# Genetic Trajectory and Clonal Evolution of Multiple Primary Lung Cancer with Lymph Node Metastasis

Chunxiang Li ( ✉ lichunxiang@cicams.ac.cn )

National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College

He Tian

Yalong Wang

Zhenlin Yang

Cancer hospital Chinese academy of medical sciences   https://orcid.org/0000-0002-6650-7782

Ping Chen

Tao Fan

National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College   https://orcid.org/0000-0002-6097-2413

Chu Xiao

Guangyu Bai

Lin Li

Bo Zheng

Jie He

# Abstract

Multiple primary lung cancer (MPLC) with lymph node metastasis (LNM) is a rare phenomenon of multifocal lung cancer. The genomic landscapes of MPLC and the clonal evolution pattern between primary lung lesions and lymph node metastasis haven't been fully illustrated. We performed whole-exome sequencing (WES) on 52 FFPE (Formalin-fixed Paraffin-Embedded) samples from 11 patients diagnosed with MPLC with LNM. Genomic profiling and phylogenetic analysis were conducted to infer the evolutional trajectory within each patient. The top 5 most frequently mutated genes in our study were TTN (76.74%), MUC16 (62.79%), MUC19 (55.81%), FRG1 (46.51%), and NBPF20 (46.51%). For 81.82% of patients (9/11), most genetic alterations harbored by pulmonary tumors were mutually exclusive, suggesting their heterogenous origins. Individually, the genetic profile of lymph node metastatic lesions overlapped with that of multiple lung cancers in different degrees but are more genetically related to specific pulmonary lesions. SETD2 was a potential metastasis biomarker of MPLC. The mean putative neo-antigen number of the primary tumor (646.5) is higher than that of lymph node metastases (300, p = 0.2416). Primary lung tumors and lymph node metastases are highly heterogenous in immune repertoires. Our findings portrayed the comprehensive genomic landscape of MPLC with LNM. We characterized the genomic heterogeneity among different tumors and offered novel clues of the clonal evolution between MPLC and their lymphatic metastases, thus advancing the treatment strategies and preventions of MPLC with LNM.

# Introduction

According to the newest data published in 2022(1), the lung cancer death rate is 21% worldwide, ranking the first in both genders among all cancer types. Over 350 people die of lung cancer every day. Lung cancer brings huge burdens to public healthcare annually(2, 3).

In recent decades, multifocal lung cancer has been an increasingly common scenario in clinical practice(4, 5). It is reported that up to 15% of lung cancer patients developed two or more lesions(6), among which multiple primary lung cancer (MPLC) is very common(7). MPLC indicates patients who developed two or more lung tumors that were originally independent, and the multiple pulmonary lesions can be synchronous or metachronous(8, 9). In 2016, the International Association for the Study of Lung Cancer (IALSC) classified multifocal lung cancer into 4 categories: (1) Second primary cancer (2) Separate Tumor Nodules (Intrapulmonary metastasis, IPM) (3) Multifocal Lung Adenocarcinoma with Ground Glass/ Lepidic (GG/L) Features and (4) Diffuse Pneumonic Type(10), among which second primary cancer and multifocal GG/L are commonly acknowledged as MPLC by clinicians. So far, there are no guidelines or widely accepted criteria in the realm of MPC.

Identifying the "dominant" lung lesions that have the most malignant nature is crucial for MPLC patients, many of whom are not qualified for radical surgical resection because of operational contraindication, such as advanced age or prolonged anesthesia risk. Pre-locating the key tumor before surgery can provide those patients with more treatment options. From the view of evolution, MPLC with lymph node metastasis (LNM) is an ideal material for studying the genetic trajectory of MPLC, thus identifying the "dominant" lung lesions. However, there are several barriers to the study of MPLC with LNM. Firstly, precise discrimination between MPLC and lung cancer intrapulmonary metastasis (IPM) has been a clinical dilemma for decades(6, 11). Both two diseases behave as multiple tumor sites in the lung, but they are of distinct staging strategies [8], treatment [9–11], and prognosis [8, 12, 13]. Secondly, MPLC with LNM is very rare in clinical practice. Thirdly, the difficulties in sample collection of MPLC with LNM hamper its studies.

What we know about MPLC with LNM is far from enough, though some researchers paid attention to it. In 2016, Gao et al. focused on an MPLC patient with one lymph node metastasis and applied targeted panel sequencing on him(12). The lymph node metastasis shared 52 common mutations with one of the pulmonary tumors but had no overlap with other lung lesions within the individual. A similar conclusion was observed in Omada et al.'s research in 2020(13), in which five MPLC with LNM patients were included.

Our study performed WES on tumor samples from 11 patients diagnosed as MPLC with LNM. We are the first to characterize the clonal evolution pattern of MPLC with lymph node metastasis using comprehensive genetic sequencing. Our findings may help grasp the mysterious nature of MPLC and identify the tumor lesion that is most likely to metastasis before surgery, thus may spur the introduction of MPLC guidelines.

We present the following article in accordance with the MDAR reporting checklist.

# Methods And Materials

**Study design and patients.** According to the 8th edition of the TNM staging system published in 2017, a total of 1458 patients during 2011 and 2019 who were diagnosed with MPLC receiving no adjuvant treatment before surgery were identified in Cancer Hospital, the Chinese Academy of Medical Sciences, and Peking Union Medical College. Under the following criteria: (1). Positive lymph node metastasis; (2). Complete access to FFPE specimens of pulmonary tumors, lymph node metastasis tumors, and matched normal samples, 11 patients were finally included in our study. Two independent certified pathologists confirmed the diagnosis by HE staining and provided the pathological details of each patient (Table 1). The clinical information was retrieved from the medical record system of our hospital (Table 2). CT (Computerized Tomography) images were obtained from the department of radiology of our hospital and two experienced radiological specialists annotated the pulmonary tumor lesions independently. The ethics committee of Cancer Hospital, Chinese Academy of Medical Science approved the study and written informed consent was obtained from each involved patient. The study was performed in accordance with the local ethical regulations and the guidelines of the Declaration of Helsinki.

**WES Sequencing.** We extracted total DNA from archived FFPE samples using QIAamp DNA FFPE Tissue Kit (Qiagen, cat. no. 56404). To enhance the tumor purity, tumor areas on FFPE slides were marked by two independent pulmonary pathologists and the non-tumor components were scraped off before DNA extraction. DNA from paired normal lung tissue samples was used to eliminate the influence of germline mutation. Isolated genomic DNA quality was verified using three combined methods: (1) DNA degradation, and contamination was monitored on 1% agarose gels. (2) DNA purity was checked using the NanoPhotometer® spectrophotometer (IMPLEN, CA, USA). (3) DNA concentration was measured by Qubit® DNA Assay Kit in Qubit® 2.0 Fluorometer

(Invitrogen, USA). WES libraries were prepared using Agilent SureSelect Human All Exon V6 kit (Agilent Technologies, CA, USA). Following the manufacturer's recommendations, index codes were added to each sample, and the clustering of the index-coded samples was performed on a cBot Cluster Generation System using Hiseq PE Cluster Kit (Illumina). DNA libraries were sequenced on the Illumina Hiseq platform with a paired-end 2×150 protocol. The WES was conducted at CapitalBio Technology Inc. Beijing, China, from January 2020.01 to October 2020.

### Data Processing and Bioinformatics.

Sequence Alignment, and Variant Calling. We trimmed and filtered raw data using Trimmomatic 0.33(14). Paired-end clean reads were aligned to the human reference sequence hg19 using the BWA-MEM algorithm (BWA version 0.7.10-r789) with default parameters(15). To guarantee the meaningful downstream analysis, duplicated sequencing reads were excluded by Picard v.2.13. In contrast, low confidence reads (reads containing adapter contamination, low-quality nucleotides, and unrecognizable nucleotide (N)) were removed by the criteria of TLOD < 10. All high-confident mutations were then annotated into the MAF format using the tool vcf2maf. We called somatic single-nucleotide variants (SNVs) and small indels using MuTect (version 3.1-0-g72492bb) and Strelka (version 1.0.14)(16, 17). All mutations in coding regions were manually checked using the Integrative Genomics Viewer (version2.3.34)(18).

Definitions of putative driver mutations. We compared all non-silent mutations with lists of lung cancer/pan-cancer potential driver genes in the COSMIC cancer gene census (September 2021) and identified the driver mutations in our data if they are in the lists.

