

SUPPLEMENTARY INFORMATION – Content list

TABLES AND FIGURES

- Table S1- 46 influential metabolomic biomarkers with statistical significance of Stage 1 lung tumor patients (mean value with SD).
- Table S2 - PCA of 10 metabolomic biomarkers in early lung tumor detection.
- Table S3 - The logistic regression models of different combined variates.
- Table S4 - 50 influential metabolomic biomarkers with statistical significance of lung tumor patients (mean value with SD).
- Table S5 - Non-parametric test of the alteration of metabolomic biomarkers in stage progression.
- Table S6 - The AUC values of ROC curve of metabolomic biomarkers in each tumor stage.
- Table S7 - Rank metabolomic biomarker features by FCBF.
- Table S8 - AUC value of machine learning models changing with the number of variates.
- Figure S1 - ROC analysis of 10 metabolomic biomarkers in early lung tumor detection.
- Figure S2 - 50 influential metabolomic biomarkers with statistical significance of lung tumor patients (mean value with SD).
- Figure S3 - ROC analysis of metabolomic biomarkers with statistical significance of adenocarcinoma and squamous carcinoma patients.
- Figure S4 - AUC value of machine learning models changing with the number of variates.

Table S1.46 influential metabolomic biomarkers with statistical significance of Stage 1 lung tumor patients(mean value with SD).

	Stage 1 lung cancer patients		Healthy individuals			Stage 1 lung cancer patients		Healthy individuals	
	Mean	SD	Mean	SD		Mean	SD	Mean	SD
GSH	0.0552	0.08758	0.0239	0.02189	Trigonelline	0.0624	0.0313	0.0927	0.0323
L-Leucine	207.8333	59.89795	173.2221	30.99564	LPC(16:0)	95.7481	37.55637	66.0326	22.60928
L-Kynurenine	1.4764	0.62948	0.9473	0.33082	LPC(18:0)	102.2146	47.90141	69.1203	37.94514
Phenylalanine	162.5185	47.96366	113.8605	21.37058	L-Glutamic acid	0.3021	0.18109	0.1995	0.04863
Creatine	0.8021	0.22692	0.6707	0.17243	PC(36:5)	7.0837	10.52985	5.7728	2.34784
L-Valine	207.5741	89.99405	126.9628	24.34529	PC(36:3)	6.1196	6.37649	7.3216	2.03146
Proline	42.2833	27.18405	15.9	5.4435	PC(36:4)	4.9199	4.53674	8.2205	2.01927
Spermidine	8.7902	2.86546	5.6814	1.15928	PE(36:4)	10.2983	11.63711	14.0449	3.40511
Ascorbic acid	1.4679	0.43496	1.1425	0.27492	Glycochenodeoxycholic acid	7.0499	5.73509	3.4542	2.11454
2-DG	0.0335	0.01979	0.0702	0.02361	Hexanoylcarnitine	10.4643	6.03788	5.3372	2.48598
Cl-Tyr	0.2859	0.15553	0.4451	0.19478	Citric acid	23.847	10.98938	45.7279	17.16728
o-Tyr	31.313	9.14382	22.6116	5.51842	Hydroxybutanedioic acid	1.709	0.94889	2.0154	0.62638
Choline	0.4156	0.21653	0.2517	0.06625	Succinic acid	3.165	1.22967	4.4542	1.11105
Serine	0.4407	0.21219	0.2926	0.07357	Oxaloacetic acid	0.2947	0.27572	0.3724	0.27388
Aminohippuric acid	6.632	4.39631	3.1251	0.74607	(E)-butenedioic acid	59.3204	15.38841	77.6209	17.93606
Palmitoyl-L-carnitine	5.4593	2.83079	1.8756	0.8058	Uric acid	36.9574	9.74601	32.7465	9.05944
4,6-dihydroxyquinoline	1.9922	0.42138	1.7496	0.24816	indole-3-lactic acid	2.615	0.97203	1.9298	0.58694
Carnitine	6.658	3.27084	3.5702	0.75905	Glucose	17.2241	12.17761	17.8014	4.66604
C16:1LPC	3.5159	2.32108	1.084	0.9121	lactate	200.1296	78.1042	216.7907	51.65387
C18:1LPC	43.6193	26.4116	25.613	15.96863	Xanthine	0.3308	0.17033	0.5212	0.14841
Xanthosine	0.0853	0.07944	0.0514	0.0237	Fumaric acid	26.7704	5.16558	37.1465	10.1389
Taurine	104.6593	36.71764	55.9512	10.89002	Ethylmalonic acid	0.419	0.38368	0.5725	0.43121
Inosine	1.1724	0.87956	1.5929	0.95401	PLA	22.6422	23.33959	39.6047	24.10801

Abbreviations: SD, standard deviation.

