Additional file 1

Optimizing the NGS-based discrimination of multiple lung cancers from the perspective of evolution

Ziyang Wang[*], Xiaoqiu Yuan[*], Ke Liu[*], Kunkun Sun[*], Olga Chervova, Yuntao Nie, Airong Yang, Yichen Jin, Jing Li, Yun Li, Fan Yang, Jun Wang, Stephan Beck, David Carbone, Kezhong Chen[#]

Table of Contents

Supplementary methods:

Sequencing workflow of NGS

The choice of sequencing platforms was made based on the patients' preferences expressed in discussing various options. Gene coverages of targeted panels were presented in Table S6.

Nucleic acid extraction

DNA was extracted from formalin-fixed, paraffin-embedded (FFPE) tissues using the QIAamp DNA FFPE Tissue Kit (QIAGEN, Valencia, Calif), while RNA was extracted from FFPE tissues and reversely transcribed into complementary DNA using the Recover ALL Total Nucleic Acid Isolation Kit (1407077) according to recommended protocols. To ensure the quality of the libraries, DNA was quantified with the Qubit dsDNA HS Assay Kit (Life Technologies, Carlsbad, Calif) and Qubit 2.0/3.0/4.0 Fluorometer (Life Technologies, USA). RNA was quantified using an SYBR Green Real-Time PCR kit (TaKaRa, Japan) in a CFX96 Real-Time PCR Detection System (Bio-Rad, United States).

The 8-genes NGS

Initial sequencing data of *EGFR*, *KRAS*, *BRAF*, *ERBB2*, and *MET* mutants were processed using Torrent Suite (5.0.4) and Coverage Analysis plugins (5.0.4.0), along with the Ion Reporter (4.4) Oncomina DNA and RNA analysis workflow. The Torrent Mapping Alignment Program and the Torrent Variant Caller plugin (5.0.4.0) were respectively applied to align DNA sequences and call variants, under default settings. Two filtering steps were used to eliminate misidentified bases and generate final calls. *ALK, ROS1,* and *RET* fusion analysis was performed by the Ion Reporter (4.4) workflow and the Torrent Mapping Alignment Program.

The 31/68/143/363/457/520/654-genes NGS

Purified DNA was first broken into fragments of around 300 bp using 5X WGS Fragmentation Mix (Qiagen, USA), followed by end repair and adapter ligation. For the 31/68/143/363/457/654-genes NGS, polymerase chain reaction (PCR) with gene targeted by the cSMART methods[1], and 96 RXN xGen Exome Research Panel v1.0 (Integrated DNA Technologies, USA) were used successively for the library preparation. NovaSeq 6000 platform (Illumina, San Diego, USA) in 150PE mode was applied for sequencing, with a depth of more than 1000x. Then We used *fastp*[2] to remove low-quality results, Burrows-Wheeler Aligner (BWA)[3] to align the filtered data to reference genome (hg19), Gencore tool[4] to manage repeated sequences, SAMtools[5] to call variants, and ANNOVAR[6] to annotate mutations. Non-synonymous SNPs with population allele frequencies lower than 0.05% and variant allele fraction (VAF) higher than 0.5% were kept for further analysis. For the 520-genes NGS, we used Nextseq500 sequencer (Illumina, Inc., USA) for paired-end sequencing and gene targeting with a depth of more than 1000x. Variations with population allele frequencies lower than 0.1% were reserved for further analysis, referring to ExAC, 1000Genomes, dbSNP, and ESP6500SI-V2 databases.

Whole exome sequencing

The library was prepared with the SureSelectXT Target Enrichment System (G7530-90000). Paired-end sequence data were generated applying the Illumina HiSeq machine with an average depth of 150×, and aligned to the reference genome (hg19) using BWA. We used the SAMtools to sort original alignment results, Picard (http://broadinstitute.github.io/picard/) to mark duplication reads, and GATK[7] to reduce false positives via local realignment and recalibration. SNVs and small indels were called by MuTect2[8] in the default setting and then annotated by ANNOVAR according to HGVS, ExAC, 1000Genomes, dbSNP, OMIM, COSMIC, and ClinVar databases. Variations with population allele frequencies lower than 0.05% were subjected to further analysis.

Figure S1. Distribution of genes covered by involved NGS panels

Distribution of each NGS panel in our study. The red part represents the overlapped genes with 654-genes panel as a reference.

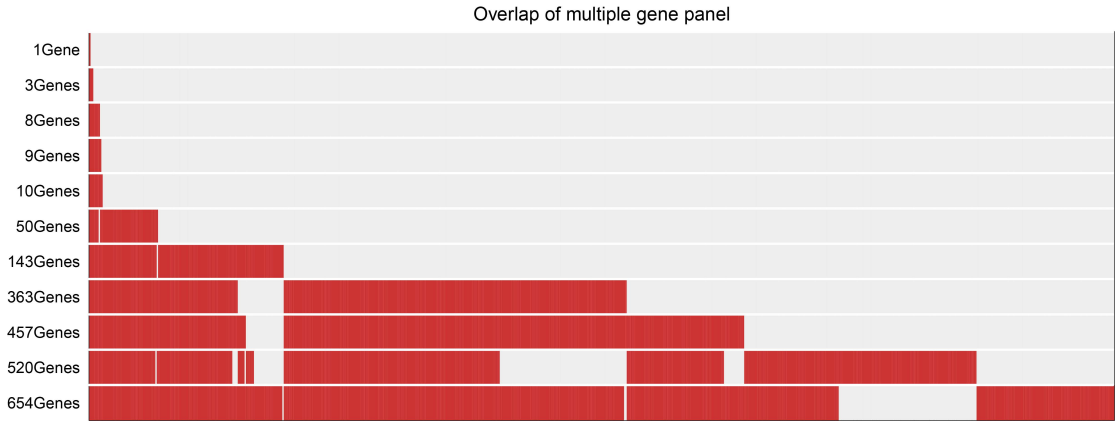


Figure S2. ROC curve for each panel with MoleA

MoleA: empirical interpretation. AUC: area under the receiver operating curve.