Mutation spectra and signature analyses. We extracted the 5' and 3' sequence context of each mutation from the hg19 reference genome and categorized the SNVs into C > A, C > G, C > T, T > A, T > C, and T > G bins according to the type of substitution and then subcategorized into 96 sub-bins according to the nucleotides preceding (5') and succeeding (3') the mutated base. The deconstructSigs package (v1.8.0) was used to infer the contributions of 30 published signatures from the Catalog of Somatic Mutations in Cancer (COSMIC) (https://cancer.sanger.ac.uk/cosmic/signatures_v2) in each sample(19). The contribution of each signature for each tumor was statistically quantified.

Phylogenetic Tree Construction and Labeling. All nonsynonymous somatic mutations were considered for the purpose of determining phylogenetic trees. We built trees using binary presence/absence matrices generated from the distribution of variants within different tumors within each patient. R Bioconductor package phangorn(20) was used to perform the parsimony ratchet method, generating unrooted trees. Branch and trunk lengths were proportional to the number of nonsynonymous mutated genes. To label the clonal phylogenetic trees for each patient, we defined 5 categories of mutations according to the driver genes lists in the COSMIC cancer gene census (September 2021): (1) Tier 1 NSCLC/lung cancer driver genes. (2) Tier 2 NSCLC/lung cancer driver genes. (3) Tier 1 pan-cancer driver genes. (4) Tier 2 pan-cancer driver genes. (5) Genes not in the list.

Unsupervised Clustering. The hclust function (the agglomeration method is "ward.D2") in R software (Version 4.0.2) was utilized to perform unsupervised clustering. According to the driver gene list from Catalog of Somatic Mutations in Cancer (COSMIC) (https://cancer.sanger.ac.uk/cosmic/signatures_v2), totally 40 genes were utilized, including 20 genes with top mutation frequencies in our data, 10 pan-cancer driver genes with top mutation frequencies except for the top 20 genes, and 10 lung cancer driver genes with top mutation frequencies except for the top 20 genes

Functional enrichment analysis. We applied the 468 genes from the MSK-IMPACT assay to our cohort and conducted GO functional enrichment analysis and showed the key genes.

Venn Diagram and Upset graph. Venn diagrams and Upset graphs were used to illustrate the mutational overlaps among multiple samples within individuals. For patients who have no more than 5 tumor samples, we drew the Venn diagrams using Omicshare (an online tool, https://www.omicshare.com/tools/Home/Soft/venn). For Patient 1 and Patient 11, each of whom has 6 tumor samples, the R software package UpSetR was applied to draw Upset graphs.

Putative Neoantigens Identification and Binding Affinity Prediction. Polysolver algorithm(21) was applied to conduct HLA typing. Non-silent mutations were used to generate a list of mutant peptides of approximately 9−11 amino acids in length with the mutated residues represented in each position.

NetMHCpan (v3.0) was used to predict the binding affinity of each mutant peptide and its corresponding wild-type peptide to the patient's germline HLA alleles(22). Candidate neoantigens were identified as those with a predicted mutant peptide binding affinity of < 500 nmol/L.

# Statistical Analysis.

Statistical analysis was performed using Fisher's exact test for categorical variables and the Wilcoxon test or Mann−Whitney U test for continuous variables. We estimated the mutual exclusivity of mutations by Monte Carlo simulation. We calculated the relevance between constant variables using Pearson's correlation or Spearman's correlation. Kaplan−Meier curves and the log-rank tests were used for the survival analysis. Statistical tests were performed in R (v3.2.0) and GraphPad (v8) We regarded a two-tailed P value < 0.05 as statistically significant.

# Result

Patient cohort. The workflow was presented in Fig. 1A. Initially, 1458 patients were selected, based on whom 1421 patients were excluded for they lacked lymph node metastasis, and 26 patients were excluded for their FFPE samples were not available. Finally,11 MPLC with LNM patients were recruited. We explored these patients from clinicopathological information, hematoxylin-eosin (HE) staining, WES sequencing, and radiological images. The ideograph of all 11 patients was displayed in Fig. 1B. Representative HE images of Patients 1 and 11 were shown in Fig. 1C (Supplementary Fig. 1).

**Clinicopathological and Demographic information.** The clinicopathological details and WES depth were in **Table 1**. We defined the pathological subtype with the highest proportion as the major pathological component. In 72.73% (8/11) patients, the multiple lung tumors were of heterogenous major pathological components, while in Patient1, 5, and 11, all their pulmonary tumors were solid subtypes. In 81.2% (9/11) patients, we captured pathological consistency between primary tumors and metastases, but in Patient7 and 10, their lymph node metastasis sites were heterogenous to primary tumors pathologically. Stage annotations were based on the proposed TNM classification criteria of lung cancers with multiple pulmonary sites by IASLC in 2016[6]. Most lung tumors were staged as T1 and T2 and most lymph node metastatic lesions were staged as N1 or N2. The average sequencing depth of tumor and normal samples were 275x (range 69x-435x) and 332x (range 63x-487x), respectively. **Table 2** exhibits demographic information. According to the proposed differentiation criteria of lung cancer with multiple pulmonary sites of involvement by IASCL in 2016(10), all patients in our study were MPLC. The average age of our cohort was 59.6 years old (range 42−76). Totally 27.27% (3/11) patients were with smoking history, and 27.27% (3/11) patients were alcohol users. Over one-third of patients (36.36%, 4/11) patients had a positive family history, especially Patient 9, who had a familial history of both lung cancer and prostate cancer. All patients were synchronous MPLC, except Patient6 (metachronous MPLC). The post-surgery recurrence/metastasis rate was 55.56% (5/9) patients. The average follow-up time of all patients is 61.1 months.

**Genomic Alterations.** We successfully conducted WES on 52 DNA samples collected from 11 patients to assess genomic alterations globally. Sequencing quality information of all the samples was in Supplementary Data 1, and the list of somatic nonsynonymous mutations was in Supplementary Data 2. The alteration spectrum is Fig. 2A (Supplementary Data 3). According to the driver gene list from COSMIC Cancer Gene Census (Sep. 2021, https://cancer.sanger.ac.uk/census, Supplementary Data 4), we marked the pan-cancer driver genes (light brown) and lung cancer/NSCLC driver genes (green). The mean tumor mutation burden (TMB) for pulmonary tumors and lymph node metastases was 12.78 and 8.94 mutations per megabase, respectively.

Among the top 20 genes with the highest alteration rates, 30% (6/20) were pan-cancer driver genes. TTN (Titin, 76.74%), MUC16 (Mucin16, 62.79%), and MUC19 (Mucin19, 55.81%) were the 3 genes with the highest alteration rates in our cohort. Pan-cancer driver genes with high alteration rates include FLNA (Filamin A, 41.86%), MUC4 (Mucin 4, 41.86%), and FAT3 (FAT Atypical Cadherin 3, 39.53%). Lung cancer/NSCLC driver genes with high alteration rates include BIRC6 (Baculoviral IAP Repeat Containing 6, 30.23%), EGFR (Epidermal Growth Factor Receptor, 30.23%), and TP53 (Tumor Protein P53, 20.93%). SETD2 (SET Domain Containing 2, Histone Lysine Methyltransferase) is the histone H3 lysine 36 histone (H3K36) methyltransferase(23) and has been reported in various solid tumors and blood malignancies(24−26). In our cohort, SETD2 alternated in all 6 tumor samples (2 primary tumors and 4 lymph node metastasis) of Patient1, including 5 splice site mutations and 1 missense mutation. Other genes alternating in both primary tumors and lymph node metastasis tumors within individuals include CALR (Calreticulin), FRG1(FSHD Region Gene 1), and ACTG1 (Actin Gamma 1).

