Appendix 1

Tree-based models are effective in capturing non-linear relationships within data. By recursively partitioning the feature space, they generate a series of decision rules that effectively handle complex pairwise and higher-order interactions, as well as non-linear relationships between features. As a result, decision trees are particularly useful for exploring non-linear associations between complex biomarkers and disease states. Ensemble methods based on decision trees retain the ability to manage non-linear data while further enhancing model robustness and predictive accuracy. Consequently, Decision Trees, Random Forest, and Gradient Boosting Trees were ultimately selected for model construction.

1. Decision Trees

Decision trees are hierarchical models that represent decisions in a tree-like structure, in which data features are recursively split into binary or multi-way partitions to generate a series of decision rules for classification. At each node, the model selects a feature and divides the data based on its values, aiming to maximize a chosen split criterion, such as information gain, Gini index, or entropy. The decision tree recursively splits starting from the root node until the sample size in each leaf node satisfies predefined stopping conditions, such as reaching the maximum depth or no further information gain.

Here, the Gini index is used as an example of a splitting criterion. To quantify the impurity of a node, the following formula is utilized:

$$G(D)=1-\sum_{i=1}^{k} p_i^2$$

Here, p_i represents the proportion of category i within the dataset D.

2. Random Forest

Random Forest is an ensemble learning method that combines multiple independently trained decision trees. During the training of each tree, bootstrap sampling is performed on the original data, and a subset of features is randomly selected for splitting at each node. The final classification result is determined through a majority voting process.

Assuming there are T decision trees, each providing a classification result of $h_t(x)$, the final classification result of the Random Forest is determined by a majority voting process.

$$H(x)=\operatorname{argmax}\sum_{t=1}^{T} I\left(h_t(x)=c\right)$$

Here, I is an indicator function that equals 1 when $h_t(x)=c$ and 0 otherwise, where c represents the class.

3. Gradient Boosting Decision Trees

Gradient Boosting Decision Trees is a gradient boosting model based on decision trees that constructs a series of decision trees iteratively to optimize the accuracy of model predictions. In each iteration, GBDT gradually reduces classification error by fitting the residuals of the previous model, with each tree used to correct the mispredictions of its predecessor.

3.1 Loss function (for classification tasks): For binary classification problems, cross-entropy loss is commonly used.

$$L(y,p)=-\left(y\log(p)+(1-y)\log(1-p)\right)$$

Here, y represents the actual class, and p denotes the predicted probability of the model.

3.2 Residual Calculation: In each iteration, GBDT fits a new decision tree by calculating the negative gradient of the objective function. For the model in the m-th iteration, the residuals (i.e., the negative gradient) are defined as:

$$r_{im}=-\left(\frac{\partial L\left(y_i, F_{m-1}(x_i)\right)}{\partial F_{m-1}(x_i)}\right)$$

Here, $F_{m-1}(x_i)$ represents the predicted value of the model from the $m-1$ iteration.

https://dx.doi.org/10.21037/tlcr-24-875

3.3 Tree Fitting: In the m-th iteration, GBDT constructs a new decision tree $h_m(x_i)$ by fitting these residuals. The new tree is used to correct the errors made by the previous iteration.

3.4 Model Update: After each iteration, the model is updated using the following formula:

$$F_m(x) = F_{m-1}(x) + \alpha h_m(x)$$

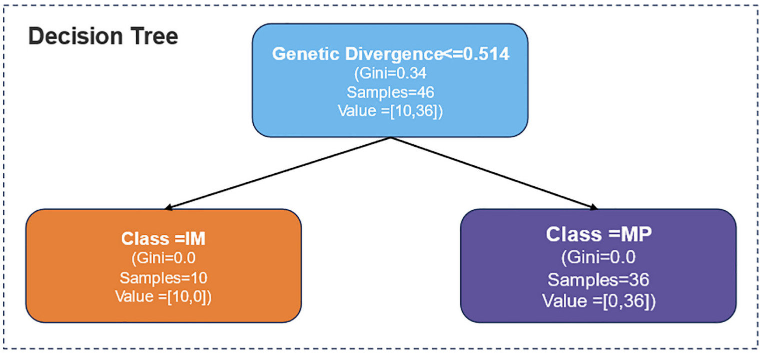
Here, α represents the learning rate, which determines the step size for each update.

3.5 Final Classification Decision: For binary classification problems, GBDT determines the class by using the sign of the final predicted value:

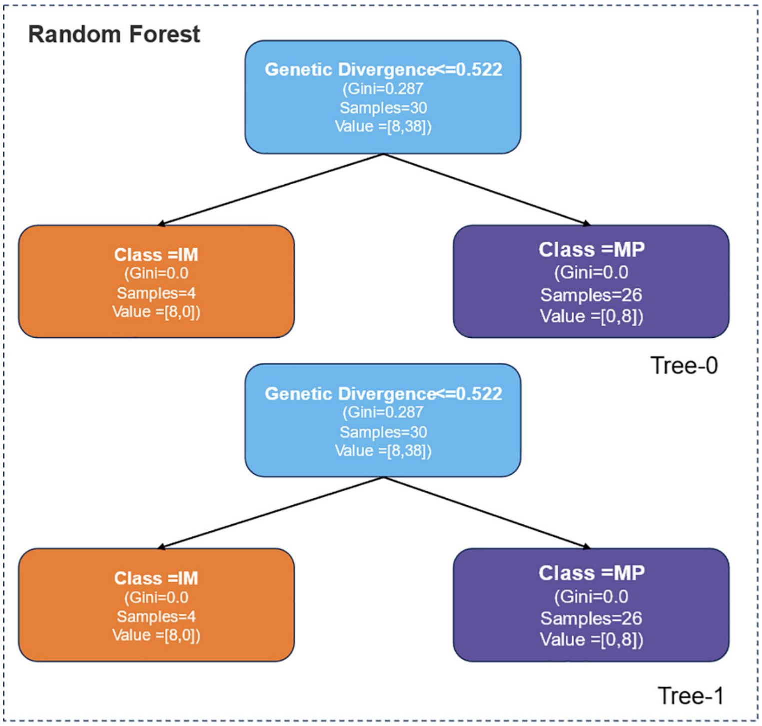
$$\hat{y} = \operatorname{sign}\left(F_M(x)\right)$$

Here, ,M represents the total number of iterations, and $F_M(x)$ denotes the final predicted score.

