor of the tumor pair had a relatively low number of detected mutations, ranging from 1 to 3 (Figure 2A). We also investigated the 56 internal negative control lesion pairs from the study cohort, and 7 out of 56 (12.5%) lesion pairs were misdiagnosed as IPMs by the NGS-based criterion, all of which had at least one lesion with ≤ 3 detected mutations (Figure 2B). These results imply that when one of the lesions has fewer detected mutations (≤ 3), the specificity of the NGS-based criterion drops significantly.

Prediction of MLA subtypes in the study cohort using the NGS-based criterion

Next, we used the NGS-based criterion to analyze the 45 lesion pairs from the testing group of the study cohort (supplementary material, Figure S1). Eighteen lesion pairs had no shared nonsynonymous mutations, while the rest of the lesion pairs shared on average ~ 2 nonsynonymous mutations (range: 1–16) (supplementary material, Table S5). A total of seven lesion pairs were judged as IPMs by the NGS-based criterion, given that they had a gene similarity score of at least 0.4 (supplementary material, Table S5). Overall, among the 101 lesion pairs in the study cohort, a total of 14 lesion pairs were classified as IPMs by the NGS-based criterion, including 7 false-positive lesion pairs from the internal control group.

Rediagnosis of MLA subtypes based on comprehensive histology assessment

As a considerable number of lesion pairs that were diagnosed as MPLAs by CHA were reclassified as IPMs by NGS, two experienced pathologists (LW and YM) independently reassessed the 113 lesions from the study cohort, especially the 14 lesion pairs with discordant results between CHA prediction and NGS prediction. All seven false-positive lesion pairs from the internal control group were still classified as MPLAs by the pathologists. For example, despite sharing two nonsynonymous mutations, both of the two lesions in patient P12 were MIAs and dominated by the lepidic growth pattern without vascular, neural, or pleural invasion (Figure 3A), which strongly suggests

Table 2. The Jaccard coefficient of the lesion pairs from the positive control cohort

Patient ID	Pair	Number of nonsynonymous mutations in lesion 1	Number of nonsynonymous mutations in lesion 2	Shared nonsynonymous mutations	Gene similarity (Jaccard)
P46	T1 versus T2	8	2	2	1
P47a	T1 versus T2	4	3	3	1
P47b	T1 versus T3	4	5	3	1
P48a	T1 versus T2	4	4	3	0.99
P48b	T1 versus T3	4	3	3	0.99
P48c	T1 versus T4	4	4	3	0.99
P48d	T1 versus T5	4	4	3	0.99
P48e	T2 versus T3	4	3	3	1
P48f	T2 versus T4	4	4	3	1
P48g	T3 versus T4	3	4	3	1
P49a	T1 versus T2	12	3	0	0.22
P49b	T1 versus T3	12	3	0	0.34
P49c	T2 versus T3	3	3	2	0.83
P50	T1 versus T2	4	4	1	1

that they were MPLAs. For the seven NGS-determined IPM lesion pairs in the testing group, four of them (i.e. P8-T1 versus T2, P31-T2 versus T3, P42-T1 versus T2, and P45-T1 versus T2) were re-diagnosed as IPMs after comprehensively examining multiple additional tissue slides from the original tumor FFPE specimens (supplementary material, Table S4). For instance, the two lesions from patient P8 shared four nonsynonymous mutations and both of them were 40% acinar, 40% lepidic, and 20% micropapillary; by reassessing different hematoxylin and eosin (H&E) slices from the two lesions, cancer cell nests were observed

within bronchioles and capillaries (Figure 3B), thus confirming that they were IPMs.

Additionally, one of the lesion pairs (T4 versus T5) from patient P17 shared eight nonsynonymous mutations and had a gene similarity score of 0.93; however, it was still classified as MPLAs even after comprehensive histological reassessment. Patient P17 harbored five different lesions, including two in the right upper lobe within the same segment (lesions P17-T4 and P17-T5), one in the right lower lobe (lesion P17-T3), and two in the left upper lobe (lesions P17-T1 and P17-T2) (Figure 3C, CT images). The histologic