The single-nucleotide variations (SNVs) displayed considerable variations across and within patients, indicating intratumor heterogeneity (Fig. 2B, Supplementary Data 5). For Patient4, 9, and 10, who were with smoking history, the majority of their sequenced samples displayed a preponderance of C > A transitions, which is associated with tobacco exposure(27). In Patient9, the SNV pattern of Patient9LN1 is highly similar to that of Patient9T1 rather than Patient 9T2, implying the metastasis dominance of Patient 9T1. The contributions of various known signatures to each sample are demonstrated in Fig. 2C (Supplementary Data 6). Signature 4 was prevalent in Patient4, 9, and 10, consistent with the fact that they were smokers. However, signature 4 was also observed in non-smoker patients, and signature 29 (representing tobacco chewing) was prevalent in a large proportion of samples, suggesting that tobacco exposure may play a critical role in the pathogenesis of MPLC with LNM. Signature 3 was identified in 93.20% (40/43) of tumor samples, indicating that DNA mismatch repair was highly involved in the etiology of MPLC with LNM. The signatures echoes between primary tumors and lymph node metastasis tumors offer clues for identifying the dominant primary tumor. For example, in Patient1, signature 22 only appeared in T1 and LN1, suggesting the unique connection between them. In Patient9, the signatures contribution of LN1 was almost identical to that of T1, implying T1 might be the source of LN1. Based on the MSK-IMPACT 468 panel genes (Supplementary Data 7), we acquired 315 genes from our data and conducted functional enrichment analysis (Fig. 2D, Supplementary Data 8). Several signaling pathways were involved, including histone modification, chromatin remolding, MAPK, and cell cycle.

Overall, the mutation spectrum of MPLC with lymph node metastasis was disconcordant with that of typical lung adenocarcinoma, indicating the unique genomic contexture of this disease. Meanwhile, primary tumor and lymph node metastasis echo in both SNVs and COSMIC signature levels, showing the potentiality of clarifying metastasis trajectory and identifying dominant primary tumor.

**Genomic Heterogeneity.** The Upset map and Venn diagram show the mutation repertoire overlap within representative individuals (Fig. 3A, Supplementary Data 9, Supplementary Fig. 2, 3.). The mean number of genes shared by all samples within individuals was 3.67 (range 1−10) per patient. FRG1B (LOC102724813) mutation was shared by all tumors of Patient1, 2, 6, and 9. Both SETD2 p.X1529_splice and STAG2 (Stromal Antigen 2) p.D1014E were harbored by Patient1T2, Patient1LN2, and Patient1LN4. SETD2 is a histone lysine methyltransferase playing a significant role in renal malignancies(28), prostate cancer(24), and NSCLC(29). STAG2 is proved to be involved with viral infection via STING signaling(30). FRG1B (LOC102724813) mutation was shared by all tumors of Patient1, 2, 6, and 9. CEP192 p.L1701F, a gene involved with PLK1 activity regulation at G2/M Transition, mitotic centrosome maturation, and bipolar assembly(31, 32), was the only mutation common to all 6 tumors of Patient11.

Mutation distribution analysis reveals that MPLC with lymph node metastasis is of high intratumor heterogeneity and the genomic complexity of pulmonary tumors and lymph node metastasis tumors is hard to gauge (Fig. 3B, Supplementary Data 10). In all patients, the mutations private to only one tumor sample account for the largest proportion (ratio range 66.40%-99.72%, median 96.08%; number range 518 to 6721, median 1596), indicating enormous heterogeneity among tumors. For Patient5 and Patient7, the percentages of mutations shared by all tumors are higher than other patients (Patient5, 19.74%, Patient7, 3.74%), we believe it results from that the two patients lack matched normal samples. Generally, pulmonary tumors take larger percentages of private mutations than lymph node metastasis tumors (pulmonary, range 37.92%-96.92%, median 53.55%; lymph node metastasis, range 0%-59.48%, median 28.66%. p = 0.054), emphasizing the significance of primary tumor management. However, in Patient2, 7, 10, and 11, lymph node metastasis tumors claim more private mutations than lung tumors, suggesting that lymph node metastasis tumors may be more complex in the genomic background. We visualized the intratumor

heterogeneity by global unsupervised clustering analysis (Fig. 3C, Supplementary Data 11). Patient11 was the only patient whose tumors were closely clustered, except Patient5 and 7, who lack paired normal samples. Pulmonary tumors and lymph node metastasis tumors from distinct individuals tended to be clustered together, respectively, implying genomic differences between primary lung tumors and lymph node metastasis sites.

**Clonal Architecture and Evolutionary Trajectory.** Filtered nonsynonymous variances from each tumor sample were used to construct phylogenetic trees using the parsimony ratchet method(33). The branch lengths were proportional to the number of nonsilent alterations within individuals (the scales were different among patients for better visual effect) and the driver genes were annotated next to the branches. Individual CT (computerized tomography) images before surgery were provided to validate the multiple pulmonary tumors. Heatmaps were used to show mutation distribution (Fig. 4, Supplementary Data 12, Supplementary Fig. 4.). Based on the driver gene list from COSMIC Cancer Gene Census (Sep. 2021, https://cancer.sanger.ac.uk/census), we annotated the driver genes to the branches.

We observed diverse evolutionary patterns between pulmonary primary tumors and lymph node metastases. Firstly, in Patient1, all 4 lymph node tumors were earlier in evolution than the two pulmonary tumors. The same phenomenon was found in Patient11LN1, indicating that lymph node metastasis could be an early event in the pathogenesis of MPLC. Secondly, for Patient4, 8, and 9, lymph node metastasis lesions were more closely related to the specific individual pulmonary tumor, suggesting these pulmonary tumors were more aggressive in nature. Thirdly, in Patient2 and 11, lymph node metastases were associated with more than one pulmonary tumor, indicating that multiple pulmonary lesions of MPLC could contribute to metastasis within individuals, for whom radical resection is necessary. Moreover, for Patient3, 5, 7, and 10, their results were consistent with Patient4, 8, and9, but needed further exploration since each of them had only one pulmonary tumor available. The three lung tumors of Patient6 were highly heterogeneous in genetic background, suggesting the intricate nature of MPLC.

Generally, the diverse evolution patterns suggested the complex nature of MPLC with LNM.

# Immune Repertoires of MPLC with LNM

Currently, the major management for MPLC is surgical resection, and the clues of chemotherapy or immunotherapy are rare. To offer new insights for the immunotherapy administration in MPLC, we performed neoantigen number prediction, neoantigen binding affinity prediction, and clone cluster calculation in Fig. 5. Overall, multiple tumors in MPLC showed extremely high heterogeneity in the immune background, which is more complicated than the mutational spectrum. Both primary pulmonary tumors and lymph node metastasis tumors could have diverse clone clusters. In Fig. 5A (Supplementary Data 13), we showed the predicted neoantigen number of each sample. For key patients (Patient 1, 2, 4, 8, 9, and 11), the mean predicted neoantigen number of primary lung cancers (mean 646.5, range 59-5186) was higher than that of lymph node metastases (mean 300, range 45-4562), but this discrepancy was not significant statistically (p = 0.2416). For all 11 patients, neoantigens private to only one tumor sample possessed absolute advantage (range 71.40%-100%, mean 99.36%). In Patient3 and 6, all their predicted neoantigens were private to one tumor (Fig. 5B Left, Supplementary Data 14). Neoantigens of primary lung tumors (range 35.07%-76.02%, mean 52.72%) hold similar proportions to that of lymph node metastases (range 2.38%-64.70%, mean 46.80%), while the two parts shared few neoantigens (range 0-28.60%, mean 0.27%), suggesting the two parts might response to immunotherapy in different manners (Fig. 5B Right, Supplementary Data 14). Binding affinity analysis shows that primary lung tumors and lymph node metastases harbor distinct neoantigens repertoires (Fig. 5C, Supplementary Data 15), implying heterogeneous tumor microenvironments and potentially various reactions to immune checkpoint blockers. The majority of sequenced samples were oligoclonal and the clonal structures of the primary tumor and metastases were in diverse patterns (Fig. 5D, Supplementary Data 16).

# Discussion

In our study, the mutation landscape of MPLC with LNM is quite different from that of east Asian LUAD patients(34) in our study, suggesting the unique nature of this disease and the necessity of further research. The most frequently mutated oncogenes in our study were MUC16, FLNA, MUC4, FAT3, SETD2, and CALR, all of which were pan-cancer driver genes, and no highly mutated lung cancer driver gene ranked in the top 20 mutations, such as EGFR, KRAS, and TP53. This deviation might result from the insufficient sequencing depth of a few samples in our study, but we have reason to infer that such inconsistencies indicate MPLC with LNM may have a unique mutation spectrum, which needs further validation in larger cohorts. SETD2 (SET Domain Containing 2) mutated in all 6 tumors of Patient1 (five of them were splice-site mutations), suggesting it may be the driving force of lymph node metastasis in MPLC. It is a histone lysine methyltransferase, closely involved with cancer behavior(35, 36), and recurrently mutated in several cancer types(24, 37). Moreover, SETD2 mutated in one lung tumor of an MPLC with LNM patient in a previous study(13), which was partly consistent with our results. TTN (Titin), the gene with the highest alteration rate in our cohort, codes a sarcomere protein which is a major player in cardiomyocyte stiffness and cardiac train sensing(38). In lung cancer, several recent studies reported that TTN mutations enhanced anti-tumor immunity and were associated with favorable responses to ICI administration(39−41). Therefore, we believe that MPLC and MPLC with LNM might be ideal substrates for ICI.