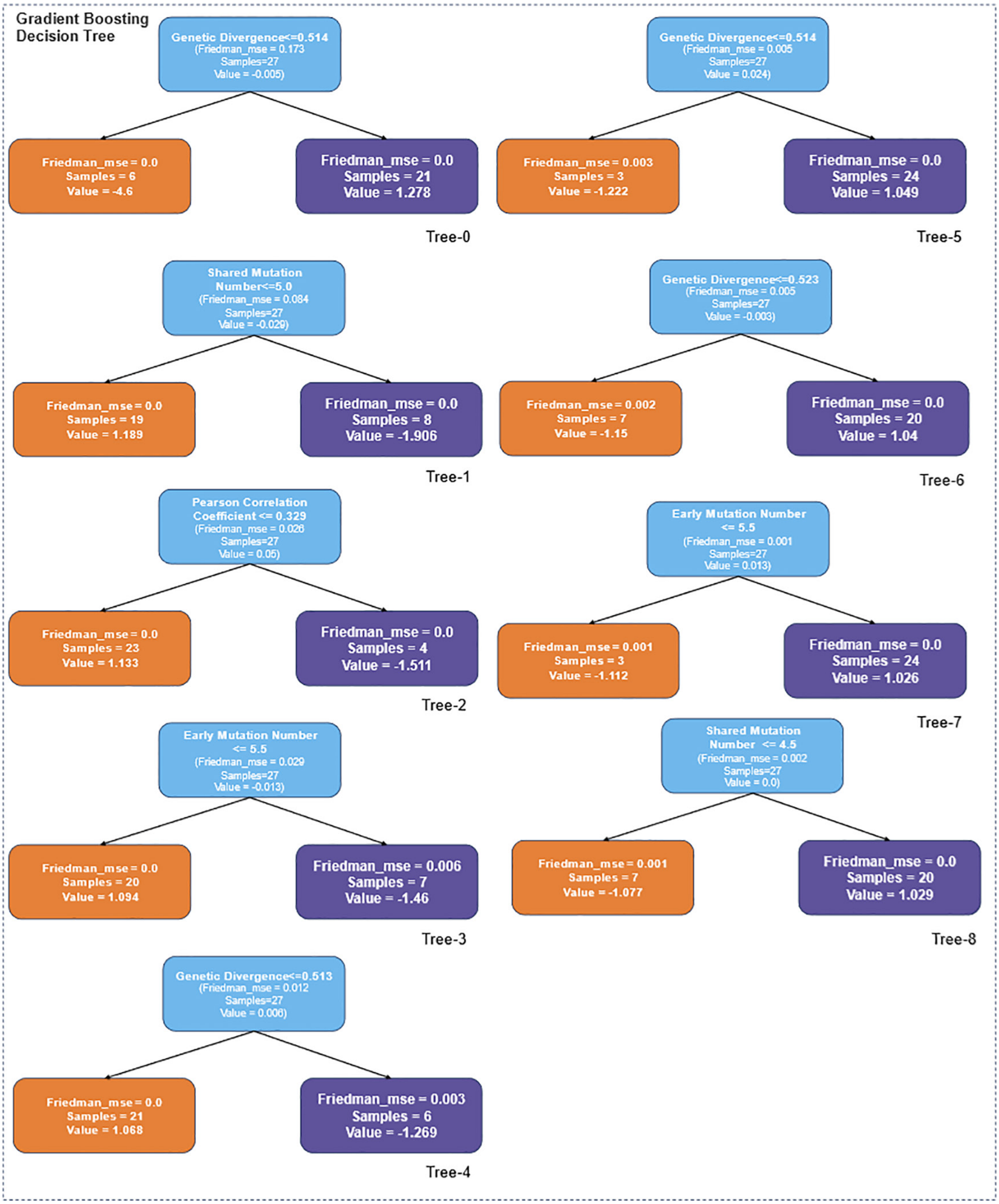
Visualization Results of the Decision Tree



Visualization Results of the Random Forest



Visualization Results of the Gradient Boosting Decision Trees



Abbreviations: MP, multiple primary lung cancer; IM, intrapulmonary metastases.
The corresponding model weight files required to reproduce and facilitate testing on independent datasets is available at the following repository: https://github.com/RocketSharks/ML_models_discriminate_multiple_NSCLCs

Figure S1 Pearson correlation analysis of mutations in paired tumors of MPLC patients in model development cohort.