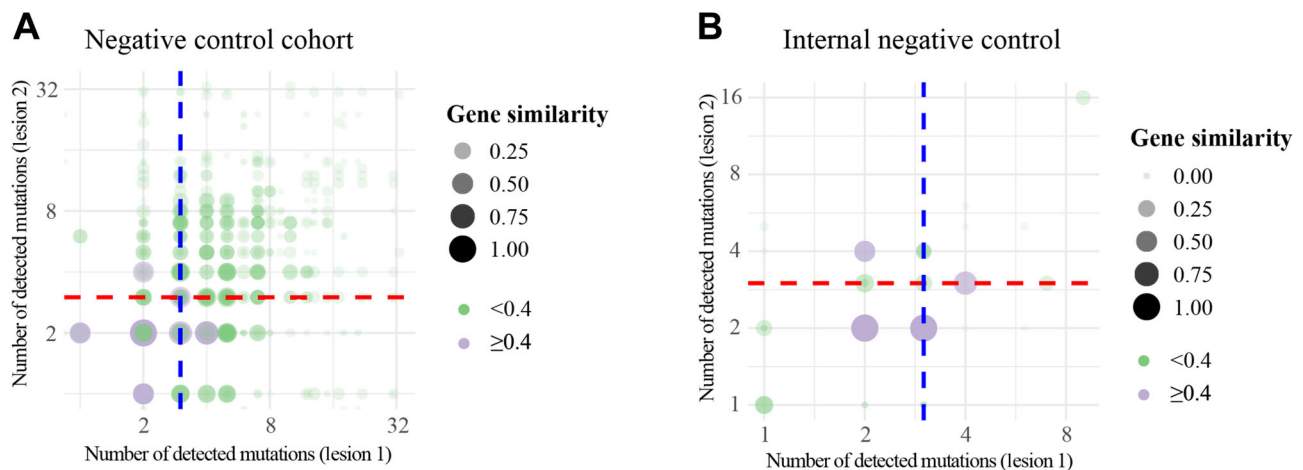


Figure 2. The misdiagnosed lesion pairs were those with low detected nonsynonymous mutations. (A) The scatter plot of lesion pairs ($n = 425$) with different numbers of detected mutations and gene similarity score for the negative control cohort. The number of detected nonsynonymous mutations in each lesion of the lesion pair is indicated on the x- and y-axis, respectively. The magnitude of gene similarity is shown by both the size and the color of the dot. The blue and red dashed lines represented the threshold of three detected mutations in lesion 1 and lesion 2, respectively. (B) The layout of the scatter plot is similar to (A), except that the analysis was performed using patients from the internal negative control group from the study cohort ($n = 56$).