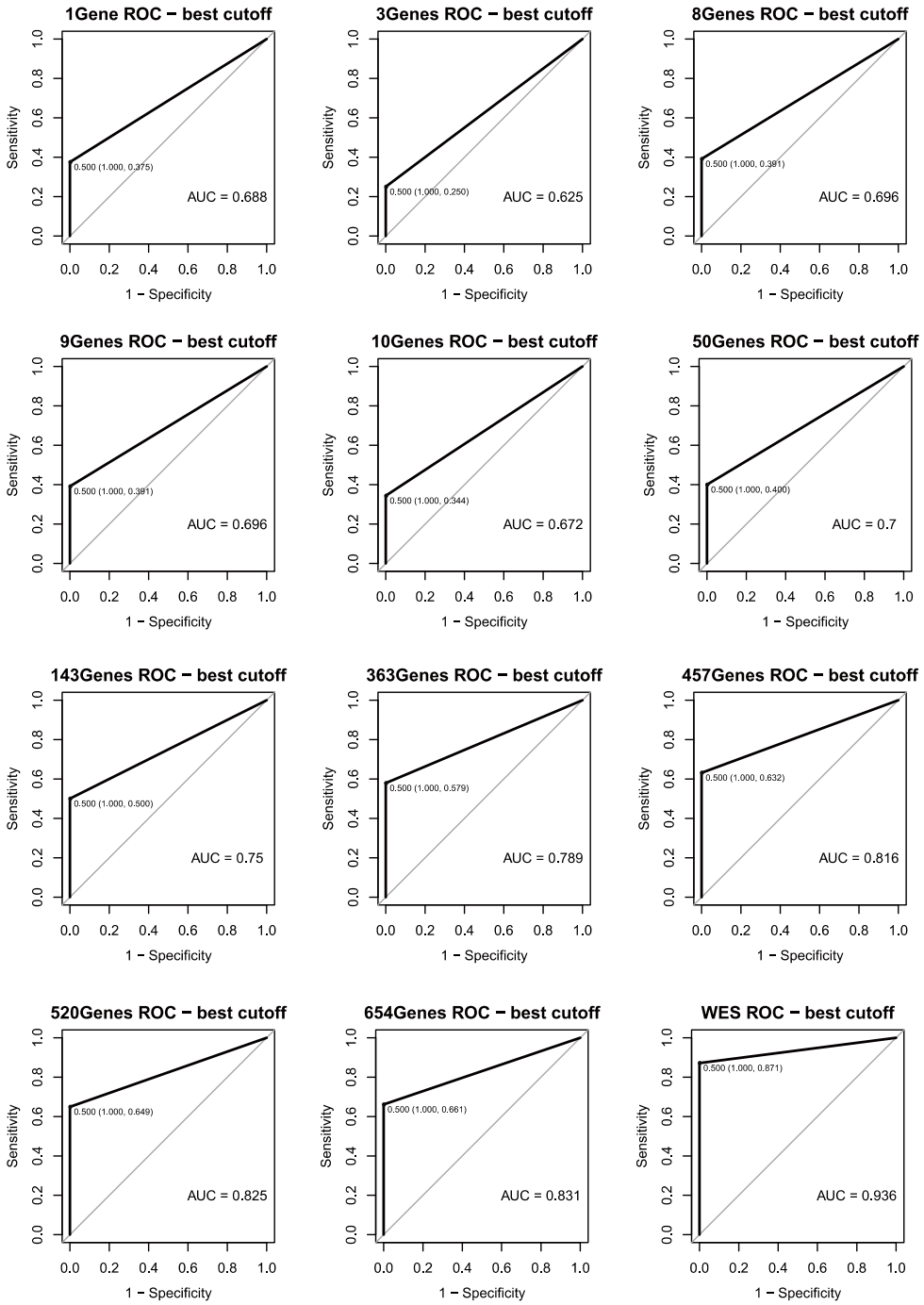


Figure S3. ROC curve for each panel with MoleB

MoleA: empirical interpretation by counting the shared mutations. MoleB: bioinformatic interpretation by calculating the clonal probability based on all the mutations.

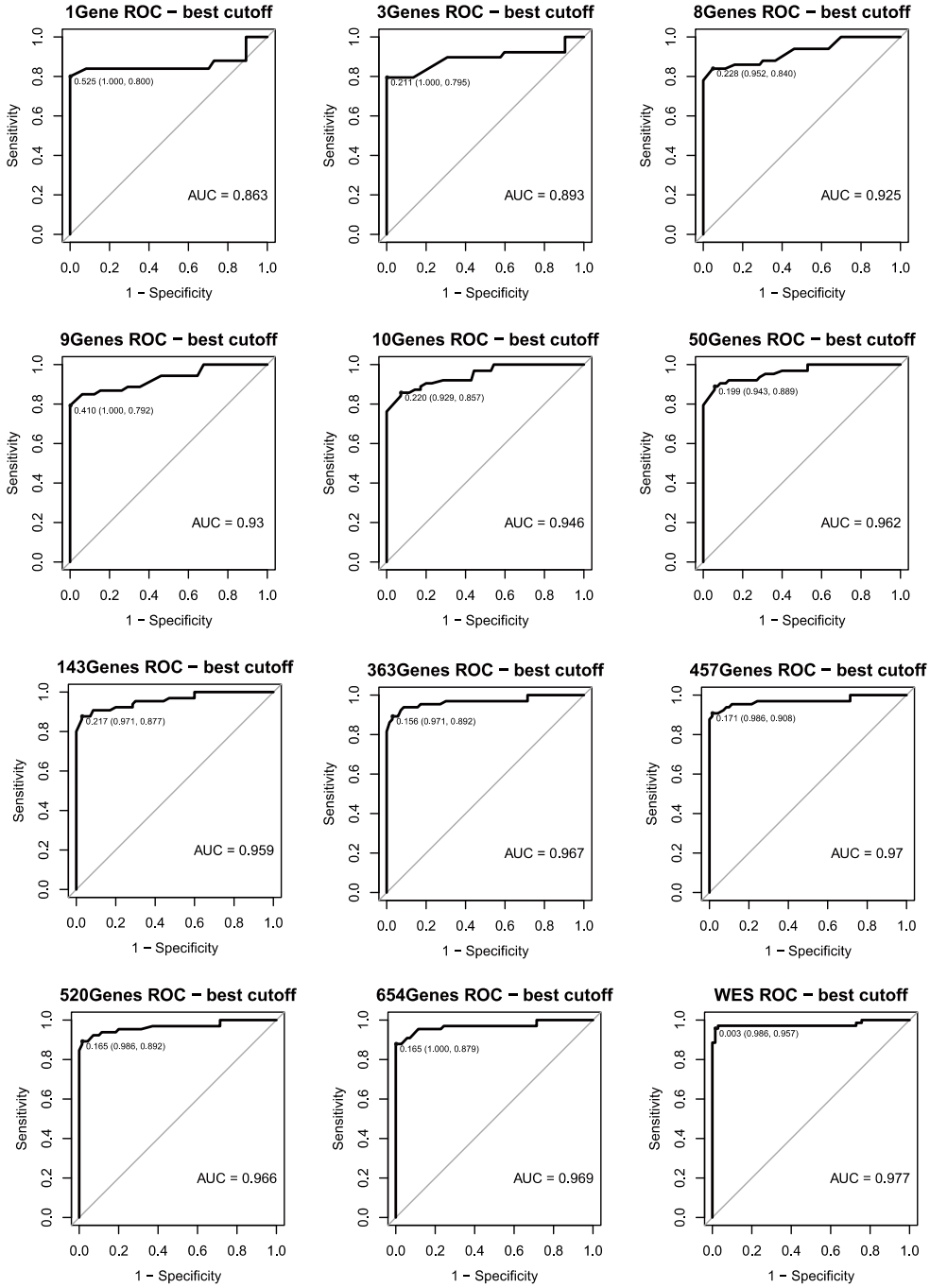


Figure S4. Survival analyses of the WES group by different panels with MoleB

MLCs: multiple lung cancers. MPLC: multiple primary lung cancers. IPM: intrapulmonary metastasis.

WES: whole-exome sequencing. MoleB: bioinformatic interpretation.

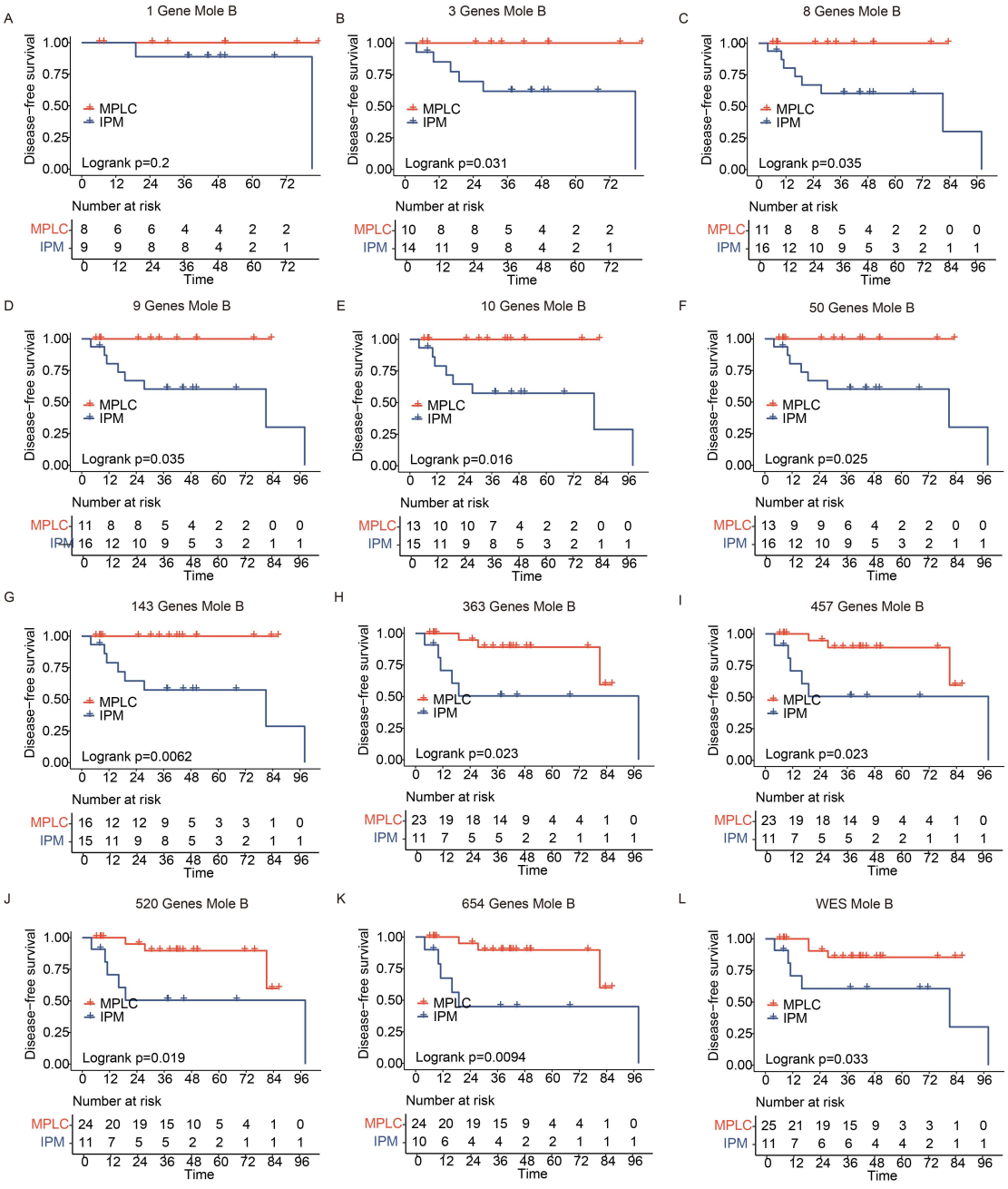


Figure S5. Survival analyses of the WES group by different panels with MoleA

MLCs: multiple lung cancers. MPLC: multiple primary lung cancers. IPM: intrapulmonary metastasis.