https://dx.doi.org/10.21037/tlcr-24-875

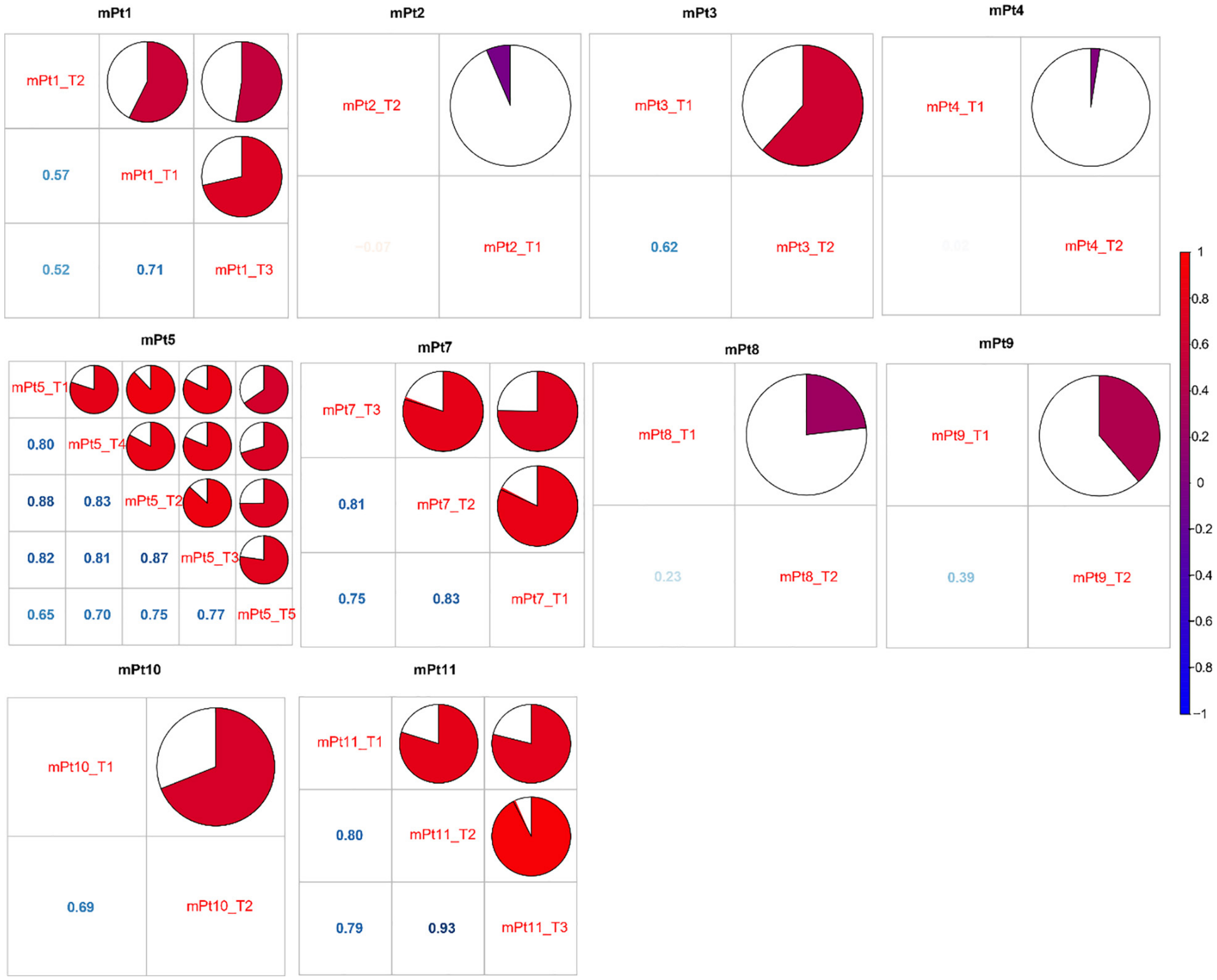


Figure S2 Pearson correlation analysis of mutations in paired tumors of IPM patients in model development cohort.