Table S2. PCA of 10 metabolomic biomarkers in early lung tumor detection.

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	6.224	62.236	62.236	6.224	62.236	62.236
2	1.294	12.945	75.181	1.294	12.945	75.181
3	.807	8.072	83.253			
4	.618	6.176	89.429			
5	.375	3.745	93.175			
6	.273	2.731	95.905			
7	.190	1.902	97.807			
8	.112	1.121	98.928			
9	.084	.840	99.768			
10	.023	.232	100.000			

Abbreviations: PCA, principal component analysis.

Table S3. The logistic regression models of different combined variates.

Logistic regression model	
Combination of two	$\text{Log}(P/(1-P)) = (-10.105) + (0.948) \text{ Spermidine} + (0.980) \text{ Amino-hippuric acid}$
Combination of three	$\text{Log}(P/(1-P)) = (-11.443) + (0.123) \text{ Proline} + (0.756) \text{ Palmitoyl-L-carnitine} + (0.094) \text{ Taurine.}$
Combination of six	$\text{Log}(P/(1-P)) = (-13.455) + (3.897) \text{ L-Kynurenine} + (0.322) \text{ Proline} + (-6.503) \text{ Spermidine} + (-2.309) \text{ Amino-hippuric acid} + (1.186) \text{ Palmitoyl-L-carnitine} + (0.745) \text{ Taurine}$
Combination of four	$\text{Log}(P/(1-P)) = (0.78) + (-1.633) \text{ L-Kynurenine} + (0.055) \text{ Proline} + (-0.268) \text{ Carnitine} + (13.086) \text{ Hypoxanthine.}$

Table S4. 50 influential metabolomic biomarkers with statistical significance of lung tumor patients (mean value with SD).

	Lung tumor patients		Healthy individuals			Lung tumor patients		Healthy individuals	
	Mean	SD	Mean	SD		Mean	SD	Mean	SD
GSH	0.0588	0.07833	0.0239	0.02189	LPC(16:0)	91.505	34.29612	66.0326	22.60928
L-Leucine	204.5373	61.23687	173.2221	30.99564	LPC(18:0)	95.1173	43.91685	69.1203	37.94514
L-Kynurenine	3.0125	16.5038	0.9473	0.33082	LPC(20:1)	2.8345	3.60471	2.2443	2.34914
Phenylalanine	158.7413	47.57185	113.8605	21.37058	L-Glutamic acid	0.3041	0.21202	0.1995	0.04863
Creatine	1.9993	13.254	0.6707	0.17243	PC(36:5)	7.0407	9.79651	5.7728	2.34784
L-Valine	203.2783	89.12877	126.9628	24.34529	PC(36:3)	6.77	7.03821	7.3216	2.03146
Proline	44.1586	29.45611	15.9	5.4435	PC(36:4)	5.0732	4.33515	8.2205	2.01927
Spermidine	8.6106	2.99739	5.6814	1.15928	PE(36:4)	11.7863	12.51869	14.0449	3.40511
Ascorbic acid	1.521	0.63584	1.1425	0.27492	Glycochenodeoxycholic acid	6.4763	5.49013	3.4542	2.11454
2-DG	0.0453	0.10691	0.0702	0.02361	Hexanoylcarnitine	9.7603	6.30752	5.3372	2.48598
Cl-Tyr	0.3141	0.17033	0.4451	0.19478	Citric acid	22.2327	10.01918	45.7279	17.16728
o-Tyr	31.9025	9.49705	22.6116	5.51842	Hydroxybutanedioic acid	1.7823	1.66618	2.0154	0.62638
Choline	0.7088	2.99064	0.2517	0.06625	Succinic acid	3.0168	1.06839	4.4542	1.11105
Serine	0.5142	0.9124	0.2926	0.07357	Oxaloacetic acid	0.3337	0.46982	0.3724	0.27388
Aminohippuric acid	7.0862	4.61527	3.1251	0.74607	(E)-butenedioic acid	57.3241	13.80873	77.6209	17.93606
Palmitoyl-L-carnitine	5.1665	2.58248	1.8756	0.8058	Uric acid	36.218	8.86447	32.7465	9.05944
4,6-dihydroxyquinoline	1.9545	0.44431	1.7496	0.24816	Uridine	10.7273	6.61985	7.3944	3.58673
Carnitine	6.4605	3.00478	3.5702	0.75905	indole-3-lactic acid	2.6793	1.14241	1.9298	0.58694
C16:1LPC	3.3382	2.2814	1.804	0.9121	Glucose	17.2445	11.68526	17.8014	4.66604
C18:1LPC	40.6131	23.57834	25.613	15.96863	lactate	196.7984	77.855	216.7907	51.65387
Xanthosine	0.3808	3.15846	0.0514	0.0237	Xanthine	1.882	16.69688	0.5212	0.14841
Taurine	100.8773	39.28674	55.9512	10.89002	Fumaric acid	26.1215	5.44013	37.1465	10.1389
Inosine	1.4477	2.49899	1.5929	0.95401	3-hydroxybutyric	9.8314	10.12331	10.4116	5.65005
Trigonelline	0.0666	0.04026	0.0927	0.0323	Ethylmalonic acid	0.4359	0.42776	0.5725	0.43121
LPE(16:0)	1.0771	0.68268	1.2147	0.51679	PLA	23.3821	25.05129	39.6047	24.10801