WES: whole-exome sequencing. MoleA: empirical interpretation.

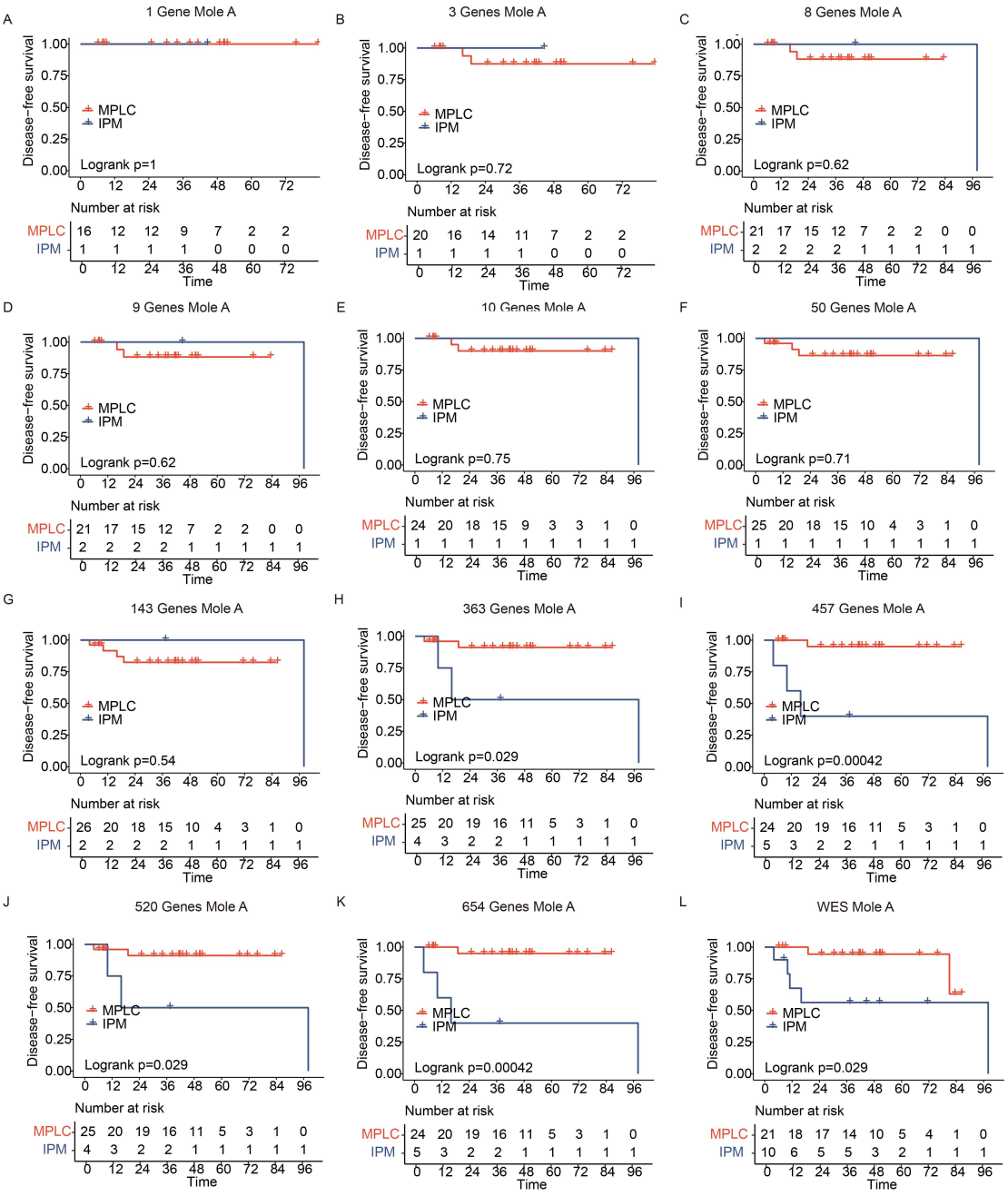
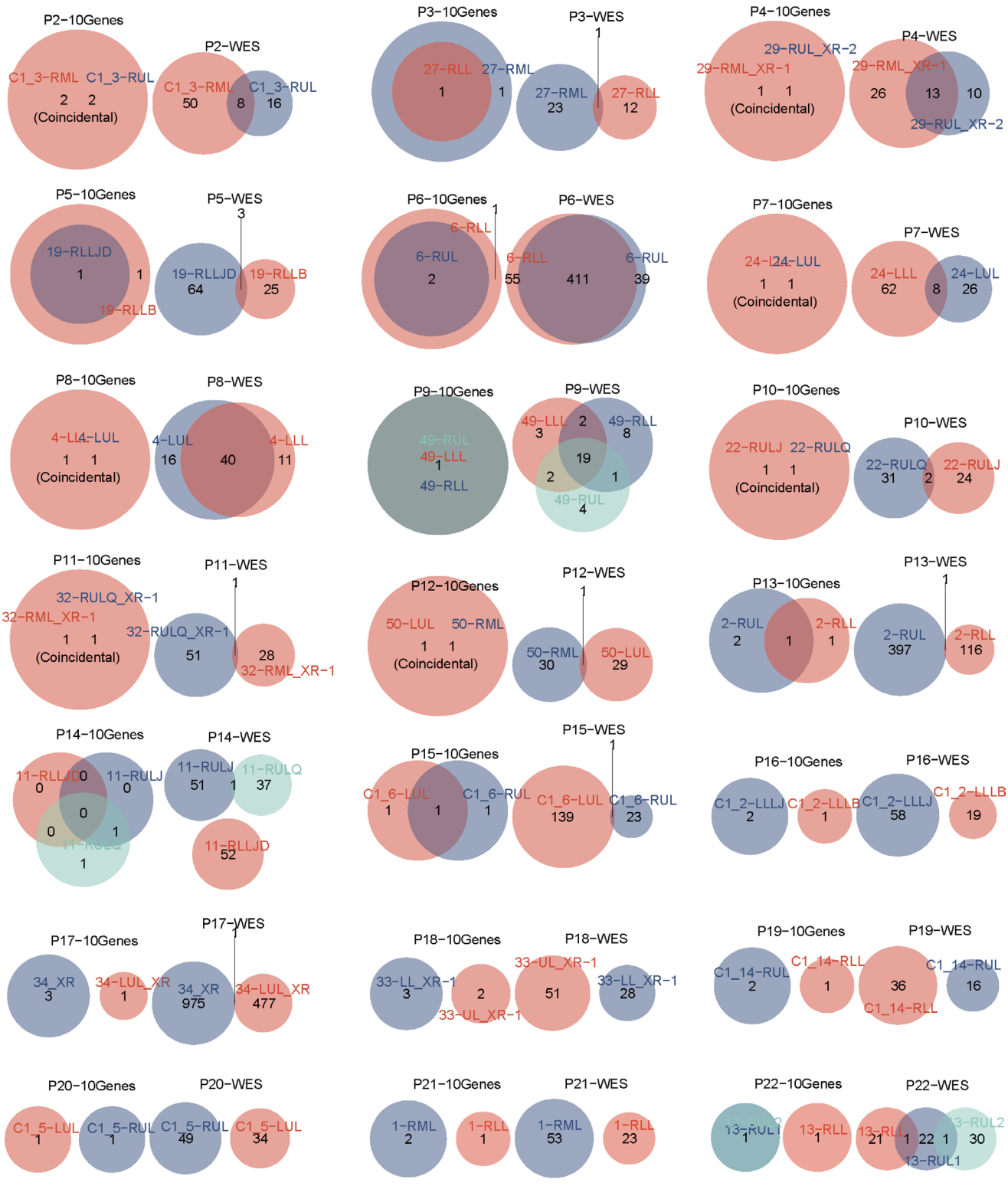


Figure S6. Venn diagram of mutations detected in each patient under 10-genes panel of WES

WES: whole-exome sequencing. RUL: right upper lobe. RML: right middle lobe. RLL: right lower lobe. LUL: left upper lobe. LLL: left lower lobe.