Our study imposed further deliberation on the current MPLC diagnosis standard. In the clinical practice of multifocal lung cancer, precise differentiation between MPLC and IPM has been a dilemma for decades(42), rendering treatment selection challenging [9−11]. Many previous studies have paid attention to this field(11, 12, 43−45) (6, 46) (6, 13, 46−48), who believed that at most 1 mutation could be shared by any two lung tumors of MPLC and tumors sharing multiple (≥ 2) mutations should be defined as IPM(11). However, in our study, the shared mutations of any two pulmonary tumors within an individual ranged from 1 to 78 (median: 20, Patient3, 5, 7, and 10 were excluded for they only had one valid lung tumor), disconcordant with previous studies, and very few overlap genes were oncogenes. This inconsistency could result from the sequencing strategies used by previous studies, most of them chose targeted sequencing which only detected cancer-related genes (11, 49, 50), while WES/WGS was applied to only a few patients(12, 43). Our research indicates that multiple pulmonary tumors of MPLC could be more like each other in the genomic contexture than what we know before and it might be arbitrary to distinguish MPLC and IPM under the current "1 mutation" criteria. Moreover, consistent with Qu et al.'s work(51), our study suggests that the absence of LNM

may not be a necessary criterion for the diagnosis of MPLC with similar tumor pathology. MPLC with LNM should be carefully diagnosed to guarantee that the patients do not miss the best treatment opportunity.

We offered novel clues for identifying the "dominant" lung tumors in MPLC and challenged the current recognitions about MPLC. At present, the clonal evolution relationships among different tumor lesions of MPLC with lymph node metastasis are unclear but of great significance, since we may catch the driving force dictating the metastasis progression, identify the tumor site that has the most potential to metastasis so that doctors can react in advance, and get a thorough understanding of the biological behavior of MPLC. However, these works are hampered by the enormous genetic heterogeneity of MPLC. While in MPLC with LNM, the genetic trajectory between pulmonary primary tumor and lymph node metastasis is an ideal material to illustrate this dilemma. In 2016, Liu et al.(12) firstly sequenced one MPLC with an LNM patient, who was adenocarcinoma. In 2020, Higuchi et al.(13) performed targeted sequencing on 5 MPLC with LNM patients, whose pulmonary tumors were all LUSC in histology. These two studies indicated that lymph node metastasis was genetically related to one specific lung lesion, suggesting metastasis dominance. In our study, we drew phylogenetic trees for each patient and inferred the evolutional process of multiple tumors, based on the theory of clonal heterogeneity and tumor evolution(52). We found the lymph node metastases could originate from both one and multiple pulmonary tumors within the individual, indicating the "dominant" tumors could be more than one. In addition, our results challenge the current understanding that MPLC is early in staging since some lymph node metastases are evolutionally earlier than pulmonary tumors.

For clinical doctors, we provided new choices for MPLC management by advancing dominant tumors identification. Currently, the most prevalent treatment for MPLC is surgical resection(53, 54), although many therapies are feasible in MPLC, including photodynamic therapy(55), stereotactic ablative radiotherapy(56), and combination of local radiofrequency ablation and melatonin(57). When the multiple tumors of MPLC are in different lobes or lungs, multiple operations become necessary, which is riskier than one-time resection. However, many MPLC patients cannot endure highly radical resections because of surgical contraindications. Therefore, if we can pre-locate and remove the dominant tumor, more patients will benefit from surgery in a safer manner. In addition, recurrence is very common for multiple lung cancer in clinical practice, the underlying reasons of which are not clear, while dominant tumor resection can be taken as a key indicator for recurrence prevention.

We offered novel insights for the immunotherapy administration in MPLC and MPLC with LNM. In the recent decade, immune checkpoint inhibitors achieved considerable success in multiple malignancies(58), including NSCLC(59) (Non-Small Cell Lung Cancer). However, MPLC and MPLC with LNM are seldom involved. In Fig. 5, we observed that the multiple tumors within individuals were highly heterogeneous in neoantigen, while disparities exist between primary tumors and metastases. It is reasonable to infer that MPLC and MPLC with LNM may be ideal audiences of ICIs (immune checkpoint inhibitors) and different strategies should be applied to primary and metastases samples.

There are mainly three obstacles in unveiling the nature of MPLC and MPLC with LNM: Firstly, the lack of guidelines or generally accepted definition of MPLC hindered the integration of data generated by different researchers who might adopt other criteria. Secondly, it is not easy to collect MPLC samples in surgery since some tumor nodules are too small to resect, and the multiple nodules might locate on a different side of the lungs. In contrast, only one side of the lung will be touched in one surgery. Thirdly, WES or WGS (whole-genome sequencing) still cannot be afforded by most scientific institutes, so most researchers used target sequencing, which could not give a profound coverage of the genomic contexture of MPLC. In our study, the incidence of MPLC with LNM was 2.54% (37/1458), and we estimate that the incidence is higher in the real world. Therefore, it is of great significance to study this malignancy regardless of several obstacles.

Several limitations existed in our study. Firstly, a few samples failed in the WES quality control, causing imperfection of some patients. Secondly, the sequencing depth of many samples is insufficient for producing highly valid results, many key mutations were missing. Thirdly, we only performed WES, without transcriptional data or information from other omics.

In conclusion, by characterizing the molecular features of MPLC with LNM, we identified potential driver genes of this disease and revealed various evolution patterns among pulmonary primary tumors and lymph node metastasis tumors. We found that lymph node metastasis might be an early event in the etiology of MPLC. Our findings push the boundaries of both MPLC and MPLC with LNM by providing new information for dominant tumor identification. Further solid evidence of MPLC and MPLC with LNM is warranted.

# Declarations

## Author Contribution Statement

H. T., C.L., and J.H. are responsible for the study design.

J.H., C.L. B.Z., and L.L are responsible for conception identification.

H.T., Y.W., and Z.Y. are responsible for the acquisition of data.

H.T., Z.Y., and P.C. are responsible for the analysis and interpretation of data.

H.T. C.L., G.B., and T.F., C.X. are responsible for drafting the manuscript.

H.T. and T.F. are responsible for revising the article critically for important intellectual content.

# References

1. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. CA Cancer J Clin. 2022;72(1):7–33.

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(10066):299–311.

3. Herbst RS, Morgensztern D, Boshoff C. The biology and management of non-small cell lung cancer. Nature. 2018;553(7689):446–54.

4. Leventakos K, Peikert T, Midthun DE, Molina JR, Blackmon S, Nichols FC, et al. Management of Multifocal Lung Cancer: Results of a Survey. J Thorac Oncol. 2017;12(9):1398–402.

5. Zheng R, Shen Q, Mardekian S, Solomides C, Wang ZX, Evans NR, 3rd. Molecular profiling of key driver genes improves staging accuracy in multifocal non-small cell lung cancer. J Thorac Cardiovasc Surg. 2020;160(2):e71-e9.

6. Murphy SJ, Harris FR, Kosari F, Barreto Siqueira Parrilha Terra S, Nasir A, Johnson SH, et al. Using Genomics to Differentiate Multiple Primaries From Metastatic Lung Cancer. J Thorac Oncol. 2019;14(9):1567–82.

7. Jiang L, He J, Shi X, Shen J, Liang W, Yang C, et al. Prognosis of synchronous and metachronous multiple primary lung cancers: systematic review and meta-analysis. Lung Cancer. 2015;87(3):303–10.

8. Zhao L, Liu C, Xie G, Wu F, Hu C. Multiple Primary Lung Cancers: A New Challenge in the Era of Precision Medicine. Cancer Manag Res. 2020;12:10361–74.

9. Wang Y, Yeung JC, Hanna WC, Allison F, Paul NS, Waddell TK, et al. Metachronous or synchronous primary lung cancer in the era of computed tomography surveillance. J Thorac Cardiovasc Surg. 2019;157(3):1196–202.