Abbreviations: SD, standard deviation.

Table S5.Non-parametric test of the alteration of metabolomic biomarkers in stage progression.

		Test statistic	Std. Error	Std. Test Statistic	Sig.	p-Value of Kruskal-Wallis		Test statistic	Std. Error	Std. Test Statistic	Sig.	p-Value of Kruskal-Wallis	
L-Kynurenine	Healthy people-Stage 1	50.808	9.174	5.538	0.000	0.000	o-Tyr	Healthy people-Stage 1	49.154	9.174	5.358	0.000	0.000
	Healthy people-Stage 2	46.665	10.678	4.370	0.000	0.000		Healthy people-Stage 2	62.799	10.678	5.881	0.000	0.000
	Healthy people-Stage 3	54.785	11.289	4.853	0.000	0.000		Healthy people-Stage 3	51.349	11.289	4.548	0.000	0.000
	Healthy people-Stage 4	72.132	26.804	2.691	0.007	0.071		Healthy people-Stage 4	14.849	26.804	0.554	0.580	1.000
	Stage 2-Stage 1	4.143	10.221	0.405	0.685	1.000		Stage 2-Stage 1	-13.644	10.221	-1.335	0.182	1.000
	Stage 2-Stage3	-8.120	12.155	-0.668	0.504	1.000		Stage 2-Stage3	11.450	12.155	0.942	0.346	1.000
	Stage 2-Stage 4	-25.467	27.180	-0.937	0.349	1.000		Stage 2-Stage 4	47.950	27.180	1.764	0.078	0.777
	Stage 1-Stage 3	-3.977	10.858	-0.366	0.714	1.000		Stage 1-Stage 3	-2.194	10.858	-0.200	0.840	1.000
	Stage 1-Stage 4	-21.324	26.625	-0.801	0.423	1.000		Stage 1-Stage 4	34.306	26.626	1.288	0.198	1.000
	Stage 3-Stage 4	-17.347	27.426	-0.632	0.527	1.000		Stage 3-Stage 4	36.500	27.426	1.331	0.183	1.000
Phenylalanine	Healthy people-Stage 1	54.327	9.174	5.922	0.000	0.000	Amino-hippuric acid	Healthy people-Stage 1	47.702	9.175	5.199	0.000	0.000
	Healthy people-Stage 2	50.891	10.678	4.766	0.000	0.000		Healthy people-Stage 2	63.135	10.678	5.913	0.000	0.000
	Healthy people-Stage 3	52.958	11.289	4.691	0.000	0.000		Healthy people-Stage 3	53.275	11.290	4.719	0.000	0.000
	Healthy people-Stage 4	12.391	26.803	0.462	0.644	1.000		Healthy people-Stage 4	31.202	26.805	1.164	0.244	1.000
	Stage 2-Stage 1	3.435	10.221	0.336	0.737	1.000		Stage 2-Stage 1	-15.433	10.221	-1.510	0.131	1.000
	Stage 2-Stage3	-2.067	12.155	-0.170	0.865	1.000		Stage 2-Stage3	9.860	12.156	0.811	0.417	1.000
	Stage 2-Stage 4	38.500	27.179	1.417	0.157	1.000		Stage 2-Stage 4	31.933	27.181	1.175	0.240	1.000
	Stage 1-Stage 3	1.369	10.858	0.126	0.900	1.000		Stage 1-Stage 3	-5.573	10.859	-0.513	0.608	1.000
	Stage 1-Stage 4	41.935	26.625	1.575	0.115	1.000		Stage 1-Stage 4	16.500	26.626	0.620	0.535	1.000
	Stage 3-Stage 4	40.567	27.425	1.479	0.139	1.000		Stage 3-Stage 4	22.073	27.427	0.805	0.421	1.000
L-Valine	Healthy people-Stage 1	57.012	9.174	6.214	0.000	0.000	Palmitoy-L-carnitine	Healthy people-Stage 1	68.038	9.175	7.416	0.000	0.000
	Healthy people-Stage 2	61.338	10.678	5.745	0.000	0.000		Healthy people-Stage 2	66.290	10.678	6.208	0.000	0.000
	Healthy people-Stage 3	51.545	11.289	4.566	0.000	0.000		Healthy people-Stage 3	56.980	5.047	5.047	0.000	0.000