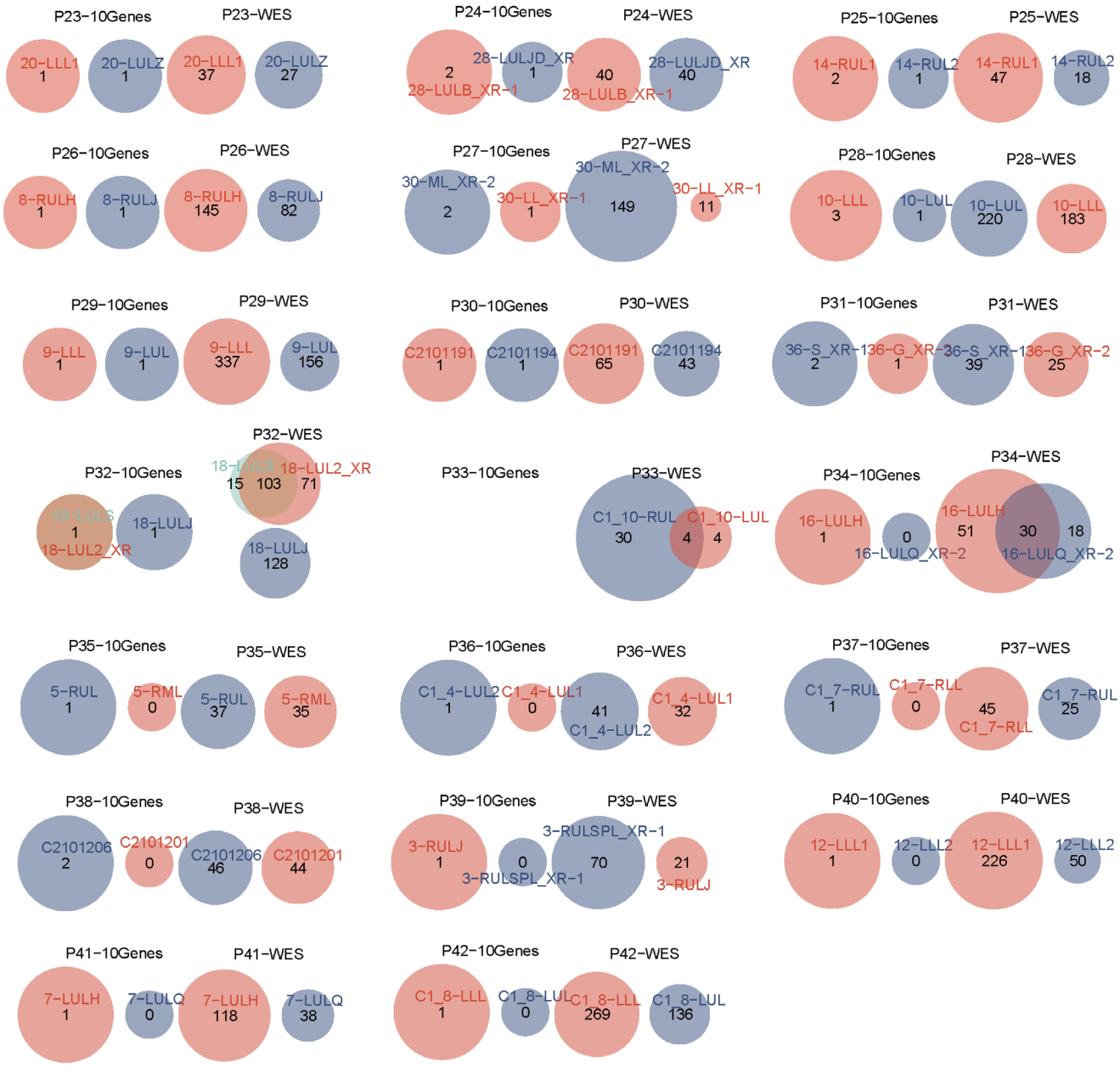

P23-10Genes
P23-WES
20-LLL1 1
20-LULZ 1
20-LLL1 37
20-LULZ 27
P24-10Genes
P24-WES
2
28-LULB_XR-1
28-LULJD_XR 1
40
28-LULB_XR-1
28-LULJD_XR 40
P25-10Genes
P25-WES
14-RUL1 2
14-RUL2 1
14-RUL1 47
14-RUL2 18
P26-10Genes
P26-WES
8-RULH 1
8-RULJ 1
8-RULH 145
8-RULJ 82
P27-10Genes
P27-WES
30-ML_XR-2 2
30-LL_XR-1 1
30-ML_XR-2 149
30-LL_XR-1 11
P28-10Genes
P28-WES
10-LLL 3
10-LUL 1
10-LUL 220
10-LLL 183
P29-10Genes
P29-WES
9-LLL 1
9-LUL 1
9-LLL 337
9-LUL 156
P30-10Genes
P30-WES
C2101191 1
C2101194 1
C2101191 65
C2101194 43
P31-10Genes
P31-WES
36-S_XR-1 2
36-G_XR-1 1
36-S_XR-1 39
36-G_XR-2 25
P32-WES
P32-10Genes
18-LULS
15
103
18-LUL2_XR 71
18-LULS 1
18-LULJ 1
18-LUL2_XR
18-LULJ 128
P33-10Genes
P33-WES
C1_10-RUL 30
4
C1_10-LUL 4
P34-WES
P34-10Genes
16-LULH 1
0
16-LULQ_XR-2
16-LULH 51
30
18
16-LULQ_XR-2
P35-10Genes
P35-WES
5-RUL 1
5-RML 0
5-RUL 37
5-RML 35
P36-10Genes
P36-WES
C1_4-LUL2 1
C1_4-LUL1 0
41
C1_4-LUL2
C1_4-LUL1 32
P37-10Genes
P37-WES
C1_7-RUL 1
C1_7-RLL 0
45
C1_7-RLL
C1_7-RUL 25
P38-10Genes
P38-WES
C2101206 2
C2101201 0
C2101206 46
C2101201 44
P39-10Genes
P39-WES
3-RULJ 1
0
3-RULSPL_XR-1
3-RULSPL_XR-1 70
21
3-RULJ
P40-10Genes
P40-WES
12-LLL1 1
12-LLL2 0
12-LLL1 226
12-LLL2 50
P41-10Genes
P41-WES
7-LULH 1
7-LULQ 0
7-LULH 118
7-LULQ 38
P42-10Genes
P42-WES
C1_8-LLL 1
C1_8-LUL 0
C1_8-LLL 269
C1_8-LUL 136

Figure S7. Analysis of the non-WES cohort verifies the superiority of optimal panels in discriminating MLCs

Similar to the analysis of the WES cohort, 49 patients in the non-WES cohort were sequenced using panels covering more than 10 genes, while the other 45 patients were tested with smaller panels that covered fewer than 9 genes. (A) Clinical features, discriminating results, and mutational profiles of the 49 patients who got sequenced by panels⩾10 genes in non-WES cohort. Here stage refers to the highest stage among the MLCs when staging all lesions separately. (B) Survival analyses stratified by clinic criteria of ACCP criteria. (C) Survival analyses stratified by CHA. (D) Survival analyses of the 49 patients diagnosed by the 10-genes panel with MoleB.

MLCs: multiple lung cancers. MPLC: multiple primary lung cancers. IPM: intrapulmonary metastasis. WES: whole-exome sequencing. MoleA: empirical interpretation. MoleB: bioinformatic interpretation. CHA: comprehensive histology assessment.