https://dx.doi.org/10.21037/tlcr-24-875

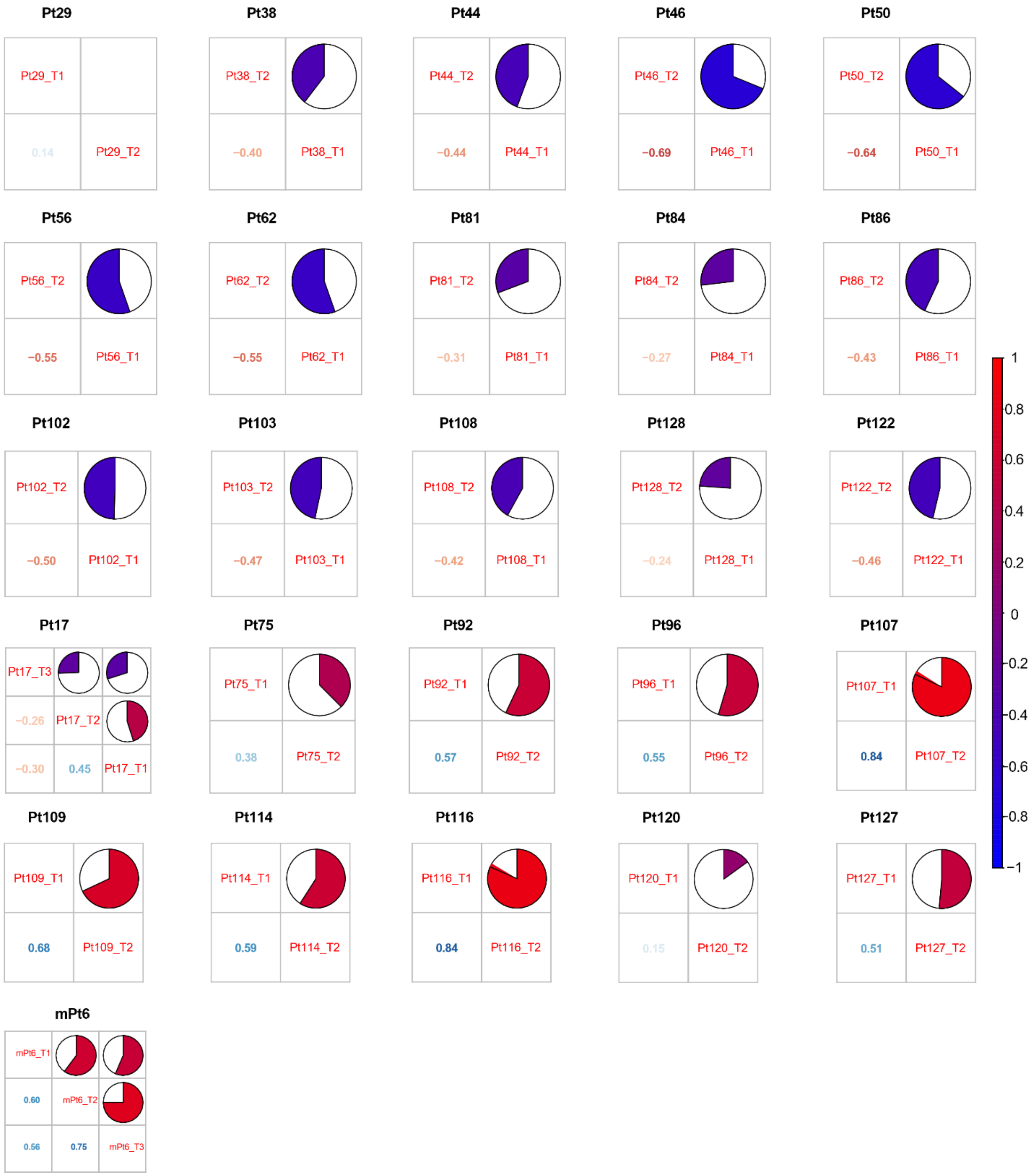


Figure S3 Pearson correlation analysis of mutations in paired tumors in model test cohort.

https://dx.doi.org/10.21037/tlcr-24-875

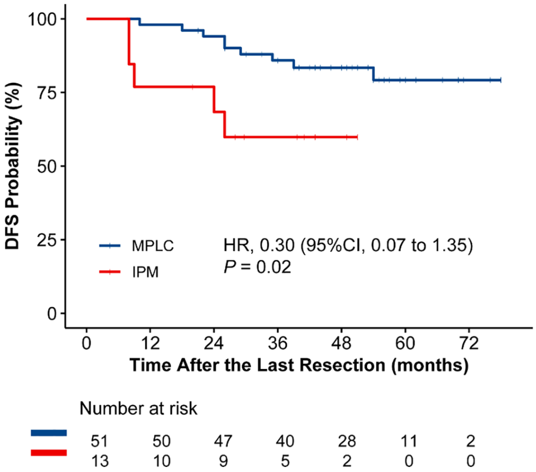


Figure S4 The Kaplan-Meier analysis of disease-free survival curve of MPLC and IPM patients without lymph nodes metastases or extra-thoracic metastasis in this study.

Table S1 Information of tumor samples from MPLC patients in model development cohort

Patients	Sample_ID	Tumor Location	Radiology Features	Tumor Size (cm)	Histology	TP53 mutation	Driver mutations	Synchronous/Metachronous	Surgical resection
Pt3	Pt3_T1	LUL	SN	1	ADC	WT	EGFR p.L858R	Synchronous	segmentectomy
	Pt3_T2	LLL	GGN	0.8	ADC	WT	WT		wedge
Pt12	Pt12_T1	RUL	GGN	0.5	MIA	WT	WT	Synchronous	segmentectomy
	Pt12_T2	RML	PSN	3	ADC	p.R158Afs*16; p.R26Afs*16; p.R119Afs*16	WT		lobectomy
Pt16	Pt16_T1	RLL	GGN	1.2	ADC	WT	WT	Synchronous	lobectomy
	Pt16_T2	RLL	SN	1.9	ADC	p.R248Q; p.R116Q; p.R89Q; p.R209Q	EGFR p.L858R		
Pt21	Pt21_T1	LLL	GGN	2	ADC	WT	EGFR p.L858R	Synchronous	segmentectomy
	Pt21_T2	RUL	SN	4	ADC	WT	EGFR19INS p.I740_K745		segmentectomy
Pt23	Pt23_T1	RUL	GGN	1.5	ADC	WT	EGFR 19Del	Synchronous	lobectomy
	Pt23_T2	RUL	GGN	0.8	ADC	WT	WT		
Pt25	Pt25_T1	RLL	SN	0.8	ADC	WT	WT	Synchronous	
	Pt25_T2	RUL	SN	2	ADC	WT	EGFR p.L858R		lobectomy
Pt32	Pt32_T1	RLL	SN	1	ADC	WT	EGFR p.L858R	Synchronous	lobectomy
	Pt32_T2	RLL	PSN	1	ADC	p.R248W; p.R116W; p.R209W; p.R89W	EGFR 19Del		
Pt33	Pt33_T1	RML	GGN	0.5	MIA	WT	WT	Synchronous	lobectomy
	Pt33_T2	RML	GGN	1.5	ADC	WT	WT		
Pt42	Pt42_T1	RUL	PSN	3	ADC	WT	WT	Synchronous	lobectomy
	Pt42_T2	RUL	GGN	1	AAH/AIS/MIA	WT	BRAF p.G469A		
Pt43	Pt43_T1	LUL	GGN	1.5	ADC	WT	EGFR 19Del	Synchronous	lobectomy
	Pt43_T2	LUL	GGN	0.8	ADC	WT	WT		
Pt45	Pt45_T1	LLL	SN	3.5	ADC	p.M237I; p.M105I; p.M198I; p.M78I	WT	Synchronous	lobectomy
	Pt45_T2	RUL	SN	2.3	ADC	p.R273C; p.R141C; p.R234C; p.R114C	KRAS p.G12C		lobectomy
Pt47	Pt47_T1	RUL	PSN	0.5	ADC	WT	WT	Synchronous	segmentectomy
	Pt47_T2	RLL	PSN	2.5	ADC	WT	EGFR 19Del		lobectomy
Pt49	Pt49_T1	RLL	GGN	0.5	ADC	WT	WT	Synchronous	wedge
	Pt49_T2	LUL	GGN	1.6	ADC	WT	EGFR p.L858R		segmentectomy
Pt53	Pt53_T1	RUL	SN	0.8	ADC	WT	WT	Synchronous	lobectomy
	Pt53_T2	RUL	GGN	1	ADC	WT	EGFR p.L858R		
	Pt53_T3	RUL	GGN	0.5	ADC	WT	WT		
Pt54	Pt54_T1	RML	GGN	2	ADC	WT	KRAS p.Q61L	Metachronous	lobectomy
	Pt54_T2	RUL	PSN	2	ADC	WT	KRAS p.G12F		lobectomy
Pt58	Pt58_T1	RUL	PSN	1.1	ADC	WT	KRAS p.G12V	Synchronous	lobectomy
	Pt58_T2	RUL	GGN	0.5	AIS/MIA	WT	EGFR p.L858R		
	Pt58_T3	RLL	PSN	1.5	ADC	WT	EGFR p.L858R		segmentectomy
Pt60	Pt60_T1	LUL	SN	0.8	ADC	WT	WT	Synchronous	segmentectomy
	Pt60_T2	LLL	GGN	2.5	ADC	p.Y234C; p.Y102C; p.Y195C; p.Y75C	EGFR p.L858R		lobectomy
Pt68	Pt68_T1	RUL	PSN	1.7	ADC	p.E258K p.E126K p.E219K p.E99K p.V73Rfs*76 p.V34Rfs*76	BRAF p.L597Q	Synchronous	segmentectomy
	Pt68_T2	RML	PSN	2.6	ADC	WT	KRAS p.G12S		lobectomy