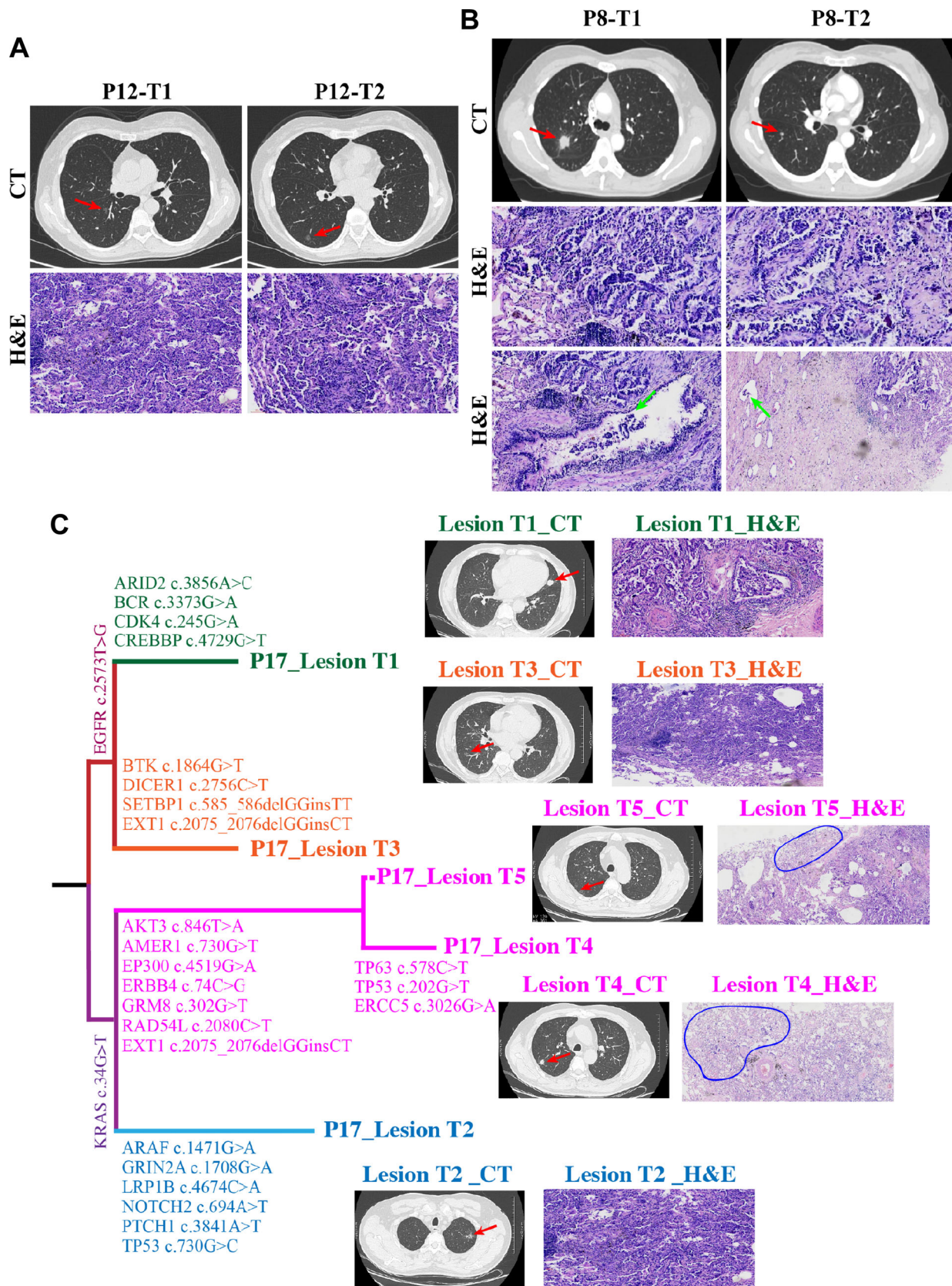


Figure 3. Legend on next page.

patterns between lesions P17-T4 and P17-T5 were different, with one being 30% lepidic, 50% micropapillary, and 20% acinar and the other being 90% lepidic and 10% micropapillary (Figure 3C, H&E images), suggesting this lesion pair should be diagnosed as MPLAs by CHA, which is significantly different from the conclusion derived from their NGS data. Notably, during the histologic reassessment, the pathologists noticed the spread of cancer cells into the alveolar space in the lung parenchyma adjacent to both of the lesions (Figure 3C, the blue circle in the H&E image of P17-T4 and P17-T5). As a result, we speculated that the lesions had spread through air spaces (STAS) to produce metastasis, which, if so, would classify this lesion pair as IPMs.

Developing a refined NGS algorithm by combining the CHA and NGS diagnosis

Given that our initial NGS-based criterion misclassified multiple lesion pairs, especially when the lesion harbored fewer detectable mutations, we developed a refined NGS algorithm by combining the CHA and NGS diagnoses. As shown in Figure 4A, because AIS and MIA lesions generally could neither be derived from nor lead to IPMs, any lesion pairs that contained AIS/MIA were classified as MPLAs. For the remaining lesion pairs, the ones with zero shared nonsynonymous mutations were classified as MPLAs, whereas those with more than 3 shared nonsynonymous mutations were classified as IPMs (Figure 4A). For lesion pairs with 1–3 shared nonsynonymous mutations, we classified lesion pairs with gene similarity score <0.4 as MPLAs; for the rest of the lesion pairs, if both lesions had more than three detected nonsynonymous mutations, they were classified as IPMs; otherwise, they were grouped based on the CHA classification (Figure 4A). Based on the refined NGS algorithm, all known MPLAs in the internal negative control group were correctly classified (Figure 4B). Furthermore, among the testing group, only five lesion pairs, including the four lesion pairs that were reassessed to be IPMs and one lesion pair of suspected IPMs, were

diagnosed as IPMs by the refined NGS algorithm (Figure 4B).