	Healthy people-Stage 4	28.771	26.804	1.073	0.283	1.000		Healthy people-Stage 4	73.140	2.729	2.729	0.006	0.064
	Stage 2-Stage 1	-4.326	10.221	-0.423	0.672	1.000		Stage 2-Stage 1	1.748	10.221	0.171	0.864	1.000
	Stage 2-Stage3	9.793	12.155	0.806	0.420	1.000		Stage 2-Stage3	9.310	12.156	0.766	0.444	1.000
	Stage 2-Stage 4	32.567	27.180	1.198	0.231	1.000		Stage 2-Stage 4	-6.850	27.181	-0.252	0.801	1.000
	Stage 1-Stage 3	5.467	10.858	0.504	0.615	1.000		Stage 1-Stage 3	11.058	10.859	1.018	0.309	1.000
	Stage 1-Stage 4	28.241	26.625	1.061	0.289	1.000		Stage 1-Stage 4	-5.102	26.626	-0.192	0.848	1.000
	Stage 3-Stage 4	22.773	27.426	0.830	0.406	1.000		Stage 3-Stage 4	-16.160	27.427	-0.589	0.556	1.000
Proline	Healthy people-Stage 1	65.201	9.175	7.107	0.000	0.000	Carnitine	Healthy people-Stage 1	55.967	9.175	6.100	0.000	0.000
	Healthy people-Stage 2	73.829	10.678	6.914	0.000	0.000		Healthy people-Stage 2	60.714	10.678	5.686	0.000	0.000
	Healthy people-Stage 3	56.495	26.805	3.428	0.000	0.000		Healthy people-Stage 3	48.270	11.290	4.276	0.000	0.000
	Healthy people-Stage 4	91.895	11.290	5.004	0.001	0.006		Healthy people-Stage 4	46.264	26.805	1.726	0.084	0.844
	Stage 2-Stage 1	-8.628	10.221	-0.844	0.399	1.000		Stage 2-Stage 1	-4.746	10.221	-0.464	0.642	1.000
	Stage 2-Stage3	17.333	12.156	1.426	0.154	1.000		Stage 2-Stage3	12.443	12.156	1.024	0.306	1.000
	Stage 2-Stage 4	-18.067	27.181	-0.665	0.506	1.000		Stage 2-Stage 4	14.450	27.181	0.532	0.595	1.000
	Stage 1-Stage 3	8.706	10.859	0.802	0.423	1.000		Stage 1-Stage 3	7.697	10.859	0.709	0.478	1.000
	Stage 1-Stage 4	-26.694	26.626	-1.003	0.316	1.000		Stage 1-Stage 4	9.704	26.626	0.364	0.716	1.000
	Stage 3-Stage 4	-35.400	27.427	-1.291	0.197	1.000		Stage 3-Stage 4	2.007	27.427	0.073	0.942	1.000
Spermidine	Healthy people-Stage 1	61.042	9.174	6.653	0.000	0.000	Taurine	Healthy people-Stage 1	68.247	9.174	7.439	0.000	0.000
	Healthy people-Stage 2	53.386	10.678	5.000	0.000	0.000		Healthy people-Stage 2	58.586	10.678	5.487	0.000	0.000
	Healthy people-Stage 3	54.750	11.290	4.850	0.000	0.000		Healthy people-Stage 3	61.262	11.290	5.426	0.000	0.000
	Healthy people-Stage 4	7.236	26.805	0.270	0.787	1.000		Healthy people-Stage 4	15.802	26.805	0.590	0.555	1.000
	Stage 2-Stage 1	7.656	10.221	0.749	0.454	1.000		Stage 2-Stage 1	9.661	10.221	0.945	0.345	1.000
	Stage 2-Stage3	-1.363	12.156	-0.112	0.911	1.000		Stage 2-Stage3	-2.677	12.156	-0.220	0.826	1.000
	Stage 2-Stage 4	46.150	27.181	1.698	0.090	0.895		Stage 2-Stage 4	42.783	27.181	1.574	0.115	1.000
	Stage 1-Stage 3	6.292	10.859	0.579	0.562	1.000		Stage 1-Stage 3	6.984	10.859	0.643	0.520	1.000
	Stage 1-Stage 4	53.806	26.626	2.021	0.043	0.433		Stage 1-Stage 4	52.444	26.626	1.970	0.049	0.489
	Stage 3-Stage 4	47.513	27.427	1.732	0.083	0.832		Stage 3-Stage 4	45.460	27.427	1.658	0.097	0.974