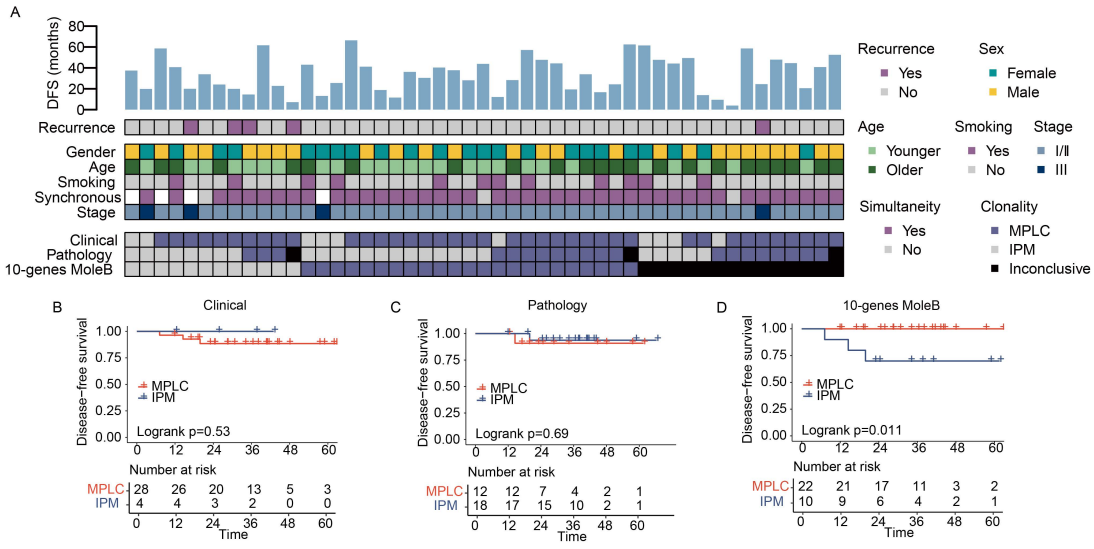


Figure S8. Distribution of VAFs of overlapped mutations in MLCs

The TumE algorithm was used to estimate overlapped mutations in MLCs sequenced at 150x mean sequencing depth and a subclone at 40% cellular fraction.

MLCs: multiple lung cancers. VAF: variant allele frequency.

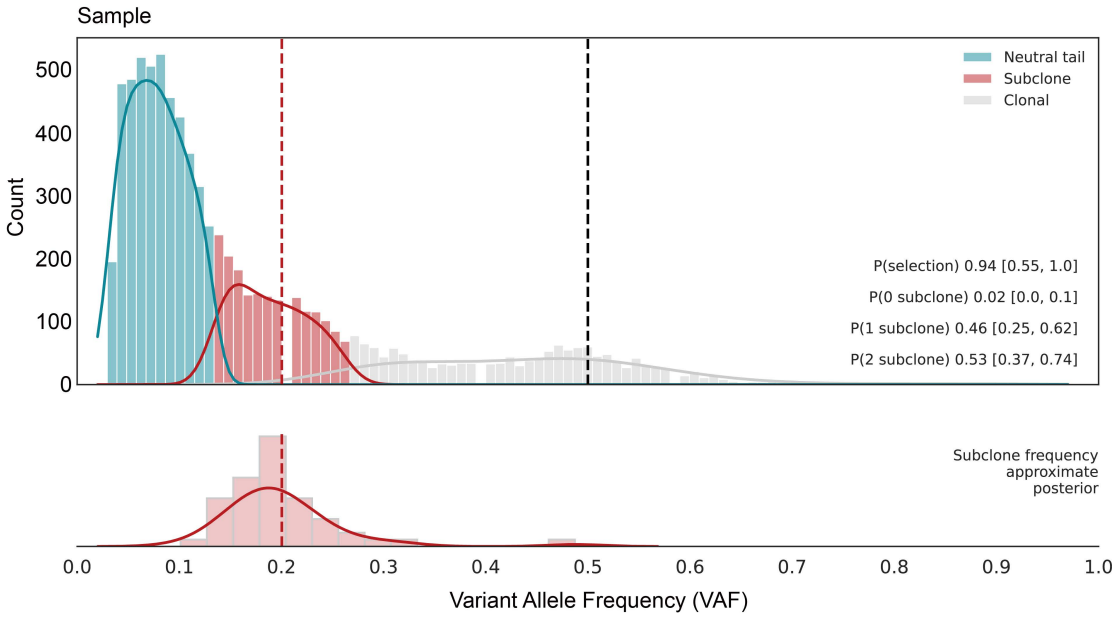
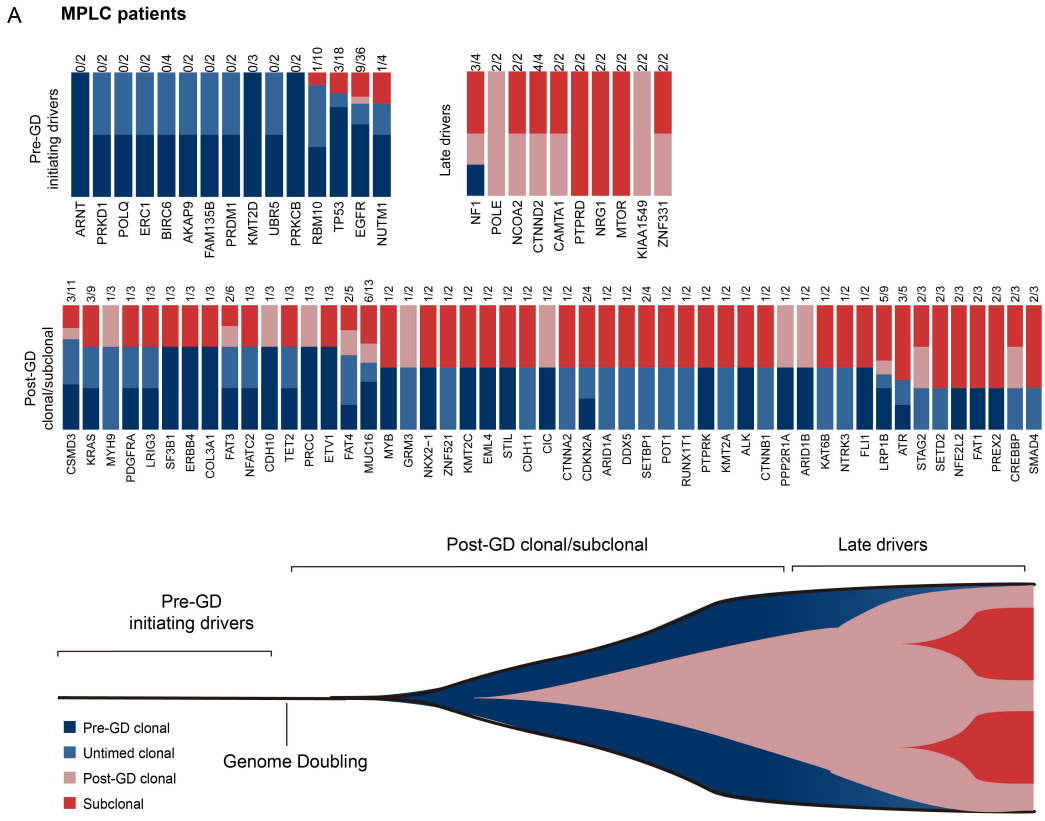


Figure S9. Timing of mutations in MLCs evolution

(A) Tumor evolution of all mutations in MPLC patients. (B) Tumor evolution of all mutations in IPM patients. Two figures show the approximate timing of driver mutations with respect to the cancer life history. Driver genes were screened against the COSMIC database. The timing of mutations is shown as bars indicating whether the events are clonal or subclonal.

MLCs: multiple lung cancers. MPLC: multiple primary lung cancers. IPM: intrapulmonary metastasis. Pre-GD: occurrence before whole-genome doubling. Post-GD: occurrence after whole-genome doubling.