Table S1 (*continued*)

Table S1 (*continued*)

Patients	Sample_ID	Tumor Location	Radiology Features	Tumor Size (cm)	Histology	TP53 mutation	Driver mutations	Synchronous/Metachronous	Surgical resection
Pt79	Pt79_T1	RUL	SN	3	IMA	WT	EGFR 19Del	Synchronous	lobectomy
	Pt79_T2	RML	SN	1	ADC	WT	KRAS p.G12V		wedge
	Pt79_T3	RLL	GGN	1	ADC	WT	EGFR 20ins; KRAS p.G12A		wedge
Pt83	Pt83_T1	RUL	PSN	2.5	ADC	WT	EGFR 19Del	Synchronous	lobectomy
	Pt83_T2	RUL	PSN	2.5	ADC	WT	EGFR p.L858R		
Pt87	Pt87_T1	RML	SN	2	ADC	p.S241F; p.S109F; p.S202F; p.S82F; p.S202F	EGFR 19Del	Synchronous	lobectomy
	Pt87_T2	LUL	GGN	2.5	ADC	WT	EGFR p.L858R		segmentectomy
Pt93	Pt93_T1	RUL	SN	2	IMA	WT	KRAS p.G13C	Synchronous	wedge
	Pt93_T2	RLL	SN	0.7	ADC	p.Q52*; p.Q13*	WT		segmentectomy
Pt94	Pt94_T1	RUL	SN	4	ADC	WT	EGFR p.L858R	Synchronous	lobectomy
	Pt94_T2	RLL	SN	2	ADC	WT	EGFR 19Del		wedge
Pt98	Pt98_T1	RUL	GGN	1	ADC	WT	EGFR 19Del	Synchronous	segmentectomy
	Pt98_T2	LLL	GGN	3	ADC	WT	EGFR p.L858R; EGFR p.E709K; CTNNB1 p.D32V		segmentectomy
Pt99	Pt99_T1	RLL	GGN	2	AIS/MIA	WT	EGFR p.L858R	Synchronous	lobectomy
	Pt99_T2	LUL	GGN	2	ADC	WT	EGFR p.L858R		wedge
	Pt99_T3	LLL	GGN	0.8	ADC	WT	EGFR 19Del		wedge
Pt104	Pt104_T1	LLL	SN	1.5	ADC	WT	EGFR 19Del	Synchronous	segmentectomy
	Pt104_T2	LUL	PSN	3.1	ADC	WT	KRAS p.G12C		wedge
Pt106	Pt106_T1	LUL	SN	2.7	ADC	WT	EGFR p.L858R	Synchronous	segmentectomy
	Pt106_T2	RML	SN	2	ADC	WT	WT		lobectomy
	Pt106_T3	RLL	GGN	0.6	ADC	WT	WT		segmentectomy
Pt110	Pt110_T1	LLL	SN	3	ADC	WT	WT	Synchronous	lobectomy
	Pt110_T2	RUL	SN	2	SCC	p.V157F; p.V25F; p.V118F	WT		lobectomy
Pt111	Pt111_T1	RUL	SN	2.9	ADC	WT	WT	Synchronous	lobectomy
	Pt111_T2	RML	SN	1.5	ADC	WT	KRAS p.G12A		lobectomy
Pt115	Pt115_T1	RLL	PSN	2	ADC	WT	EGFR 19Del	Synchronous	segmentectomy
	Pt115_T2	RUL	SN	1.5	ADC	WT	EGFR p.L858R		lobectomy
Pt117	Pt117_T1	RUL	SN	2.5	ADC	WT	EGFR p.L861Q	Synchronous	segmentectomy
	Pt117_T2	RLL	SN	2	ADC	WT	EGFR p.L858R		segmentectomy
Pt118	Pt118_T1	LUL	PSN	3	ADC	p.E339*; p.E207*; p.E300*; p.E180*;	EGFR p.L858R	Synchronous	segmentectomy
	Pt118_T2	LLL	SN	4	ADC	p.G245V; p.G113V; p.G206V; p.G86V	EGFR p.V774M; EGFR p.H773L		segmentectomy
Pt119	Pt119_T1	LUL	PSN	2.5	ADC	p.C135F; p.C3F; p.C96F	EGFR p.G719S	Synchronous	lobectomy
	Pt119_T2	LLL	SN	1	ADC	WT	EGFR p.L858R; PIK3CA p.C420R		wedge
Pt123	Pt123_T1	RUL	PSN	2.3	ADC	WT	EGFR p.L858R	Metachronous	lobectomy
	Pt123_T2	LUL	PSN	0.9	ADC	WT	WT		segmentectomy
Pt124	Pt124_T1	RUL	GGN	1.5	ADC	WT	MAP2K1 p.E102_I103delEI	Synchronous	lobectomy
	Pt124_T2	RUL	GGN	2	AIS/MIA	WT	WT		
Pt125	Pt125_T1	LUL	GGN	1.6	ADC	WT	KRAS p.G12A	Synchronous	wedge
	Pt125_T2	LLL	SN	3	ADC	WT	KRAS p.G12V		lobectomy