PFS analysis for patients stratified by the two different NGS approaches

Because MPLA patients are expected to have better clinical outcomes than IPM patients [26], we performed PFS analyses for patients stratified by the two different MLA diagnostic approaches. For the 45 patients in the study cohort, the median follow-up time was 20 months (mean = 20.24; range: 14–26 months), and none of the patients died according to the last follow-up. The patients were grouped as IPMs if they harbored at least one IPM lesion pair; otherwise, they were grouped as MPLAs. Based on the NGS-based criterion, 10 patients were classified as IPM and they had significantly worse PFS than the remaining 35 patients ($p = 0.0026$; Figure 4C). On the other hand, by the refined NGS algorithm, only five patients were diagnosed as IPM, and the separation of PFS between MPLA and IPM was even more significant ($p < 0.0001$; Figure 4D). Strikingly, for the five patients (i.e. P12, P19, P25, P33, and P34) who were diagnosed as IPM by the NGS-based criterion but were diagnosed as MPLA by the refined NGS algorithm, none of these patients had disease progression according to the last follow-up, suggesting that these five patients were likely to represent false-positive diagnosis resulting from fewer detected mutations in their tumor lesion.

Discussion

NGS is widely used in clinical practice for guiding treatment with tyrosine-kinase inhibitors [27]. Recently, with the development of immune checkpoint inhibitor (ICI) therapy, TMB is considered as a promising predictive biomarker for ICI treatment [28], and broad-panel NGS is recommended by oncologists to

Figure 3. Phylogenetic and imaging analyses of patients reassessed by CHA. (A) Patient P12 harbored two lesions, one in the right lower lobe superior segment (P12-T1) and the other in the right lower lobe basal segment (P12-T2). The red arrows on the CT scans indicate the positions of the tumors. (B) Patient P8 harbored two lesions, one in the right upper lobe posterior segment (P8-T1) and the other in the right lower lobe basal segment (P8-T2). The red arrows on the CT scans indicate the positions of the tumors. The green arrows on the H&E images indicate the cancer cell nests within bronchioles and capillaries. (C) Patient P17 harbored five lesions, including two in the right upper lobe (lesions P17-T4 and P17-T5), one in the right lower lobe (lesion P17-T3), and two in the left upper lobe (lesions P17-T1 and P17-T2). Certain genetic alterations during tumor evolution are labeled at the stem and branches of the phylogenetic tree. CT scans of the five lesions are shown on the left, with red arrows to indicate the positions of the tumors. H&E images are illustrated on the right, and evidence of metastasis by STAS is highlighted using blue circles.