Table S6. The AUC values of ROC curve of metabolomic biomarkers in each tumor stage.

	Stage 1	Stage 2	Stage 3
L-Kynurenine	0.825	0.824	0.839
Phenylalanine	0.848	0.832	0.852
L-Valine	0.876	0.885	0.833
Proline	0.923	0.952	0.902
Spermidine	0.89	0.833	0.879
o-Tyr	0.822	0.883	0.852
Amino-hippuric acid	0.811	0.903	0.848
Palmitoyl-L-carnitine	0.906	0.935	0.931
Carnitine	0.848	0.888	0.861
Taurine	0.92	0.881	0.937

Abbreviations: ROC, receiver operating characteristic; AUC, area under the curve.

Table S7. Rank metabolomic biomarker features by FCBF.

	Score
Taurine	0.655
Palmitoyl-L-carnitine	0.482
proline	0.470
2-DG	0.446
PE(36:4)	0.353
Xanthine	0.303
PC(36:5)	0.290
Citric acid	0.280

Abbreviations: FCBF, fast correlation-based filter.

Table S8. AUC value of machine learning models changing with the number of variates.

	1	2	3	4	5	6	7	8
KNN	0.947	0.959	0.964	0.971	1.000	0.988	0.988	1.000
SVM	0.925	0.927	0.961	0.995	1.000	1.000	1.000	1.000
Random Forest	0.945	0.952	0.960	0.968	0.996	1.000	0.998	0.997
Neural Network	0.895	0.917	0.944	0.944	0.999	0.991	0.991	0.998
Naïve Bayes	0.869	0.926	0.964	0.990	0.998	1.000	1.000	1.000
AdaBoost	0.833	0.859	0.859	0.887	0.887	0.899	0.899	0.913

Abbreviations: AdaBoost, Adaptive Boosting; SVM, support vector machines; KNN, k-nearest neighbor; AUC, area under the curve.

Figure S1. ROC analysis of 10 metabolomic biomarkers in early lung tumor detection.