10. Detterbeck FC, Nicholson AG, Franklin WA, Marom EM, Travis WD, Girard N, et al. The IASLC Lung Cancer Staging Project: Summary of Proposals for Revisions of the Classification of Lung Cancers with Multiple Pulmonary Sites of Involvement in the Forthcoming Eighth Edition of the TNM Classification. J Thorac Oncol. 2016;11(5):639–50.

11. Chang JC, Alex D, Bott M, Tan KS, Seshan V, Golden A, 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(23):7113–25.

12. Liu Y, Zhang J, Li L, Yin G, Zhang J, Zheng S, et al. Genomic heterogeneity of multiple synchronous lung cancer. Nat Commun. 2016;7:13200.

13. Higuchi R, Nakagomi T, Goto T, Hirotsu Y, Shikata D, Yokoyama Y, et al. Identification of Clonality through Genomic Profile Analysis in Multiple Lung Cancers. J Clin Med. 2020;9(2).

14. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114–20.

15. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 2010;26(5):589–95.

16. 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(3):213–9.

17. Saunders CT, Wong WS, Swamy S, Becq J, Murray LJ, Cheetham RK. Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. Bioinformatics. 2012;28(14):1811–7.

18. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. 2013;14(2):178–92.

19. Rosenthal R, McGranahan N, Herrero J, Taylor BS, Swanton C. DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. Genome Biol. 2016;17:31.

20. Schliep KP. phangorn: phylogenetic analysis in R. Bioinformatics. 2011;27(4):592–3.

21. Shukla SA, Rooney MS, Rajasagi M, Tiao G, Dixon PM, Lawrence MS, et al. Comprehensive analysis of cancer-associated somatic mutations in class I HLA genes. Nat Biotechnol. 2015;33(11):1152–8.

22. Nielsen M, Andreatta M. NetMHCpan-3.0; improved prediction of binding to MHC class I molecules integrating information from multiple receptor and peptide length datasets. Genome Med. 2016;8(1):33.

23. Chen R, Zhao WQ, Fang C, Yang X, Ji M. Histone methyltransferase SETD2: a potential tumor suppressor in solid cancers. J Cancer. 2020;11(11):3349–56.

24. Yuan H, Han Y, Wang X, Li N, Liu Q, Yin Y, et al. SETD2 Restricts Prostate Cancer Metastasis by Integrating EZH2 and AMPK Signaling Pathways. Cancer Cell. 2020;38(3):350 − 65.e7.

25. Mar BG, Chu SH, Kahn JD, Krivtsov AV, Koche R, Castellano CA, et al. SETD2 alterations impair DNA damage recognition and lead to resistance to chemotherapy in leukemia. Blood. 2017;130(24):2631–41.

26. Huang KK, McPherson JR, Tay ST, Das K, Tan IB, Ng CC, et al. SETD2 histone modifier loss in aggressive GI stromal tumours. Gut. 2016;65(12):1960–72.

27. Swanton C, Govindan R. Clinical Implications of Genomic Discoveries in Lung Cancer. N Engl J Med. 2016;374(19):1864–73.

28. Gerlinger M, Rowan AJ, Horswell S, Math M, Larkin J, Endesfelder D, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. N Engl J Med. 2012;366(10):883–92.

29. Kim IK, McCutcheon JN, Rao G, Liu SV, Pommier Y, Skrzypski M, et al. Acquired SETD2 mutation and impaired CREB1 activation confer cisplatin resistance in metastatic non-small cell lung cancer. Oncogene. 2019;38(2):180–93.

30. Ding S, Diep J, Feng N, Ren L, Li B, Ooi YS, et al. STAG2 deficiency induces interferon responses via cGAS-STING pathway and restricts virus infection. Nat Commun. 2018;9(1):1485.

31. Joukov V, Walter JC, De Nicolo A. The Cep192-organized aurora A-Plk1 cascade is essential for centrosome cycle and bipolar spindle assembly. Mol Cell. 2014;55(4):578–91.

32. Chinen T, Yamazaki K, Hashimoto K, Fujii K, Watanabe K, Takeda Y, et al. Centriole and PCM cooperatively recruit CEP192 to spindle poles to promote bipolar spindle assembly. J Cell Biol. 2021;220(2).

33. Murugaesu N, Wilson GA, Birkbak NJ, Watkins T, McGranahan N, Kumar S, et al. Tracking the genomic evolution of esophageal adenocarcinoma through neoadjuvant chemotherapy. Cancer Discov. 2015;5(8):821–31.

34. Chen J, Yang H, Teo ASM, Amer LB, Sherbaf FG, Tan CQ, et al. Genomic landscape of lung adenocarcinoma in East Asians. Nat Genet. 2020;52(2):177–86.

35. Niu N, Lu P, Yang Y, He R, Zhang L, Shi J, et al. Loss of Setd2 promotes Kras-induced acinar-to-ductal metaplasia and epithelia-mesenchymal transition during pancreatic carcinogenesis. Gut. 2020;69(4):715–26.

36. Chen BY, Song J, Hu CL, Chen SB, Zhang Q, Xu CH, et al. SETD2 deficiency accelerates MDS-associated leukemogenesis via S100a9 in NHD13 mice and predicts poor prognosis in MDS. Blood. 2020;135(25):2271–85.

37. Comprehensive molecular characterization of clear cell renal cell carcinoma. Nature. 2013;499(7456):43–9.

38. Loescher CM, Hobbach AJ, Linke WA. Titin (TTN): from molecule to modifications, mechanics and medical significance. Cardiovasc Res. 2021.

39. Su C, Wang X, Zhou J, Zhao J, Zhou F, Zhao G, et al. Titin mutation in circulatory tumor DNA is associated with efficacy to immune checkpoint blockade in advanced non-small cell lung cancer. Transl Lung Cancer Res. 2021;10(3):1256–65.

40. Wang Z, Wang C, Lin S, Yu X. Effect of TTN Mutations on Immune Microenvironment and Efficacy of Immunotherapy in Lung Adenocarcinoma Patients. Front Oncol. 2021;11:725292.

41. Xie X, Tang Y, Sheng J, Shu P, Zhu X, Cai X, et al. Titin Mutation Is Associated With Tumor Mutation Burden and Promotes Antitumor Immunity in Lung Squamous Cell Carcinoma. Front Cell Dev Biol. 2021;9:761758.

42. Mitsudomi T, Yatabe Y, Koshikawa T, Hatooka S, Shinoda M, Suyama M, et al. Mutations of the P53 tumor suppressor gene as clonal marker for multiple primary lung cancers. J Thorac Cardiovasc Surg. 1997;114(3):354–60.

43. Ma P, Fu Y, Cai MC, Yan Y, Jing Y, Zhang S, et al. Simultaneous evolutionary expansion and constraint of genomic heterogeneity in multifocal lung cancer. Nat Commun. 2017;8(1):823.

44. Roepman P, Ten Heuvel A, Scheidel KC, Sprong T, Heideman DAM, Seldenrijk KA, 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(4):436–45.

45. Takahashi Y, Shien K, Tomida S, Oda S, Matsubara T, Sato H, et al. Comparative mutational evaluation of multiple lung cancers by multiplex oncogene mutation analysis. Cancer Sci. 2018;109(11):3634–42.

46. Murphy SJ, Aubry MC, Harris FR, Halling GC, Johnson SH, Terra S, et al. Identification of independent primary tumors and intrapulmonary metastases using DNA rearrangements in non-small-cell lung cancer. J Clin Oncol. 2014;32(36):4050–8.

47. Sozzi G, Miozzo M, Pastorino U, Pilotti S, Donghi R, Giarola M, et al. Genetic evidence for an independent origin of multiple preneoplastic and neoplastic lung lesions. Cancer Res. 1995;55(1):135–40.

48. Takamochi K, Oh S, Matsuoka J, Suzuki K. Clonality status of multifocal lung adenocarcinomas based on the mutation patterns of EGFR and K-ras. Lung Cancer. 2012;75(3):313–20.

49. Pei G, Li M, Min X, Liu Q, Li D, Yang Y, et al. Molecular Identification and Genetic Characterization of Early-Stage Multiple Primary Lung Cancer by Large-Panel Next-Generation Sequencing Analysis. Front Oncol. 2021;11:653988.

50. Park E, Ahn S, Kim H, Park SY, Lim J, Kwon HJ, et al. Targeted Sequencing Analysis of Pulmonary Adenocarcinoma with Multiple Synchronous Ground-Glass/Lepidic Nodules. J Thorac Oncol. 2018;13(11):1776–83.