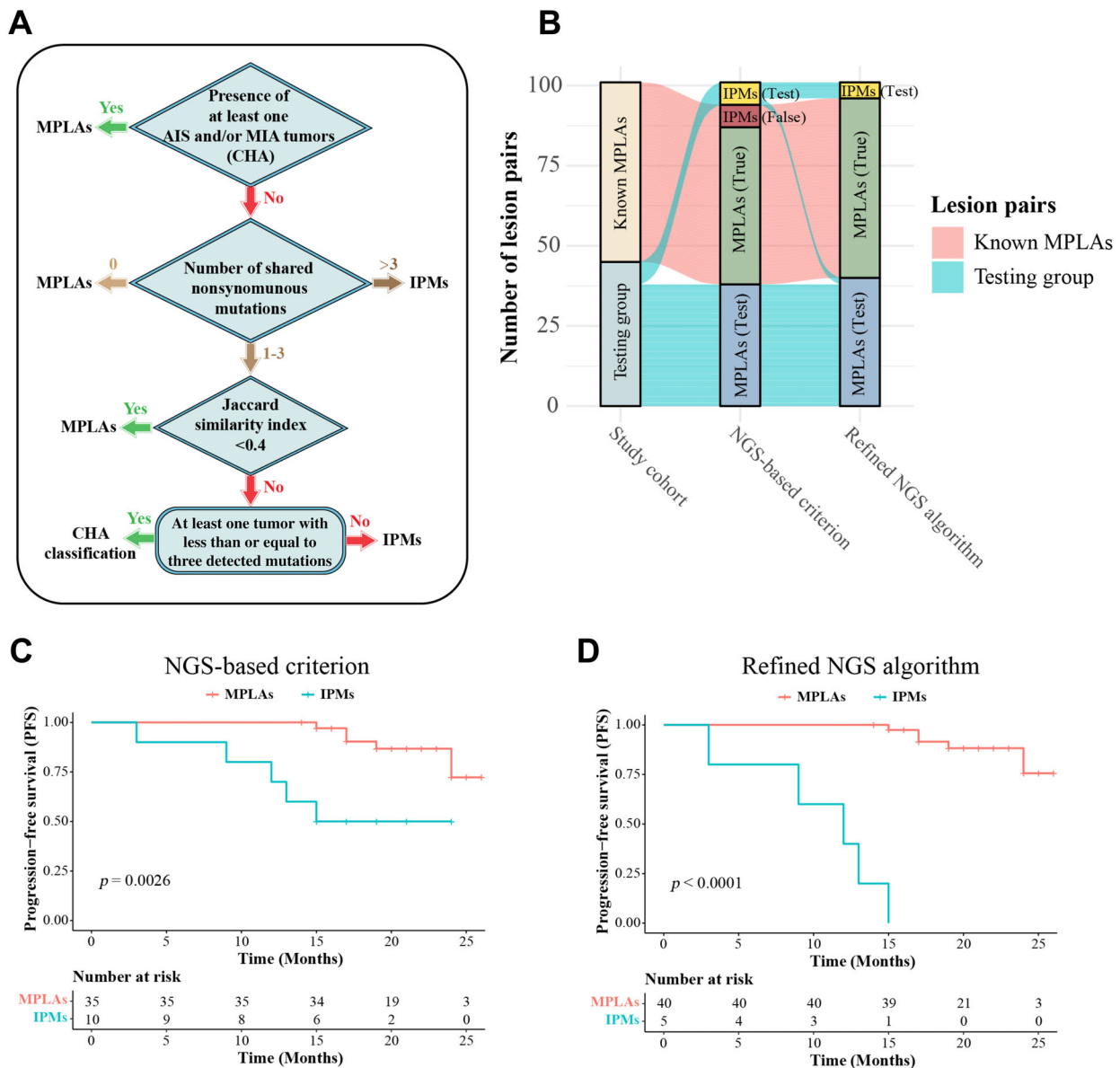


Figure 4. The refined NGS algorithm improves distinction between MPLAs and IPMs. (A) Diagram of the refined NGS algorithm for MLA classification. (B) Ggalluvial plot illustrating the diagnosis of lesion pairs in the study group by either the initial NGS-based criterion or the refined NGS algorithm. The ‘known MPLAs’ were all the lesion pairs from the internal control group, ‘MPLAs (True)’ denotes all the lesion pairs that were from the internal control group and were correctly diagnosed as MPLAs, ‘IPMs (False)’ denoted all the lesion pairs that were from the internal control group and were incorrectly diagnosed as IPMs, and ‘MPLAs (Test)’ and ‘IPM (Test)’ were from the testing group, whose true MLA subtypes were unknown. (C and D) Kaplan–Meier curves of disease-free survival in 45 MLA patients from the study cohort in strata of MLA subtypes determined by either (C) the NGS-based criterion or (D) the refined NGS algorithm.

calculate TMB. Our study developed a novel NGS algorithm that can objectively and accurately classify MLAs. Specifically, the NGS algorithm correctly diagnosed all the known MPLAs and IPMs in the positive and negative control groups, while in the testing group, the NGS-determined IPM patients had

significantly worse PFS than NGS-determined MPLA patients ($p < 0.0001$), which further supports the high accuracy of the NGS-based diagnostic algorithm for MLA subtyping.

Determining the clonal relationship between MLA lesions has major implications for staging and treatment,

especially for patients without lymph node and distant metastases. We selected MLA patients without lymph node and distant metastases as the study group. To the best of our knowledge, this is the largest study to explore the performance of broad-panel NGS for distinguishing IPMs from MPLAs in synchronous MLA patients with pNOM0 disease. For the past two decades, molecular methods have been intensively investigated to elucidate the clonal relationship among different lesions within a patient. For example, aCGH has long been used as a gold standard to distinguish IPMs from MPLAs [9,29,30]. Additionally, sequencing certain driver genes or a small panel of genes was employed in multiple previous studies [31,32]. With the rapid development of NGS technology in recent years, more and more studies started to utilize broad-panel NGS, whole exome sequencing (WES), or whole genome sequencing, which has multiple advantages over aCGH or small-panel sequencing [22], to investigate multifocal tumors. For example, by using MSK-IMPACT (a panel of 505 cancer-related genes) in a clinical sequencing cohort, Chang *et al* found that results from the broad-panel sequencing could be used to supplement the traditional histologic approaches to classify multifocal NSCLC [22]. Similarly, Li *et al* utilized WES to help distinguish multicentric origin lesions from intramural metastatic lesions for multifocal esophageal squamous cell carcinomas, especially for tumors without well-defined histological features [33]. Based on broad-panel NGS results, our newly developed NGS-based diagnostic algorithm demonstrated promising potential to facilitate the distinction between MPLAs and IPMs with high diagnostic accuracy. Furthermore, the current expansion of large-scale molecular pathology allows for a broader application of our NGS-based diagnostic algorithm as more and more patients receive comprehensive genomic profiling at the time of surgical resection.