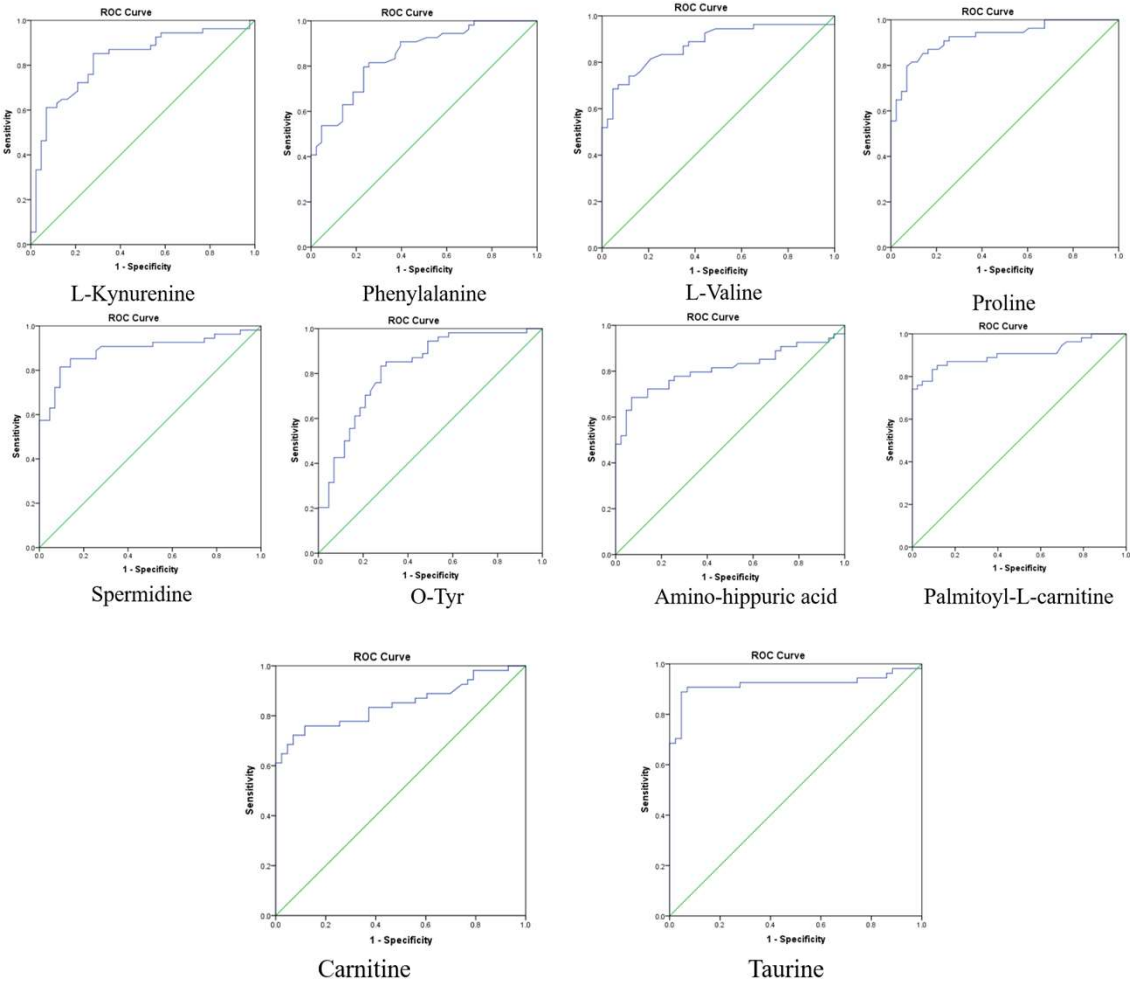


Figure S2. 50 influential metabolomic biomarkers with statistical significance of lung tumor patients (mean value with SD).