51. Qu R, Tu D, Ping W, Zhang N, Fu X. Synchronous Multiple Lung Cancers with Lymph Node Metastasis and Different EGFR Mutations: Intrapulmonary Metastasis or Multiple Primary Lung Cancers? Onco Targets Ther. 2021;14:1093–9.

52. McGranahan N, Swanton C. Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future. Cell. 2017;168(4):613–28.

53. Zhang Y, Wang Y, Lv C, Shu X, Wang J, Yang Q. Clinical analysis of 56 cases of simultaneous bilateral video-assisted thoracoscopic surgery for bilateral synchronous multiple primary lung adenocarcinoma. J Thorac Dis. 2018;10(12):6452–7.

54. Liu M, He W, Yang J, Jiang G. Surgical treatment of synchronous multiple primary lung cancers: a retrospective analysis of 122 patients. J Thorac Dis. 2016;8(6):1197–204.

55. Usuda J, Ichinose S, Ishizumi T, Hayashi H, Ohtani K, Maehara S, et al. Management of multiple primary lung cancer in patients with centrally located early cancer lesions. J Thorac Oncol. 2010;5(1):62–8.

56. Chang JY, Liu YH, Zhu Z, Welsh JW, Gomez DR, Komaki R, et al. Stereotactic ablative radiotherapy: a potentially curable approach to early stage multiple primary lung cancer. Cancer. 2013;119(18):3402–10.

57. Li M, Hao B, Zhang M, Reiter RJ, Lin S, Zheng T, et al. Melatonin enhances radiofrequency-induced NK antitumor immunity, causing cancer metabolism reprogramming and inhibition of multiple pulmonary tumor development. Signal Transduct Target Ther. 2021;6(1):330.

58. Sanmamed MF, Chen L. A Paradigm Shift in Cancer Immunotherapy: From Enhancement to Normalization. Cell. 2018;175(2):313–26.

59. Uprety D, Mandrekar SJ, Wigle D, Roden AC, Adjei AA. Neoadjuvant Immunotherapy for NSCLC: Current Concepts and Future Approaches. J Thorac Oncol. 2020;15(8):1281–97.

## Tables

Table 1 Tumor Identification, Pathological Subtype, and Sequencing information of 11 patients

| Patient ID | Sample ID | Tumor Location | Tumor Size (mm) | Histology | Pathological Subtype | | | | | TNM Stage | WES Depth/X (Mean) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | A% | P% | MP% | S% | L% | | |
| 1 | N | RLL | — | Normal | — | | | | | — | 346 |
| | T1 | RLL | 17*13*13 | ADC | | | | 100 | | T2 | 383 |
| | T2 | RLL | 10*10*9 | ADC | 10 | | | 90 | | T1b | 306 |
| | LN1 | Lower Lobe Parabronchial | — | ADC | | | | 100 | | N1 | 289 |
| | LN2 | Region 7 | — | ADC | | | | 100 | | N2 | 311 |
| | LN3 | Upper Lobe Parabronchial | — | ADC | | | | 100 | | N1 | 334 |
| | LN4 | Middle Lobe Parabronchial | — | ADC | 10 | 10 | | 80 | | N1 | 327 |
| 2 | N | RUL | — | Normal | — | | | | | — | 321 |
| | T1 | RUL | 12*10*10 | ADC | | 80 | 20 | | | T1b | 313 |
| | T2 | RUL | 10*10*7 | ADC | 100 | | | | | T1a | 221 |
| | LN1 | Extrapulmonary | — | ADC | 100 | | | | | N1 | 351 |
| | LN2 | Region 4R and 2R | — | ADC | 30 | 70 | | | | N2 | 313 |
| | LN3 | Region 11 | — | ADC | 100 | | | | | N1 | NA |
| 3 | N | RUL | — | Normal | — | | | | | | 367 |
| | T1 | LLL | 13*12*14 | ADC | 20 | 80 | | | | T1b | 295 |
| | T2 | RUL | 24*15*15 | ADC | | | 100 | | | T1c | NA |
| | LN1 | LUL Parabronchial | — | ADC | | | 100 | | | N1 | 261 |
| | LN2 | LLL Extrapulmonary | — | ADC | | | 100 | | | N1 | 183 |
| | LN3 | Region 3A | — | ADC | | | 100 | | | N2 | NA |
| 4 | N | LUL | — | Normal | — | | | | | — | 360 |
| | T1 | LUL | 25*15*10 | ADC | 30 | | | 70 | | T1c | 330 |
| | T2 | LUL | 22*9*3 | ADC | 60 | | | 40 | | T1c | 331 |
| | T3 | LUL | 12*10*10 | ADC | 30 | | | 70 | | T1c | 292 |
| | LN1 | Region 5 and 6 | | ADC | 20 | | | 80 | | N2 | 300 |
| 5 | N | RUL | — | Normal | — | | | | | — | NA |
| | T1 | RUL | 26*17*10 | ADC | | | | 100 | | T2 | 261 |
| | T2 | RUL | 30*26*20 | ADC | | | | 100 | | T1c | NA |
| | LN1 | Region 10 | — | ADC | | | | 100 | | N1 | 258 |
| | LN2 | Region 11 | — | ADC | | | | 100 | | N1 | 269 |
| | LN3 | Region 4R | — | ADC | | | | 100 | | N2 | 274 |
| 6 | N | RLL | — | Normal | — | | | | | — | 358 |
| | T1 | LUL | 25*22*20 | ADC | 10 | 10 | | 70 | 10 | T2 | 276 |
| | T2 | LLL | 15*13*5 | ADC | | | | 70 | 30 | T2 | 348 |
| | T3 | RLL | 9*6*5 | ADC | | 80 | | 20 | | T1a | 331 |
| | LN1 | Region 5 | — | ADC | 10 | 10 | | 80 | | N2 | NA |
| 7 | N | RLL | — | Normal | — | | | | | — | NA |
| | T1 | RLL | 10*5*3 | ADC | | 30 | 70 | | | T1a | 234 |
| | T2 | RUL | 30*25*30 | ADC | 40 | | | | 60 | T1c | NA |
| | LN1 | Extrapulmonary | — | ADC | 100 | | | | | N1 | 435 |
| | LN2 | Region 4R | — | ADC | 100 | | | | | N2 | 299 |
| 8 | N | RUL | — | Normal | — | | | | | — | 299 |

| | Sample | Location | Size | Pathology | A | P | MP | S | L | Stage | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | T1 | RUL | 15*13*10 | ADC | 90 | 10 | | | | T1b | 281 |
| | T2 | RUL | 29*20*18 | ADC | 30 | 30 | | 40 | | T2 | 306 |
| | LN1 | Region 4R | — | ADC | | 10 | | 90 | | N2 | 365 |
| 9 | N | RLL | — | Normal | — | | | | | — | 487 |
| | T1 | RLL | 65*35*35 | ADC | | | | 100 | | T3 | 328 |
| | T2 | RUL | 14*12*2 | ADC | | | | 80 | 20 | T3 | 363 |
| | LN1 | Subcarinal | — | ADC | | | | 100 | | N2 | 306 |
| 10 | N | LUL | — | Normal | — | | | | | — | 392 |
| | T1 | LUL | 30*30*22 | ADC | | 100 | | | | T2 | 237 |
| | T2 | LUL | 10*5 | ADC | 95 | 5 | | | | T1a | NA |
| | LN1 | Parabronchial | — | ADC | | | | 100 | | N1 | 344 |
| | LN2 | Parabronchial | — | ADC | 100 | | | | | N1 | 270 |
| | LN3 | Region 5 | — | ADC | | | | 100 | | N2 | 394 |
| 11 | N | LUL | | Normal | — | | | | | — | 63 |
| | T1 | LUL | 32*28*15 | ADC | | 5 | | 95 | | T2 | 72 |
| | T2 | LUL | 17*10*7 | ADC | | 20 | | 80 | | T1b | 69 |
| | LN1 | Extrapulmonary | — | ADC | | | | 100 | | N1 | 92 |
| | LN2 | Intrapulmonary | — | ADC | | | 10 | 90 | | N2 | 70 |
| | LN3 | Proximal Posterior Basal | — | ADC | | | | 100 | | N1 | 84 |
| | LN4 | Region 10 | — | ADC | | | | 100 | | N1 | 115 |