A
MPLC patients
Pre-GD initiating drivers
0/2 0/2 0/2 0/2 0/4 0/2 0/2 0/2 0/3 0/2 0/2 1/10 3/18 9/36 1/4
ARNT PRKD1 POLQ ERC1 BIRC6 AKAP9 FAM135B PRDM1 KMT2D UBR5 PRKCB RBM10 TP53 EGFR NUTM1
Late drivers
3/4 2/2 2/2 4/4 2/2 2/2 2/2 2/2 2/2 2/2 2/2
NF1 POLE NCOA2 CTNND2 CAMTA1 PTPRD NRG1 MTOR KIAA1549 ZNF331
Post-GD clonal/subclonal
3/11 3/9 1/3 1/3 1/3 1/3 1/3 2/6 1/3 1/3 1/3 1/3 2/5 6/13 1/2 1/2 1/2 1/2 1/2 1/2 1/2 1/2 1/2 2/4 1/2 2/4 1/2 1/2 1/2 1/2 1/2 1/2 1/2 1/2 1/2 1/2 1/2 1/2 1/2 1/2 5/9 3/5 2/3 2/3 2/3 2/3 2/3 2/3 2/3
CSMD3 KRAS MYH9 PDGFRA LRIG3 SF3B1 ERBB4 COL3A1 FAT3 NFATC2 CDH10 TET2 PRCC ETV1 FAT4 MUC16 MYB GRM3 NKX2−1 ZNF521 KMT2C EML4 STIL CDH11 CIC CTNNA2 CDKN2A ARID1A DDX5 SETBP1 POT1 RUNX1T1 PTPRK KMT2A ALK CTNNB1 PPP2R1A ARID1B KAT6B NTRK3 FLI1 LRP1B ATR STAG2 SETD2 NFE2L2 FAT1 PREX2 CREBBP SMAD4
Post-GD clonal/subclonal
Late drivers
Pre-GD initiating drivers
Pre-GD clonal
Untimed clonal
Post-GD clonal
Subclonal
Genome Doubling

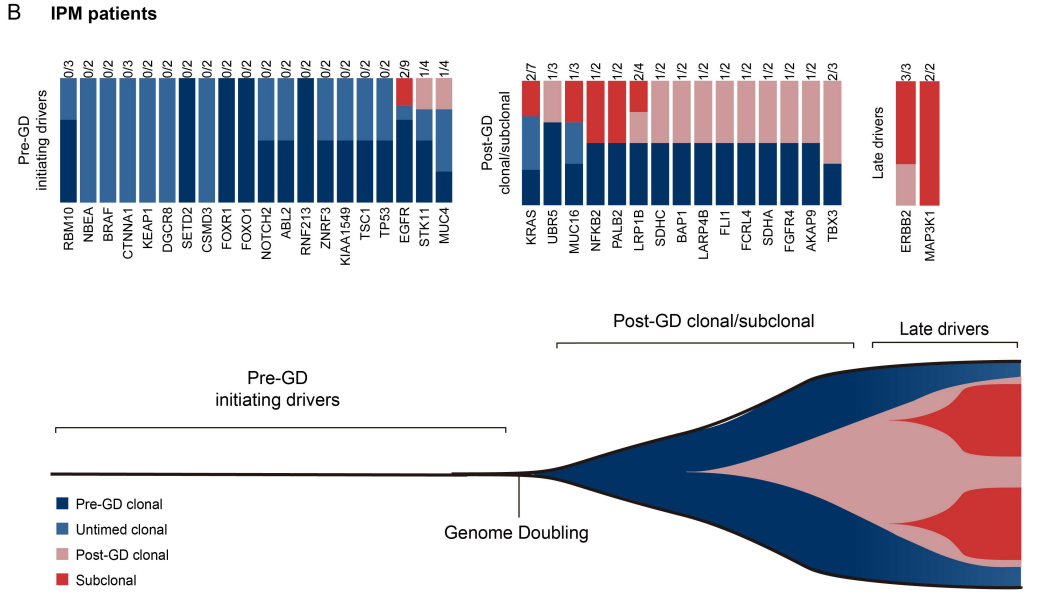
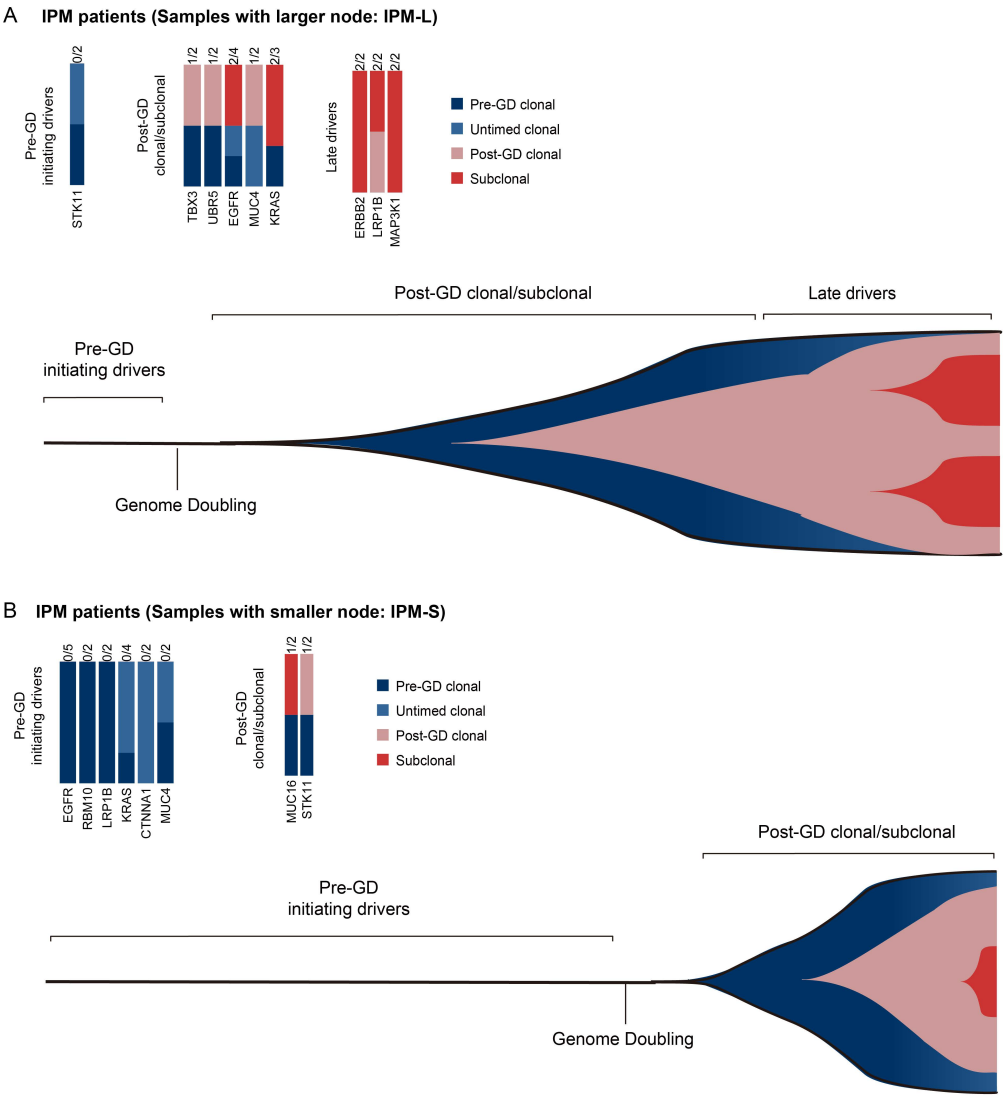
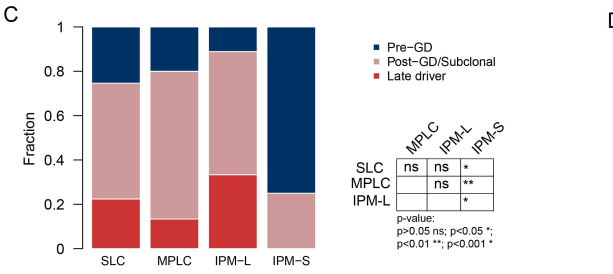

B
IPM patients
Pre-GD initiating drivers
0/3 0/2 0/2 0/3 0/2 0/2 0/2 0/2 0/2 0/2 0/2 0/2 0/2 0/2 0/2 0/2 2/9 1/4 1/4
RBM10 NBEA BRAF CTNNA1 KEAP1 DGCR8 SETD2 CSMD3 FOXR1 FOXO1 NOTCH2 ABL2 RNF213 ZNRF3 KIAA1549 TSC1 TP53 EGFR STK11 MUC4
Post-GD clonal/subclonal
2/7 1/3 1/3 1/2 1/2 2/4 1/2 1/2 1/2 1/2 1/2 1/2 1/2 1/2 2/3
KRAS UBR5 MUC16 NFKB2 PALB2 LRP1B SDHC BAP1 LARP4B FLI1 FCRL4 SDHA FGFR4 AKAP9 TBX3
Late drivers
3/3 2/2
ERBB2 MAP3K1
Post-GD clonal/subclonal
Late drivers
Pre-GD initiating drivers
Pre-GD clonal
Untimed clonal
Post-GD clonal
Subclonal
Genome Doubling

Figure S10. Timing of mutations in IPM evolution and temporal analysis between different lung cancer cohorts

(A) Tumor evolution of all mutations in larger nodules of IPM patients. (B) Tumor evolution of all mutations in smaller nodules of IPM patients. (C) Distributions of driver mutations in cancer evolution in SLC, MPLC, IPM-L and IPM-S. (D) Distributions of driver mutations in cancer evolution in SLC and IPM-L.