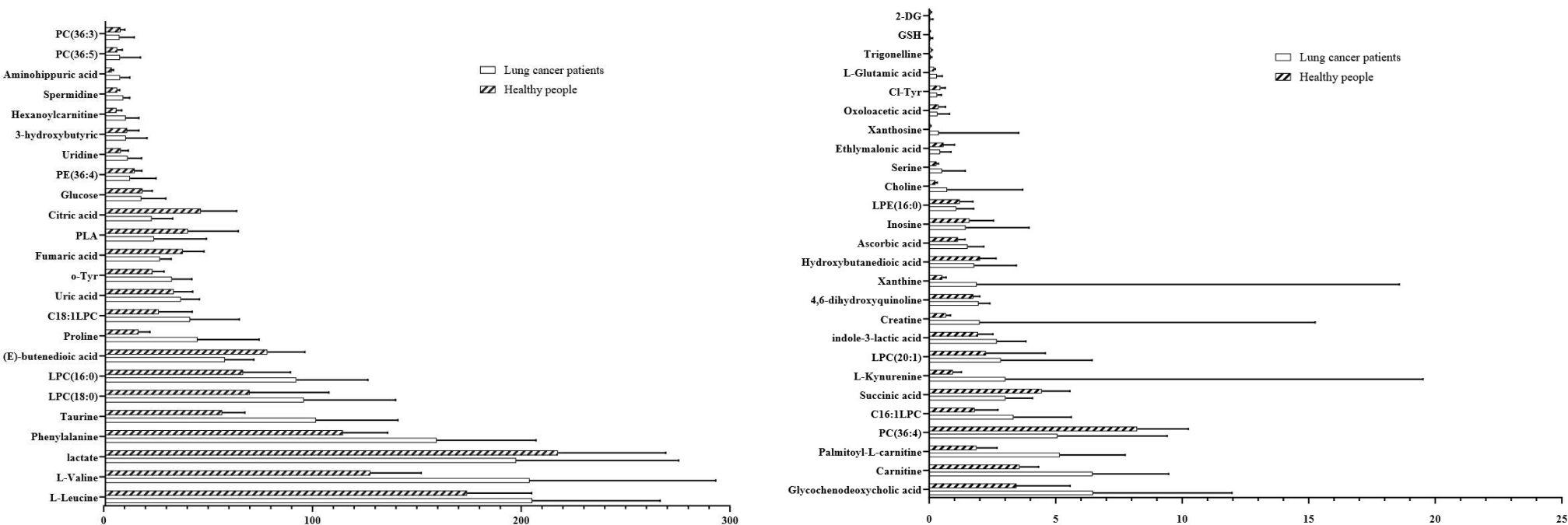
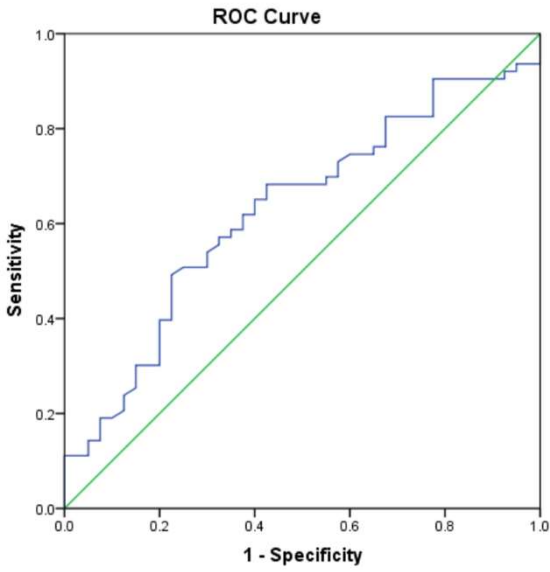


Figure S3. ROC analysis of metabolomic biomarkers with statistical significance of adenocarcinoma and squamous carcinoma patients.

Hippuric acid



Hypoxanthine

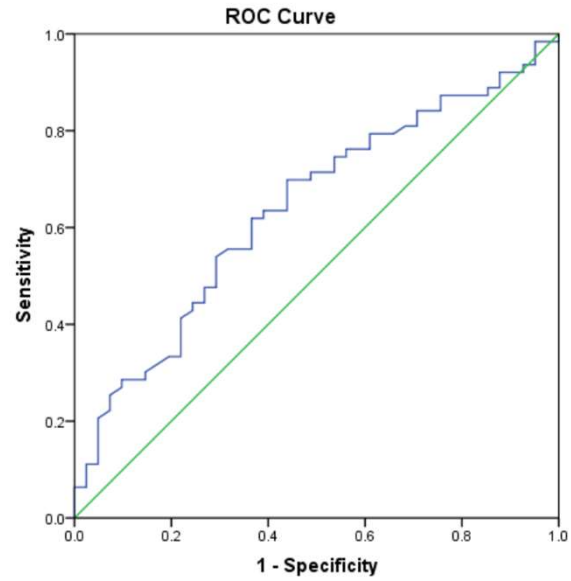


Figure S4. AUC value of machine learning models changing with the number of variates.