N, normal sample; T, tumor sample; LN, lymph node metastasis
RUL, right upper lobe; RLL, right lower lobe; LUL, left upper lobe; LLL, left lower lobe
ADC, adenocarcinoma
A, Acinar; P, Papillary; MP, Micropapillary; S, Solid; L, Lepidic
Failed, WES was not performed because of failure in cDNA library construction
NA, unknown because of WES failure.
H, high; M, middle; L, low; M-L, middle to low; H-M, high to middle

Table 2 Demographic information and clinical characteristics of 11 patients with MPLC

| Patient ID | Age/Gender | Ethnicity | Smoking, PY | Alcohol, y | Family History | Pattern | Surgery | Comprehensive Classification* | Adjunct Therapy | Patient Status# | Recur. Dis. Met |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 42/F | Han | 0 | 0 | Yes | Syn | Thoracotomy | Primary | ALK Targeted therapy | Alive | Rec |
| 2 | 68/F | Han | 0 | 0 | Yes | Syn | VATS | Primary | None | Alive | Non |
| 3 | 45/F | Han | 0 | 0 | No | Syn | VATS | Primary | EGFR Targeted therapy | Alive | Bon Me |
| 4 | 56/M | Han | 35 | 0 | No | Syn | VATS | Primary | Chemotherapy | Alive | Non |
| 5 | 76/M | Han | 0 | 20 | No | Syn | VATS | Primary | Chemotherapy +EGFR Targeted therapy | NA | NA |
| 6 | 48/F | Han | 0 | 0 | No | Meta | VATS | Primary | None | Dead | NA |
| 7 | 64/F | Han | 0 | 0 | No | Syn | VATS | Primary | None | Alive | Non |
| 8 | 52/M | Han | 0 | 25 | Yes | Syn | VATS | Primary | None | Dead | Rec |
| 9 | 65/M | Han | 45 | 30 | Yes | Syn | Thoracotomy | Primary | None | Dead | Rec |
| 10 | 73/M | Han | 33 | 0 | No | Syn | VATS | Primary | EGFR Targeted therapy +Radiotherapy | Alive | Bra Me |
| 11 | 67/M | Han | 0 | 0 | No | Syn | VATS | Primary | Capecitabine+Icotinib | Alive | Non |

F, female; M, male. PY, packed-years, y, years. Syn, synchronous; Meta, metachronous. NA, unknown. Recur, Recurrence. VATS, video-assisted thoracoscopic s
* Patients are classified according to the 8th TNM staging system published in 2017.
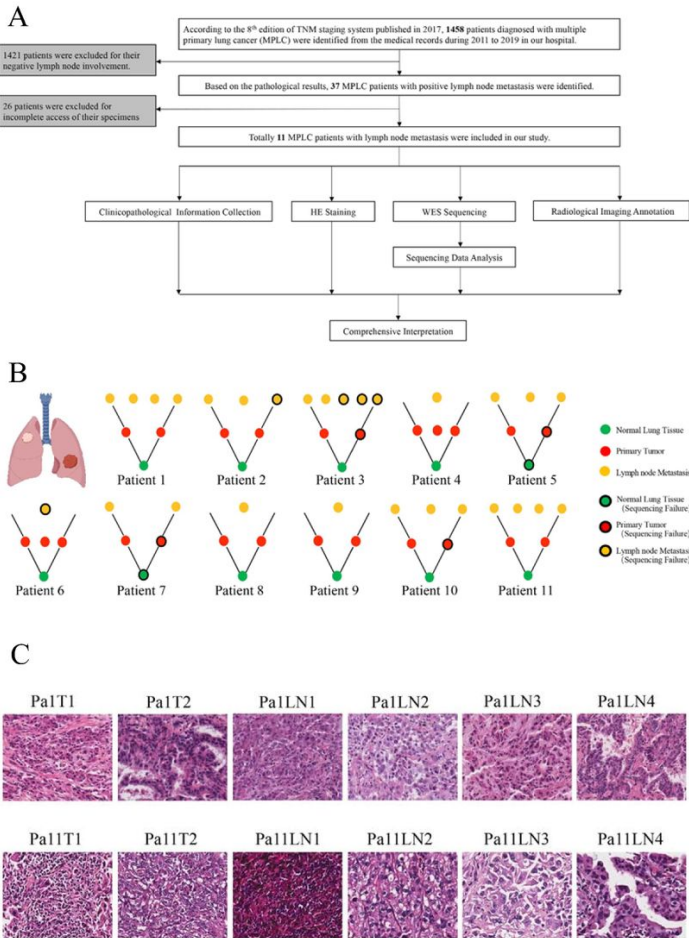# Until the last follow-up time

# Figures

## Figure1



# Figure 1

## Study design and patient description

**A.** The flow chart showed the design and workflow of our research.

**B.** The cartoons showed the concept of multiple primary lung cancer, the different colors of the two tumors indicated they are genetically independent. Pattern diagrams constituted by dots of different colors showed the tumor architecture within each patient. The green, red, and orange dots represented standard samples, primary lung tumors, and lymph node metastases sites. The black circle indicated that the sample existed but failed to produce valid data in WES.

**C.** Representative HE staining image of patient 1 and patient 11.
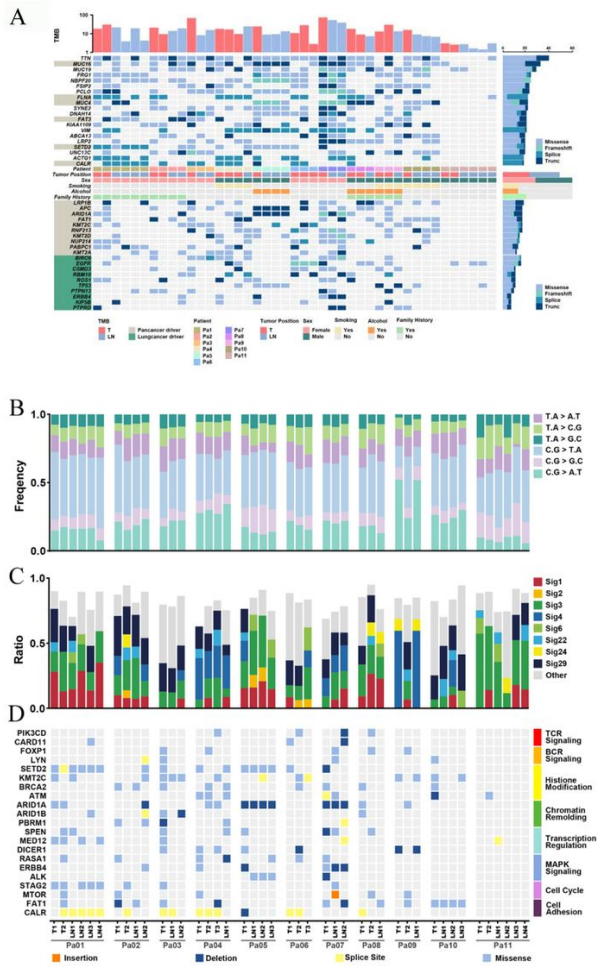
**Figure 2**

**Mutational landscape of MPLC with lymph node metastasis.**

**A.** Significantly mutated genes (SMGs). Upper: Top 20 genes with highest mutation frequency. Middle: Demographic and clinical information of the 11 patients. Bottom: Highly mutated pan-cancer driver genes (light brown) and lung cancer driver genes (green), ranked according to their frequency. Driver gene identification is based on the COSMIC Cancer Gene Census (Sep. 2021, https://cancer.sanger.ac.uk/census).

**B.** Six-type mutation spectra of all samples.

**C.** Signature contributions of all samples based on the COSMIC database.

**D.** Putative driver genes (based on MSK-IMPACT 468 gene panel) with somatic mutations in all 11 patients were classified according to the functional categories
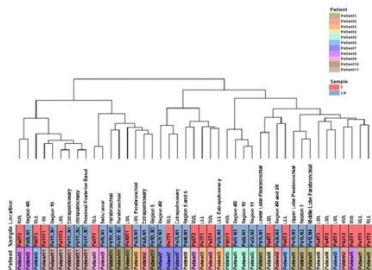
**Figure3**

A
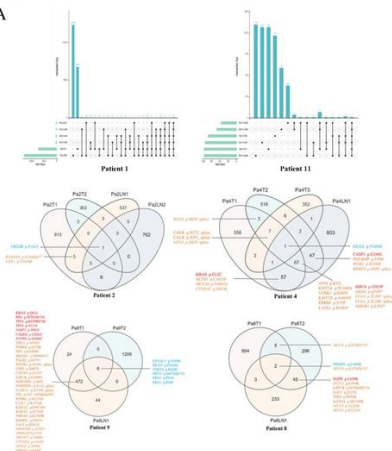


B
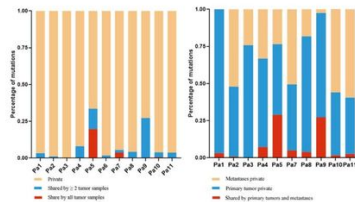


C



Figure 3

Genomic heterogeneity among different samples.