MPLC: multiple primary lung cancers. IPM: intrapulmonary metastasis. IPM-L: the larger lesion (primary). IPM-S: the smaller lesion (metastasis). SLC: solitary lung cancer. Pre-GD: occurrence before whole-genome doubling. Post-GD: occurrence after whole-genome doubling.


A
IPM patients (Samples with larger node: IPM-L)
Pre-GD initiating drivers
Post-GD clonal/subclonal
Late drivers
STK11
TBX3
UBR5
EGFR
MUC4
KRAS
ERBB2
LRP1B
MAP3K1
Pre-GD clonal
Untimed clonal
Post-GD clonal
Subclonal
Post-GD clonal/subclonal
Late drivers
Pre-GD initiating drivers
Genome Doubling
B
IPM patients (Samples with smaller node: IPM-S)
EGFR
RBM10
LRP1B
KRAS
CTNNA1
MUC4
MUC16
STK11
Post-GD clonal/subclonal
Pre-GD initiating drivers
Genome Doubling


C
Fraction
SLC
MPLC
IPM−L
IPM−S
Pre−GD
Post−GD/Subclonal
Late driver
p-value:
p>0.05 ns; p<0.05 *;
p<0.01 **; p<0.001 *

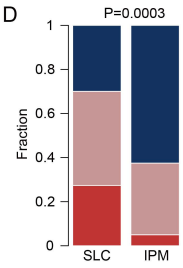

D
P=0.0003
Fraction
SLC
IPM

Table S1. Genes contained in each panel and the number of patients sequenced by each panel in our cohort

This table is provided as an excel file (additional file 2).

Table S2. The optimal parameters of the mutation.rem function of the "Clonality" package and the cutoffs for the calculated probabilities

Panel	Optimal parameters			Optimal cutoff of probability	Specificity	Sensitivity
	μ	σ	π			
1 Gene	0.59	0.94	0.46	0.52	1	0.8
3 Genes	0.6	0.93	0.45	0.21	1	0.79
8 Genes	0.56	0.93	0.47	0.23	0.95	0.84
9 Genes	0.46	0.94	0.47	0.41	1	0.79
10 Genes	0.63	0.89	0.47	0.22	0.93	0.86
50 Genes	0.61	0.84	0.44	0.2	0.94	0.89
143 Genes	0.49	0.81	0.44	0.22	0.97	0.88
341 Genes	0.31	0.72	0.44	0.1	0.93	0.92
363 Genes	0.23	0.72	0.43	0.16	0.97	0.89
457 Genes	0.19	0.7	0.45	0.17	0.99	0.91
468 Genes	0.19	0.7	0.45	0.17	0.99	0.91
520 Genes	0.24	0.68	0.45	0.17	0.99	0.89
654 Genes	0.2	0.68	0.44	0.16	1	0.88
WES	-0.11	0.69	0.45	0.003	0.99	0.96

Table S3. Studies included in the systematic review

First author [Year]	Sample (patients / tumors/ pairs)	NGS panel (genes)	Clonal relatedness		Clinicopathology method[b]	Inconclusive		Changed diagnosis by NGS	Survival analysis	
			Interpretation	Special MPLC definition[a]		NGS	Clinicopathology		NGS	Clinicopathology
Zheng [2020] [9]	18/41/18	4	Empirical	--	HPE	17% (3/18)	6% (1/18)	53% (8/15)	NA	
Donfrancesco [2020][10]	24/50/24	14	Empirical	--	CHA+IHC	13% (3/24)	0%	38% (9/24)	PFS: P=0.054	PFS: P=0.60
Takahashi [2018][11]	37/82/37	20	Empirical	--	HPE	0%	17 (46%)	59% (22/17)	OS: P=0.21	OS: P=0.70
Belardinilli [2021][12]	10/24/10	22	Empirical	Sharing one high-frequency driver mutation plus different mutation(s)	CHA+IHC	10% (1/10)	0%	22% (2/9)	NA	
Goodwin [2021][13]	40/80/40	24	Bio-informatic	Clonality package	HPE	0%	0%	18% (7/40)	OS: P<0.05	OS: P=0.1
Bruehl [2022][14]	32/64/32	35	Empirical	Sharing high-frequency driver mutation (s)	HPE	25% (8/32)	NA	17% (4/24)	OS: P=0.78	NA
Patel [2017][15]	11/31/11	50	Empirical	Using distant metastasis to establish the reference concordant rate	CHA	0%	27% (3/11)	27% (3/11)	OS: P=0.09	NA
Saab [2017][16]	18/52/18	50	Empirical	--	HPE+IHC	6% (1/18)	33% (6/18)	6% (1/17)	NA	
Roepman [2018][17]	50/111/57	50	Empirical	Sharing high-frequency driver mutation (s)	CHA+IHC	4% (2/53 pairs)	0%	37% (19/51 pairs)	OS: P=0.061	OS: P=0.644
Ezer [2021][18]	61/131/61	52	Bio-informatic	Self-proposed algorithm	CHA	13% (8/61)	13% (8/61)	23% (12/53)	NA	
Higuchi [2020][19]	37/76/37	53	Empirical	--	HPE	0%	0%	30% (11/37)	NA	
Chang [2019][20]	60/128/76	341-468	Bio-informatic	Clonality package	CHA	0%	30% (23/76)	22% (17/76 pairs)	PFS: P=0.197	NA
Liu [2020][21]	16/40/16	464	Empirical	--	HPE	0%	6% (1/16)	14% (2/14)	NA	
Wang [2021][22]	51/116/51	605	Empirical	Sharing one high-frequency driver mutation ± one low-frequency mutation	HPE	0%	0% exclude all IPM	0%	NA	
Pei [2021][23]	30/67/45	808	Empirical	Sharing only EGFR L858R	Martini	0%	0%	47% (14/30)	NA	

a. In empirical interpretation, MPLC was usually defined as cases with different mutations or as cases where one tumor had a mutation and the other tumor was wild-type. Here, "Special" refers to the definition which was supplemented by the original literature aside from the above-mentioned definition.

b. HPE refers to a pathological assessment method that evaluates the main pathological types and subtypes but does not belong to CHA.

MLCs: multiple lung cancers. CHA: comprehensive histology assessment. IHC: immunohistochemistry. NGS: next-generation sequencing. OS: overall survival. PFS: progression-free survival. NA: not available or not applicable.

Table S4. HR values of OS (MPLC versus IPM, classified by empirical or bioinformatic interpretation)

	First Author [year]	MPLC	IPM	HR (95% CI)	P value
Empirical interpretation	Chang [2019]	41	19	0.742 [0.146, 3.766]	0.197
	Donfrancesco [2020]	14	10	0.602 [0.066, 5.458]	0.054
	Ezer [2021]	46	15	0.377 [0.046, 3.067]	0.4
	Roepman [2018]	34	23	0.467 [0.05, 4.345]	0.061
	Takahashi [2018]	17	20	0.515[0.013, 19.664]	0.21
	Goodwin[2021	33	7	0.11 [0.03, 0.4]	<0.05
	Belardinilli [2021]	8	2	0.239[0.034, 1.684]	0.151
	Bruehl [2022]	20	7	0.457 [0.084, 2.499]	0.57
	Patel [2017]	8	3	0 [0, 1.4615397059142E+265]	0.09
Bioinformatic interpretation	Belardinilli [2021]	7	4	3.981 [0.1413, 112.2]	-
	Bruehl [2022]	20	7	0.2788 [0.04169, 1.864]	-
	Patel [2017]	7	1	0.08454 [0.004879, 1.465]	-

Note: All data were obtained by grabbing and calculating the survival curves from available studies in the part of systematic review.