Abbreviations: T1, tumor 1; T2, tumor 2;T3, tumor 3; LLL, left low lung lobe; LUL, left up lung lobe; RUL, right up lung lobe; RML, right middle lung lobe; RLL, right low lung lobe; GGN, ground-glass nodule; SN, solid nodule; PSN, part-solid nodule; AAH, atypical adenomatous hyperplasia; AIS, adenocarcinoma in situ; MIA, minimally invasive adenocarcinoma; IAC, invasive adenocarcinoma; ADC, adenocarcinoma; SCC, squamous cell carcinoma; WT, Wide type.

Table S2 Information of tumor samples from IPM patients in model development cohort

Patients	Sample_ID	Tumor Location	Radiology Features	Tumor Size (cm)	Histology	TP53 Mutation	Driver Mutations	Synchronous/Metachronous	Surgical resection
mPt1	mPt1_T1	RUL	SN	4	ADC	WT	WT	Synchronous	wedge
	mPt1_T2	RLL	SN	2	ADC	p.R273L; p.R141L; p.R234L; p.R114L	WT		wedge
	mPt1_T3	pleural	SN	2	metastatic carcinoma	WT	WT		
mPt2	mPt2_T1	RML	SN	5	ADC	WT	EGFR p.G719C; EGFR p.S768I	Synchronous	lobectomy
	mPt2_T2	pleural	SN	2	metastatic carcinoma	WT	EGFR p.G719C; EGFR p.S768I		
mPt3	mPt3_T1	RUL	SN	7	ADC	WT	WT	Synchronous	lobectomy
	mPt3_T2	pleural	SN	2	ADC	WT	WT		
mPt4	mPt4_T1	RUL	SN	5.4	ADC	WT	EGFR 19Del	Synchronous	lobectomy
	mPt4_T2	pleural	SN	1.6	metastatic carcinoma	WT	EGFR 19Del		
mPt5	mPt5_T1	LUL	SN	4.2	ADC	WT	KRAS p.G12D	Synchronous	NA
	mPt5_T2	LUL	SN	0.7	ADC	WT	KRAS p.G12D		
	mPt5_T3	LUL	SN	2.5	ADC	WT	KRAS p.G12D		
	mPt5_T4	LLL	SN	0.8	ADC	WT	KRAS p.G12D		
	mPt5_T5	pleural	SN	1.7	ADC	WT	KRAS p.G12D		
mPt7	mPt7_T1	RUL	SN	3	ADC	WT	EGFR p.L858R	Synchronous	lobectomy
	mPt7_T2	RUL	SN	0.7	ADC	WT	EGFR p.L858R		
	mPt7_T3	RUL	SN	0.6	ADC	WT	EGFR p.L858R		
mPt8	mPt8_T1	RUL	SN	3	ADC	p.Y163C; p.Y31C; p.Y124C; p.Y4C	EGFR 19Del	Synchronous	lobectomy
	mPt8_T2	RML	SN	1	ADC	p.Y163C; p.Y31C; p.Y124C; p.Y4C	EGFR 19Del		
mPt9	mPt9_T1	RUL	SN	2	ADC	WT	EGFR p.L858R	Synchronous	lobectomy
	mPt9_T2	RLL	SN	2	ADC	WT	EGFR p.L858R		wedge
mPt10	mPt10_T1	RML	SN	2.5	ADC	p.Y220C; p.Y88C; p.Y181C; p.Y61C	EGFR 19Del	Synchronous	lobectomy
	mPt10_T2	RLL	SN	1	ADC	p.Y220C; p.Y88C; p.Y181C; p.Y61C	EGFR 19Del		segmentectomy
mPt11	mPt11_T1	LLL	SN	3.5	ADC	WT	ERBB2 20ins	Synchronous	lobectomy
	mPt11_T2	LLL	SN	1.5	ADC	WT	ERBB2 20ins		
	mPt11_T3	LLL	SN	1	ADC	WT	ERBB2 20ins		

Abbreviations: T1, tumor 1; T2, tumor 2;T3, tumor 3; LLL, left low lung lobe; LUL, left up lung lobe; RUL, right up lung lobe; RML, right middle lung lobe; RLL, right low lung lobe; GGN, ground-glass nodule; SN, solid nodule; PSN, part-solid nodule; AAH, atypical adenomatous hyperplasia; AIS, adenocarcinoma in situ; MIA, minimally invasive adenocarcinoma; IAC, invasive adenocarcinoma; ADC, adenocarcinoma; SCC, squamous cell carcinoma; WT, Wide type; NA, not avaliable.