The CHA approach is commonly used for classifying MLAs in clinical practice. However, the results from CHA diagnosis usually suffer from high variability and discordant results have sometimes been observed between CHA prediction and NGS prediction [22]. Consistent with this, although all 45 patients, as well as 101 lesion pairs, in our study cohort were diagnosed to be MPLAs by initial CHA analysis, 4 patients were reclassified as IPMs after comprehensive histological reassessment. In addition, one patient (P17) was suspected to harbor an IPM lesion pair. Histologically, the lepidic pattern is considered as the earlier morphologic change of adenocarcinoma and always represents primary lesions [34]. The lepidic content in lesions P17-T4 and P17-T5 of patient P17 were 30 and 90%, respectively, so CHA determined

the lesion pair as MPLAs. As the NGS results strongly suggested that the lesion pair were IPMs, the pathologists rechecked multiple pathological sections and found evidence of metastasis due to STAS. STAS is a newly defined pattern of invasion in lung adenocarcinoma, which can be observed in early-stage adenocarcinoma and have an impact on prognosis [35–38]. Therefore, we speculate that the lepidic pattern, in some situations, represents a kind of STAS-caused metastasis. In addition, we observed highly frequent concordant mutations in MIA lesions, particularly in patients with IPMs. However, we did not find STAS in their pathological sections. Similar observations were previously reported by Li *et al* [39], although the underlying pathogenesis was still unclear.

A limitation of our study was the lack of long-enough follow-up time; none of the patients died based on the last follow-up, so we do not have overall survival data for the patients. We have scheduled patient follow-up every 3–6 months, and we intend to conduct a subsequent study when the survival data become more mature. Another limitation was that we did not have an independent MPLA patient cohort with known MPLA subtypes to further validate our NGS algorithm. Indeed, this is the limitation for most MPLA studies, as currently there is no gold standard to classify MPLA subtypes. Also, because all of our tumor samples were FFPE specimens, many of which have been stored for 3–4 years, these samples are not suitable for RNA-sequencing analysis to further confirm some of the genetic alterations we identified, such as the splicing variants. Additionally, our positive control cohort was relatively small compared with the study and negative control cohorts. This is because most parietal pleura metastases are accompanied by pleural effusion, and these patients are not suitable for surgical resection. Therefore, parietal pleura metastases samples with matched primary tumors are relatively rare, and we were able to recruit only five patients from our hospital over 2 years. Lastly, all the analyses were performed in a single clinical center, and future multicenter investigations are necessary to further confirm the accuracy and robustness of the diagnostic algorithm to fully elucidate its clinical utility.

In conclusion, we have developed a novel NGS algorithm for MLA subtyping, which could accurately and robustly distinguish MPLAs from IPMs, and whose results were also consistent with the clinical expectation in these patients. Our results illustrate that broad-panel NGS could complement traditional CHA diagnosis by reducing the risk of erroneous staging in MLA patients, helping better direct the therapeutic schedule, and improving the accuracy of prognosis prediction.

Acknowledgements

The authors thank Dr. Jun Nakajima (Department of Thoracic Surgery, The University of Tokyo Graduate School of Medicine) for his useful suggestions. This work was supported by Scientific Research of Department of Science and Technology of Liaoning province (grant number: 2017225035), which is used for data collection, analysis, and interpretation. This work is not paid by a pharmaceutical company or other agency.

Author contributions statement

XZ and SX designed the research. XZ, XF, CS, LW, YM, LW and SX were involved in data acquisition and data curation. XZ, XF, CS, LW, YM, LW, YX, PY, XR and XW conducted the data analysis. XZ, XF, CS, YX, LW, YM, LW and SX were involved in the interpretation of the data. All authors wrote and revised the paper. All authors read and approved the final paper.

Ethics approval and consent to participate

This study was approved by the Institutional Research Ethics Committee of The First Hospital of China Medical University (ethical number: AF-SOP-07-1.1-01). All patients signed informed consent forms prior to sample collection and consented to the publication of related clinical information and data. This study was performed in accordance with the Declaration of Helsinki.