Table S5. Discrimination of MLCs by pathology or NGS in included studies

Author [year]	Diagnosis		Pathology			Samples
			MPLCs	IPMs	IN	
Chang [2019]	NGS	MPLCs	8	0	0	11 patients
		IPMs	0	0	3	
		IN	0	0	0	
Chang [2019]		MPLCs	23	11	0	57 pairs
		IPMs	0	9	0	
		IN*4	11	3	0	
Chang [2019]		MPLCs	9	1	7	37 patients
		IPMs	8	6	6	
		IN	0	0	0	
Chang [2019]		MPLCs	45	6	0	76 pairs
		IPMs	11	14	0	
		IN	0	0	0	
Donfrancesco [2020]		MPLCs	9	6	0	24 patients
		IPMs	2	4	0	
		IN	2	1	0	
Higuchi [2020]		MPLCs	27	2	0	37 patients
		IPMs	4	4	0	
		IN	0	0	0	
Zheng [2020]		MPLCs	6	4	1	18 patients
		IPMs	0	4	0	
		IN	0	3	0	
Liu [2020]		MPLCs	12	0	0	16 patients
		IPMs	2	1	1	
		IN	0	0	0	
Wang [2021]		MPLCs	44	0	0	51 patients
		IPMs	7	0	0	
		IN	0	0	0	
Chang [2019]		MPLCs	14	4	0	32 patients
		IPMs	0	6	0	
		IN	3	5	0	
Chang [2019]		MPLCs	8	1	0	11 pairs
		IPMs	1	0	0	
		IN	1	0	0	
Chang [2019]		MPLCs	38	4	4	61 patients
		IPMs	2	4	1	
		IN	8	0	0	

Table S6. Discriminating results of the different methods applied to each patient in included studies

This table is provided as an excel file (additional file 2).

Table S7. Conclusive rates of different discriminating methods subsampled from included studies

Panels used in included studies	Clinical	Pathology	Empirical interpretation	Bioinformatic interpretation	Bioinformatic interpretation			
					sub3 genes	sub9 genes	sub10 genes	sub50 genes
below 10 genes	100.00 %	NA	100	91.75%	91.72 %	90.45 %	87.90%	77.07%
around 20 genes	96.77%	99.17%	98.14%	97.52%	NA	93.94 %	75.25%	64.65%
around 50 genes	100.00 %	83.33%	93.18%	89.39%	NA	87.88 %	81.82%	77.27%
around 300 genes	99.99%	97.78%	82.22%	88.89%	NA	NA	NA	80.00%

Table S8. Unique Driver mutations that occur at different times in MLCs compared to an Asian cohort of solitary NSCLC

Driver	Timing in MPLCs
ARNT	Pre-GD initiating drivers
PRKD1	Pre-GD initiating drivers
ERC1	Pre-GD initiating drivers
PRDM1	Pre-GD initiating drivers
PRKCB	Pre-GD initiating drivers
MYH9	Post-GD clonal/subclonal
LRIG3	Post-GD clonal/subclonal
ERBB4	Post-GD clonal/subclonal
COL3A1	Post-GD clonal/subclonal
TET2	Post-GD clonal/subclonal
PRCC	Post-GD clonal/subclonal
ETV1	Post-GD clonal/subclonal
MYB	Post-GD clonal/subclonal
NKX2-1	Post-GD clonal/subclonal
EML4	Post-GD clonal/subclonal
STIL	Post-GD clonal/subclonal
CIC	Post-GD clonal/subclonal
CTNNA2	Post-GD clonal/subclonal
DDX5	Post-GD clonal/subclonal
PTPRK	Post-GD clonal/subclonal
KMT2A	Post-GD clonal/subclonal
ALK	Post-GD clonal/subclonal
CTNNB1	Post-GD clonal/subclonal
PPP2R1A	Post-GD clonal/subclonal
KAT6B	Post-GD clonal/subclonal
STAG2	Post-GD clonal/subclonal
NFE2L2	Post-GD clonal/subclonal
NF1	Late drivers
NCOA2	Late drivers
CTNND2	Late drivers
CAMTA1	Late drivers
KIAA1549	Late drivers
ZNF331	Late drivers

References

1. Lv W, Wei X, Guo R, et al. Noninvasive prenatal testing for Wilson disease by use of circulating single-molecule amplification and resequencing technology (cSMART). *Clin Chem*. 2015;61(1):172-181. doi:10.1373/clinchem.2014.229328
2. Chen S, Zhou Y, Chen Y, Gu J. Fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 2018;34(17):i884-i890. doi:10.1093/bioinformatics/bty560
3. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754-1760. doi:10.1093/bioinformatics/btp324
4. Chen S, Zhou Y, Chen Y, et al. Gencore: An efficient tool to generate consensus reads for error suppressing and duplicate removing of NGS data. *BMC Bioinformatics*. 2019;20(Suppl 23):1-8. doi:10.1186/s12859-019-3280-9
5. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25(16):2078-2079. doi:10.1093/bioinformatics/btp352
6. Wang K, Li M, Hakonarson H. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38(16):1-7. doi:10.1093/nar/gkq603
7. McKenna A, Hanna M, Banks E, et al. The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20(9):1297-1303. doi:10.1101/gr.107524.110
8. Cibulskis K, Lawrence MS, Carter SL, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol*. 2013;31(3):213-219. doi:10.1038/nbt.2514
9. Zheng R, Shen Q, Mardekian S, Solomides C, Wang ZX, Evans NR. Molecular profiling of key driver genes improves staging accuracy in multifocal non–small cell lung cancer. *Journal of Thoracic and Cardiovascular Surgery*. 2020;160(2):e71-e79. doi:10.1016/j.jtcvs.2019.11.126
10. Donfrancesco E, Yvorel V, Casteillo F, et al. Histopathological and molecular study for synchronous lung adenocarcinoma staging. *Virchows Archiv*. 2020;476(6):835-842. doi:10.1007/s00428-019-02736-0

11. Takahashi Y, Shien K, Tomida S, et al. Comparative mutational evaluation of multiple lung cancers by multiplex oncogene mutation analysis. *Cancer Sci*. 2018;109(11):3634-3642. doi:10.1111/cas.13797

12. Belardinilli F, Pernazza A, Mahdavian Y, et al. A multidisciplinary approach for the differential diagnosis between multiple primary lung adenocarcinomas and intrapulmonary metastases. *Pathol Res Pract*. 2021;220(November 2020):153387. doi:10.1016/j.prp.2021.153387

13. Goodwin D, Rathi V, Conron M, Wright GM. Genomic and Clinical Significance of Multiple Primary Lung Cancers as Determined by Next-Generation Sequencing. *Journal of Thoracic Oncology*. 2021;16(7):1166-1175. doi:10.1016/j.jtho.2021.03.018

14. Bruehl FK, Doxtader EE, Cheng YW, Farkas DH, Farver C, Mukhopadhyay S. Does histological assessment accurately distinguish separate primary lung adenocarcinomas from intrapulmonary metastases? A study of paired resected lung nodules in 32 patients using a routine next-generation sequencing panel for driver mutations. *J Clin Pathol*. 2022;75(6):390-396. doi:10.1136/jclinpath-2021-207421

15. Patel SB, Kadi W, Walts AE, et al. Next-Generation Sequencing: A Novel Approach to Distinguish Multifocal Primary Lung Adenocarcinomas from Intrapulmonary Metastases. *Journal of Molecular Diagnostics*. 2017;19(6):870-880. doi:10.1016/j.jmoldx.2017.07.006

16. Saab J, Zia H, Mathew S, Kluk M, Narula N, Fernandes H. Utility of Genomic Analysis in Differentiating Synchronous and Metachronous Lung Adenocarcinomas from Primary Adenocarcinomas with Intrapulmonary Metastasis. *Transl Oncol*. 2017;10(3):442-449. doi:10.1016/j.tranon.2017.02.009

17. 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. *Journal of Molecular Diagnostics*. 2018;20(4):436-445. doi:10.1016/j.jmoldx.2018.02.007

18. Ezer N, Wang H, Corredor AG, et al. Integrating NGS-derived mutational profiling in the diagnosis of multiple lung adenocarcinomas. *Cancer Treat Res Commun*. 2021;29(October):1-10. doi:10.1016/j.ctarc.2021.100484

19. Higuchi R, Nakagomi T, Goto T, et al. Identification of clonality through genomic profile

analysis in multiple lung cancers. *J Clin Med*. 2020;9(2). doi:10.3390/jcm9020573

20. 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. *Clinical Cancer Research*. 2019;25(23):7113-7125. doi:10.1158/1078-0432.CCR-19-1700

21. Liu J, Mao G, Li Y, et al. Targeted deep sequencing helps distinguish independent primary tumors from intrapulmonary metastasis for lung cancer diagnosis. *J Cancer Res Clin Oncol*. 2020;146(9):2359-2367. doi:10.1007/s00432-020-03227-5

22. Wang X, Gong Y, Yao J, et al. Establishment of Criteria for Molecular Differential Diagnosis of MPLC and IPM. *Front Oncol*. 2021;10(January):1-10. doi:10.3389/fonc.2020.614430

23. Pei G, Li M, Min X, 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(May):1-12. doi:10.3389/fonc.2021.653988