Table S3 Information of tumor samples from patients in model test cohort

Patients	Sample_ID	Tumor Location	Radiology Features	Tumor Size (cm)	Histology	TP53 Mutation	Driver Mutations	Synchronous/Metachronous	Surgical resection
Pt17	Pt17_T1	RUL	PSN	4	ADC	WT	EGFR p.L858R	Synchronous	lobectomy
	Pt17_T2	RUL	GGN	1	ADC	WT	EGFR p.L858R		
	Pt17_T3	RML	SN	1	ADC	WT	EGFR p.L858R		wedge
Pt29	Pt29_T1	RLL	PSN	1.4	ADC	WT	EGFR p.L858R	Synchronous	wedge
	Pt29_T2	LUL	PSN	2.1	ADC	WT	EGFR p.L858R		segmentectomy
Pt38	Pt38_T1	RUL	PSN	1.5	ADC	WT	EGFR p.L858R	Synchronous	lobectomy
	Pt38_T2	RML	SN	2	ADC	p.C275F; p.C143F; p.C236F; p.C116F	EGFR p.L858R		wedge
Pt44	Pt44_T1	RLL	SN	0.8	ADC	WT	WT	Synchronous	lobectomy
	Pt44_T2	RLL	SN	3	ADC	WT	WT		
Pt46	Pt46_T1	RUL	SN	3	SCC	p.R213Dfs*34; p.R81Dfs*34; p.R174Dfs*34; p.R54Dfs*34	WT	Metachronous	lobectomy
	Pt46_T2	RML	SN	1.7	SCC	p.E298*; p.E166*; p.E259*; p.E139*	WT		lobectomy
Pt50	Pt50_T1	LUL	SN	2	SCC	p.H179Y; p.H47Y; p.H140Y; p.H20Y	WT	Synchronous	lobectomy
	Pt50_T2	LUL	GGN	0.5	SCC	p.G245R; p.G113R; p.G206R; p.G86R	WT		
Pt56	Pt56_T1	RUL	PSN	1.2	ADC	p.A159S; p.A27S; p.A120S; p.V157G; p.V25G; p.V118G	EGFR p.L858R	Synchronous	lobectomy
	Pt56_T2	RLL	PSN	1.5	ADC	WT	EGFR p.L858R; EGFR p.T790M		lobectomy
Pt62	Pt62_T1	RUL	SN	3.5	ADC	p.R196*; p.R64*; p.R157*; p.R37*	EGFR p.L858R	Synchronous	lobectomy
	Pt62_T2	RLL	GGN	1	ADC	WT	EGFR p.L858R		segmentectomy
Pt75	Pt75_T1	RML	SN	2	ADC	p.R248L; p.R116L; p.R209L; p.R89L; p.R209L	EGFR 20ins	Synchronous	lobectomy
	Pt75_T2	RLL	SN	1	ADC	p.R248L; p.R116L; p.R209L; p.R89L; p.R209L	EGFR 20ins		segmentectomy
Pt81	Pt81_T1	LUL	SN	3	ADC	WT	EGFR 19Del	Metachronous	lobectomy
	Pt81_T2	RLL	SN	1.5	ADC	p.Y126C; p.Y87C	EGFR 19Del		segmentectomy
Pt84	Pt84_T1	LLL	PSN	1.5	ADC	WT	EGFR p.L858R	Synchronous	lobectomy
	Pt84_T2	LLL	SN	2	ADC	WT	EGFR p.L858R		
Pt86	Pt86_T1	RUL	GGN	2	ADC	WT	EGFR p.L858R	Synchronous	wedge
	Pt86_T2	LUL	GGN	1.5	ADC	WT	EGFR p.L858R; EGFR p.T790M		wedge
Pt92	Pt92_T1	RUL	SN	4	ADC	p.R342*; p.R210*; p.R303*; p.R183*; p.R303*	WT	Metachronous	lobectomy
	Pt92_T2	RML	SN	4	ADC	p.R342*; p.R210*; p.R303*; p.R183*; p.R303*	WT		lobectomy
Pt96	Pt96_T1	RLL	GGN	1	ADC	WT	EGFR p.L858R	Synchronous	lobectomy
	Pt96_T2	RLL	GGN	1	ADC	WT	EGFR p.L858R		
Pt102	Pt102_T1	RLL	SN	1	ADC	WT	EGFR 19Del	Synchronous	wedge
	Pt102_T2	RUL	SN	2	ADC	WT	EGFR 19Del		lobectomy
Pt103	Pt103_T1	LUL	PSN	2.5	ADC	WT	EGFR p.L858R; EGFR p.R776H	Synchronous	lobectomy
	Pt103_T2	LLL	SN	1	ADC	WT	EGFR p.L858R		wedge
Pt107	Pt107_T1	LUL	SN	2	ADC	p.S241Pfs*6; p.S109Pfs*6; p.S202Pfs*6; p.S82Pfs*6	EGFR p.L858R	Synchronous	lobectomy
	Pt107_T2	LUL	SN	0.9	ADC	p.S241Pfs*6; p.S109Pfs*6; p.S202Pfs*6; p.S82Pfs*6	EGFR p.L858R		

Table S3 (*continued*)