Data availability statement

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

References

- Mascalchi M, Comin CE, Bertelli E, *et al.* Screen-detected multiple primary lung cancers in the ITALUNG trial. *J Thorac Dis* 2018; **10**: 1058–1066.
- Amin MB, Greene FL, Edge SB, *et al.* The Eighth Edition AJCC Cancer Staging Manual: continuing to build a bridge from a population-based to a more “personalized” approach to cancer staging. *CA Cancer J Clin* 2017; **67**: 93–99.
- Ettinger DS, Wood DE, Aisner DL, *et al.* Non-small cell lung cancer, version 3.2022, NCCN Clinical Practice Guidelines in Oncology. *J Natl Compr Canc Netw* 2022; **20**: 497–530.
- Martini N, Melamed MR. Multiple primary lung cancers. *J Thorac Cardiovasc Surg* 1975; **70**: 606–612.
- Girard N, Deshpande C, Lau C, *et al.* Comprehensive histologic assessment helps to differentiate multiple lung primary nonsmall cell carcinomas from metastases. *Am J Surg Pathol* 2009; **33**: 1752–1764.
- Travis WD, Brambilla E, Noguchi M, *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**: 244–285.
- Rami-Porta R, Asamura H, Travis WD, *et al.* Lung cancer – major changes in the American Joint Committee on Cancer eighth edition cancer staging manual. *CA Cancer J Clin* 2017; **67**: 138–155.
- Sun W, Liu Y, Liu XY, *et al.* Significance of nonmucinous lepidic component with mild nuclear atypia in the discrimination of multiple primary lung cancers from intrapulmonary metastases. *Int J Clin Exp Pathol* 2014; **7**: 7583–7596.
- Sun W, Feng L, Yang X, *et al.* Clonality assessment of multifocal lung adenocarcinoma by pathology evaluation and molecular analysis. *Hum Pathol* 2018; **81**: 261–271.
- Homer RJ. Pathologists’ staging of multiple foci of lung cancer: poor concordance in absence of dramatic histologic or molecular differences. *Am J Clin Pathol* 2015; **143**: 701–706.
- Fonseca A, Detterbeck FC. How many names for a rose: inconsistent classification of multiple foci of lung cancer due to ambiguous rules. *Lung Cancer* 2014; **85**: 7–11.
- Wang X, Wang M, MacLennan GT, *et al.* Evidence for common clonal origin of multifocal lung cancers. *J Natl Cancer Inst* 2009; **101**: 560–570.
- van Rens MT, Eijken EJ, Elbers JR, *et al.* p53 mutation analysis for definite diagnosis of multiple primary lung carcinoma. *Cancer* 2002; **94**: 188–196.
- Asmar R, Sonett JR, Singh G, *et al.* Use of oncogenic driver mutations in staging of multiple primary lung carcinomas: a single-center experience. *J Thorac Oncol* 2017; **12**: 1524–1535.
- Shen C, Wang X, Tian L, *et al.* Microsatellite alteration in multiple primary lung cancer. *J Thorac Dis* 2014; **6**: 1499–1505.
- Shen C, Wang X, Tian L, *et al.* “Different trend” in multiple primary lung cancer and intrapulmonary metastasis. *Eur J Med Res* 2015; **20**: 17.
- Patel SB, Kadi W, Walts AE, *et al.* Next-generation sequencing: a novel approach to distinguish multifocal primary lung adenocarcinomas from intrapulmonary metastases. *J Mol Diagn* 2017; **19**: 870–880.
- Eguren-Santamaria I, Sanchez-Bayona R, Patino-Garcia A, *et al.* Targeted DNA sequencing for assessing clonality in multiple lung tumors: a new approach to an old dilemma. *Lung Cancer* 2018; **122**: 120–123.
- Peng L, Zeng Z, Teng X, *et al.* Genomic profiling of synchronous triple primary tumors of the lung, thyroid and kidney in a young female patient: a case report. *Oncol Lett* 2018; **16**: 6089–6094.
- Yu F, Tang J, Liu W, *et al.* Genetic profiling of synchronous multiple primary lung carcinomas presenting as ground-glass opacities. *J Clin Oncol* 2019; **37**: e20034.