**A.** Upset map and Venn diagrams of representative patients showing the distribution of nonsynonymous somatic mutations among different tumors within individuals (Patient 1 and 11: Upset Venn diagrams. Patient 2, 4, 8, and 9: Venn diagrams). The putative pathogenic mutations were marked with different colors and typefaces according to the oncogene list in COSMIC Cancer Gene Census (https://cancer.sanger.ac.uk/census). Orange, Pan-cancer driver gene; Red, NSCLC/lung cancer driver gene; Blue, genes are not pan-cancer driver genes or NSCLC/lung cancer driver gene. Roman type, Tie 1 gene in COSMIC (gene possessing a documented activity relevant to cancer, along with evidence of promoting oncogenic transformation); Italic, Tie 2 gene in COSMIC (genes with strong indications of a role in cancer but with less extensive available evidence).

**B.** Distributions of mutations. Left: percentages of mutations shared by

all the tumor samples (Red), two or more tumor samples (Blue), and private to only one tumor sample (Yellow) in each patient, respectively. Right: percentages of mutations shared by primary lung tumors and lymph node metastases (Red), private to primary lung tumor samples (Blue), and private to lymph node metastasis samples (Yellow), regardless of tumor numbers.

**D.** The unsupervised clustering of all 11 patients based on non-synonymous mutations. All 11 patients are distinguished with different colors. Red, primary tumor. Blue, lymphatic metastasis tumors.
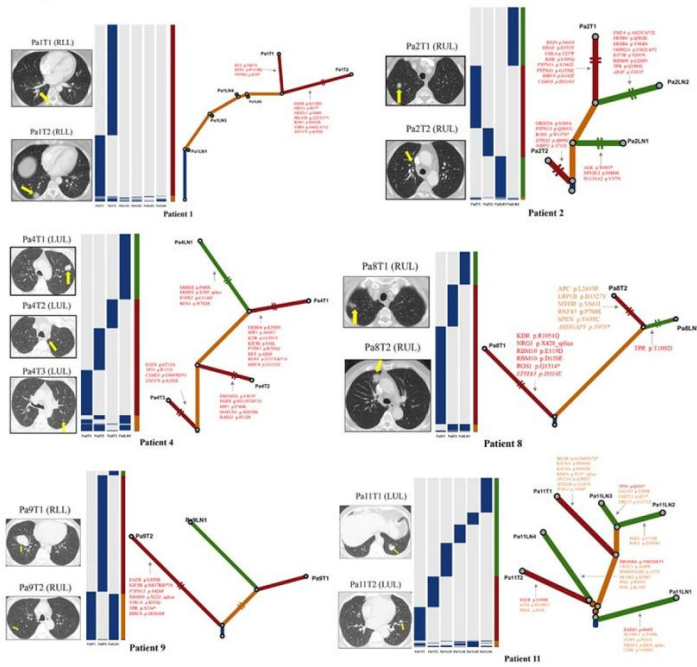
# Figure4



Figure 4

**Clonal architecture and evolution of representative patients.**

**Left:** CT (computerized tomography) images of each patient, with yellow arrows marking the tumor's location. **Middle:** Heatmaps show the distribution of all non-silent mutations; presence (blue), absence (gray). Colum next to the heatmap shows the distribution of mutations; mutation present in all tumors (blue), shared in more than one but not all tumors (orange), in one pulmonary tumor (red), and in one lymph node metastasis tumor (green). **Right:** Phylogenetic trees based on the distribution of all detected mutations. Trunk and branch lengths are proportional to the number of non-silent mutations acquired. Putative driver genes are indicated next to the trunk or with an arrow pointing to the branches where they were detected. Orange, Pan-cancer driver gene; Red, known NSCLC/lung cancer driver gene; Roman type and Italic represent tie 1 gene and tie 2 genes in COSMIC, respectively.
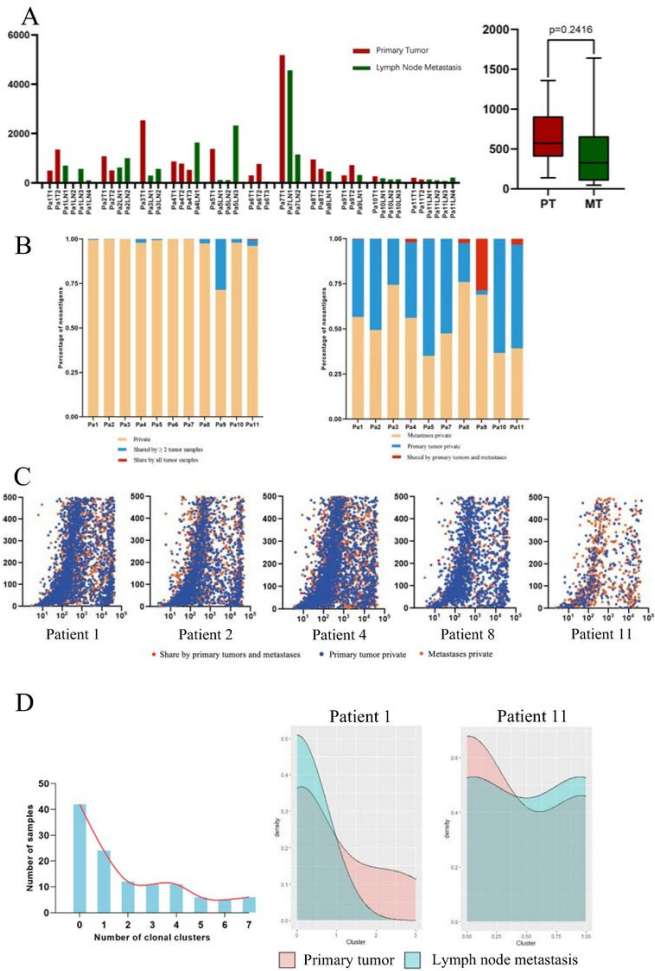
# Figure5



Figure 5

**Immunogenicity heterogeneity across and within individuals.**

**A.** Left: The number of all predicted neoantigens in each tumor of 11 patients. Right: The difference in the mean number of potential neoantigens in primary tumors and lymph node metastasis tumors based on Patient1, 2, 4, 8, 9, and 11 (p value=0.2416).

**B.** Distributions of predicted neoantigens. Left: percentages of neoantigens shared by all the tumor samples (Red), two or more tumor samples (Blue), and private to only one tumor sample (Yellow) in each patient, respectively. Right: percentages of neoantigens shared by primary lung tumors and lymph node metastases (Red), private to primary lung tumor samples (Blue), and private to lymph node metastasis samples (Yellow), regardless of tumor numbers.

**C.** Predictions of neoantigen binding affinity across all 9-11 amino acids peptides generated from nonsynonymous mutations and the matched wild-type peptides using NetMHCpan algorithms. Red, neoantigens shared by both primary and lymph node metastasis tumors; Blue, neoantigens private to primary lung tumors; Orange, neoantigens private to lymph node metastasis tumor.

**D.** Histograms and density plots of SciClone inferred clonal clusters. Left: Clonal clusters of all 11 patients. Right: Clonal clusters density plots of Patient1 and 11.

## Supplementary Files

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

- SupplementaryData1.xlsx
- SupplementaryData2.xlsx
- SupplementaryData3.xlsx
- SupplementaryData4.xlsx
- SupplementaryData5.xlsx
- SupplementaryData6.xlsx

- SupplementaryData7.xlsx
- SupplementaryData8.xlsx
- SupplementaryData9.xlsx
- SupplementaryData10.xlsx
- SupplementaryData11.xlsx
- SupplementaryData12.xlsx
- SupplementaryData13.xlsx
- SupplementaryData14.xlsx
- SupplementaryData15.xlsx
- SupplementaryData16.xlsx
- SupplementaryFigure1.pdf
- SupplementaryFigure2.pdf
- SupplementaryFigure3.pdf
- SupplementaryFigure4.pdf
- SupplementaryFigure5.pdf
- SupplementaryFigure6.pdf