Table S3 *(continued)*

Patients	Sample_ID	Tumor Location	Radiology Features	Tumor Size (cm)	Histology	TP53 Mutation	Driver Mutations	Synchronous/Metachronous	Surgical resection
Pt108	Pt108_T1	RLL	SN	2.5	ADC	WT	EGFR p.L858R; EGFR p.T790M	Synchronous	segmentectomy
	Pt108_T2	RUL	SN	2.1	ADC	WT	EGFR p.L858R		segmentectomy
Pt109	Pt109_T1	RLL	SN	1.5	ADC	WT	WT	Synchronous	wedge
	Pt109_T2	LLL	SN	4	ADC	WT	WT		lobectomy
Pt114	Pt114_T1	RLL	SN	1.5	ADC	WT	WT	Synchronous	segmentectomy
	Pt114_T2	RLL	SN	1.5	ADC	WT	WT		
Pt116	Pt116_T1	LUL	SN	2.5	SCC	p.Y220C; p.Y88C; p.Y181C; p.Y61C	WT	Synchronous	lobectomy
	Pt116_T2	LUL	SN	2.5	SCC	p.Y220C; p.Y88C; p.Y181C; p.Y61C	WT		
Pt120	Pt120_T1	RML	SN	1.9	ADC	WT	ERBB2 20ins	Synchronous	lobectomy
	Pt120_T2	RLL	SN	2.2	ADC	WT	ERBB2 20ins		segmentectomy
Pt122	Pt122_T1	RML	SN	3	ADC	WT	EGFR p.L858R	Synchronous	lobectomy
	Pt122_T2	LUL	GGN	2	ADC	WT	EGFR p.L858R		segmentectomy
Pt127	Pt127_T1	RUL	GGN	2	ADC	p.R273C; p.R141C; p.R234C; p.R114C	EGFR 19Del	Synchronous	lobectomy
	Pt127_T2	RUL	SN	1.2	ADC	p.R273C; p.R141C; p.R234C; p.R114C	EGFR 19Del		
Pt128	Pt128_T1	RUL	PSN	1.8	ADC	WT	EGFR p.L858R	Synchronous	wedge
	Pt128_T2	RLL	PSN	1.8	ADC	p.V157F; p.V25F; p.V118F	EGFR p.L858R		lobectomy
mPt6	mPt6_T1	RML	SN	2.5	ADC	WT	WT	Synchronous	wedge
	mPt6_T2	RLL	SN	1.2	ADC	WT	WT		wedge
	mPt6_T3	RLL	SN	1.5	ADC	WT	WT		wedge

Abbreviations: T1, tumor 1; T2, tumor 2;T3, tumor 3; T4, tumor 4;T5, tumor 5; LLL, left low lung lobe; LUL, left up lung lobe; RUL, right up lung lobe; RML, right middle lung lobe; RLL, right low lung lobe; GGN, ground-glass nodule; SN, solid nodule; PSN, part-solid nodule; AAH, atypical adenomatous hyperplasia; AIS, adenocarcinoma in situ; MIA, minimally invasive adenocarcinoma; IAC, invasive adenocarcinoma; ADC, adenocarcinoma; SCC, squamous cell carcinoma; WT, Wide type.

Table S4 Information of MPLC patients with a family history of cancer and their corresponding germline alteration genes

Patients	Family cancer	P/LP	Germline Mutation Genes
Pt3	lung cancer (younger sister)	positive	*FLG, NCAM2, SLC12A3*
Pt21	rectal cancer (mother)	negative	–
Pt32	lung cancer (younger brother)	negative	–
Pt38	colon cancer (mother)	positive	*LMNA, SLC22A5*
Pt42	lung cancer (father, younger sister)	positive	*EARS2, RGS16*
Pt43	lung cancer (younger brother)	positive	*APBB1*
Pt45	pancreatic cancer (older sister)	positive	*SLC34A3, SRD5A2*
Pt47	esophageal cancer (father)	negative	–
Pt54	gastric cancer (father)	positive	*ACADVL, LY96*
Pt58	lung cancer (mother, sister, brother, paternal uncle), liver cancer (father)	positive	*FLG, ROBO2, SLC52A3*
Pt60	gastric cancer, renal cell carcinoma, prostate cancer (father)	positive	*TBX6*
Pt62	lung cancer (daughter)	negative	–
Pt79	lung cancer (daughter)	negative	–
Pt81	lung cancer (brother, paternal aunt, paternal cousin)	negative	–
Pt83	esophageal cancer (brother)	negative	–
Pt87	lung cancer (older sister)	negative	–
Pt108	colon cancer (younger sister)	positive	*BUB1B*
Pt110	breast cancer (mother)	negative	–
Pt115	lung cancer (younger brother), pancreatic cancer (older brother)	positive	*GJB2*
Pt118	lung cancer (older brother)	positive	*CYP4V2*
Pt119	lung cancer (mother, older sister, older brother)	positive	*ABCA4, BSND*
Pt123	lung cancer (younger sister)	positive	*SCRIB, SLC22A5*
mPt4	lung cancer (older sister)	negative	–

Table S5 Correlation between MPLC patient clinical information and germline mutation prevalence

	Univariate analysis		
	OR	95% CI	P
Age of diagnosis	1.06	0.96–1.16	0.24
Sex (male *vs.* female)	6.84	0.61–76.91	0.119
Smoking status (never *vs.* ever)	12.26	0.96–156.66	0.054
Family history of cancer (yes *vs.* no)	0.21	0.05–0.97	0.046
Dual tumors (yes *vs.* no)	4.02	0.4–40.2	0.237

https://dx.doi.org/10.21037/tlcr-24-875