21. Qiu T, Li W, Zhang F, *et al.* Major challenges in accurate mutation detection of multifocal lung adenocarcinoma by next-generation sequencing. *Cancer Biol Ther* 2020; **21**: 170–177.
22. Chang JC, Alex D, Bott M, *et al.* Comprehensive next-generation sequencing unambiguously distinguishes separate primary lung carcinomas from intrapulmonary metastases: comparison with standard histopathologic approach. *Clin Cancer Res* 2019; **25**: 7113–7125.
23. Detterbeck FC, Franklin WA, Nicholson AG, *et al.* The IASLC Lung Cancer Staging Project: background data and proposed criteria to distinguish separate primary lung cancers from metastatic foci in patients with two lung tumors in the forthcoming eighth edition of the TNM classification for lung cancer. *J Thorac Oncol* 2016; **11**: 651–665.
24. Reiter JG, Makohon-Moore AP, Gerold JM, *et al.* Reconstructing metastatic seeding patterns of human cancers. *Nat Commun* 2017; **8**: 14114.
25. Hu Z, Li Z, Ma Z, *et al.* Multi-cancer analysis of clonality and the timing of systemic spread in paired primary tumors and metastases. *Nat Genet* 2020; **52**: 701–708.
26. Lv J, Zhu D, Wang X, *et al.* The value of prognostic factors for survival in synchronous multifocal lung cancer: a retrospective analysis of 164 patients. *Ann Thorac Surg* 2018; **105**: 930–936.
27. Friedman AA, Letai A, Fisher DE, *et al.* Precision medicine for cancer with next-generation functional diagnostics. *Nat Rev Cancer* 2015; **15**: 747–756.
28. Galuppini F, Dal Pozzo CA, Deckert J, *et al.* Tumor mutation burden: from comprehensive mutational screening to the clinic. *Cancer Cell Int* 2019; **19**: 209.
29. Girard N, Ostrovnya I, Lau C, *et al.* Genomic and mutational profiling to assess clonal relationships between multiple non-small cell lung cancers. *Clin Cancer Res* 2009; **15**: 5184–5190.
30. Arai J, Tsuchiya T, Oikawa M, *et al.* Clinical and molecular analysis of synchronous double lung cancers. *Lung Cancer* 2012; **77**: 281–287.
31. Roepman P, Ten Heuvel A, Scheidel KC, *et al.* Added value of 50-gene panel sequencing to distinguish multiple primary lung cancers from pulmonary metastases: a systematic investigation. *J Mol Diagn* 2018; **20**: 436–445.
32. Liu M, He WX, Song N, *et al.* Discrepancy of epidermal growth factor receptor mutation in lung adenocarcinoma presenting as multiple ground-glass opacities. *Eur J Cardiothorac Surg* 2016; **50**: 909–913.
33. Li W, Cheng N, Zhao Z, *et al.* Molecular characteristics of multifocal esophageal squamous cell carcinomas to discriminate multicentric origin from intramural metastasis. *J Pathol* 2022; **258**: 395–407.
34. Travis WD, Brambilla E, Burke AP, *et al.* Introduction to the 2015 World Health Organization classification of tumors of the lung, pleura, thymus, and heart. *J Thorac Oncol* 2015; **10**: 1240–1242.
35. Kadota K, Nitadori JI, Sima CS, *et al.* Tumor spread through air spaces is an important pattern of invasion and impacts the frequency and location of recurrences after limited resection for small stage I lung adenocarcinomas. *J Thorac Oncol* 2015; **10**: 806–814.
36. Warth A, Muley T, Kossakowski CA, *et al.* Prognostic impact of intra-alveolar tumor spread in pulmonary adenocarcinoma. *Am J Surg Pathol* 2015; **39**: 793–801.
37. Masai K, Sakurai H, Sukeda A, *et al.* Prognostic impact of margin distance and tumor spread through air spaces in limited resection for primary lung cancer. *J Thorac Oncol* 2017; **12**: 1788–1797.
38. Uruga H, Fujii T, Fujimori S, *et al.* Semiquantitative assessment of tumor spread through air spaces (STAS) in early-stage lung adenocarcinomas. *J Thorac Oncol* 2017; **12**: 1046–1051.
39. Li R, Li X, Xue R, *et al.* Early metastasis detected in patients with multifocal pulmonary ground-glass opacities (GGOs). *Thorax* 2018; **73**: 290–292.

SUPPLEMENTARY MATERIAL ONLINE

Figure S1. Flowchart of the study design

Table S1. Demographic characteristics of the patients with unifocal lung adenocarcinoma

Table S2. Jaccard coefficients of the tumor pairs derived from the patients with unifocal lung adenocarcinomas

Table S3. The tumor content of lesions from the positive control and study cohorts

Table S4. Clinicopathological characteristics of the lesions in the study group

Table S5. The gene similarity score and the CHA-re-diagnosed results for the lesion pairs in the study cohort