Supplementary

Table S1 The diagnostic performance of this panel in four data sets

Data sets	AUC	95% CI	Sensitivity	Specificity
Training set	0.999	0.997–1.000	0.985	1.000
Validation set 1	0.965	0.920–1.000	0.917	1.000
Validation set 2	0.963	0.924–1.000	0.922	0.970
Integrated data set	0.967	0.945–0.989	0.934	0.964

The above results were obtained using the linear regression algorithm and the "pROC" R package. AUC, area under the ROC curve; ROC, receiver operating characteristic; CI, confidence interval.

Table S2 The NSCLC diagnostic ability of this panel under different algorithms

Algorithm	Training set	Validation set 1	Validation set 2	Integrated data set
Linear SVM				
AUC	0.999	0.925	0.944	0.969
Sensitivity	0.985	0.861	0.877	0.916
Specificity	1.000	1.000	0.970	0.976
Random forest				
AUC	0.999	0.957	0.958	0.981
Sensitivity	0.970	0.917	0.923	0.934
Specificity	1.000	1.000	0.909	0.952

Linear SVM algorithm and random forest algorithm were performed on MetaboAnalyst 5.0. NSCLC, non-small cell lung cancer; SVM, support vector machine; AUC, area under the ROC curve; ROC, receiver operating characteristic.

Table S3 A total of 893 DEGs in the turquoise module

PARPBP	SLC5A9	C7	IL6	AC144831.1
AC096921.2	HJURP	ITLN1	COL6A6	WISP2
CIP2A	CDCA3	PBK	CDC45	PTGDS
AC093278.2	HMMR	NPNT	PPP1R15A	GUCY1A2
HRCT1	FPR2	HSPC324	BMP2	TBX2
ADCY4	SPI1	ADAMTSL3	GINS4	EMCN
CPB2	CDK1	NDC80	WDHD1	MAOB
PPIAP39	NEK2	FGD5	GRASP	FIBIN
DENND2A	CCDC69	CLIC3	LYVE1	PCDH12
KANK2	AC011899.2	GATA6-AS1	NKAIN1	SCN4B
ODF3L1	WIF1	NPM3	BCL6B	CLSPN
AC013457.1	KLB	DCSTAMP	ZC3HAV1L	ECEL1P2
EFNA4	RBP4	MIR3677	GLIPR2	ADAMTS7P3
TDRD5	AC093787.1	SLC39A8	NDNF	SCN1A
SFTA1P	VWF	SGCA	SLC12A9-AS1	SPC24
FAM13C	KCNT2	AC009093.3	FLRT3	AC020907.1
PLPP2	HSPB6	EPAS1	NCAPH	DEPP1
IQCN	FHL1	HBA1	GAREM2	AC079630.1
SIGLEC11	ACP5	IGSF10	FBLN5	PLEK2
KIF2C	HMGA1	POLE2	PDE1B	AL445524.1
TGFBR3	AC012213.4	FFAR4	AGTR2	DUOX1
YBX2	ADAMTSL4	CDCA5	METTL7A	CCNB1
AURKB	GPD1	TESMIN	AC018755.4	CENPH
ABCA3	CYYR1	AOX1	PRKCE	BRIP1
SVEP1	SMIM25	VEPH1	CLDN18	GPRIN1
RTKN2	PLAC1	CTHRC1	TICRR	TNNC1
PCAT6	ZP3	KANK3	DIO2	KIF20A
SLC19A3	AQP1	MKI67	GNG4	GATA6
LRRC36	MAP1LC3C	PPP1R14A	AL035409.1	EFCC1
ARHGEF39	AL355388.1	TFAP2A-AS1	HMGB3	RASGRP4

Table S3 (*continued*)

Table S3 (*continued*)

CHEK1	ABI3BP	RRM2	SNHG4	DNA2
GPR146	CDO1	NMUR1	PEAR1	VSIG4
ASNS	NME1	CDCA8	COBL	RNU6-529P
AGTR1	STIL	TMEM150B	RASIP1	PCOLCE2
A2M	C3orf86	FGFR4	IL3RA	CA3
AP001972.5	LGI3	ORC1	ART4	AL133215.2
ADRB1	CD300LF	ETV4	CLEC1A	AQP9
LINC01290	EZH2	JPH2	OGDHL	ANOS1
MAGI2-AS3	CACNA2D2	PARAL1	CD36	SUSD2
NEIL3	FENDRR	PCLAF	ANGPT1	LAMP3
ST8SIA6	KIF14	CXCR1	OR7E47P	MMP9
AC027288.3	PTPRM	RAMP2	AP003469.2	KIF18B
AURKA	ZFP36	COL6A5	CCL2	APOBR
PPBP	LINC02555	BDNF	CLDN5	GPER1
AC016205.1	MYOC	S100A3	MIR3945HG	WDR62
SPOCK2	FRMD3	RECQL4	PLXNB3	DPEP2
CALCRL	FOXD3-AS1	GRK5	C5AR1	MYRF
CDCA2	MDK	ANKRD1	AC008268.1	FABP4
TEDC2	MCM10	RAI2	FEN1	RGCC
PLPP4	STEAP1	C1QTNF2	GPA33	SLC18A2
CCL23	KNL1	LINC02154	CLEC12A	CDKN2A
EDNRB	FANCI	GKN2	C1QA	TRIP13
GDF10	HELLS	ITGA8	GPBAR1	ABCA8
ADAMTS8	AC027288.2	AC025166.1	CXorf36	ADGRE5
SPTBN2	CFP	LINC01936	MYADM	FAM107A
AC112777.1	FAM110D	FCN3	IGF2BP3	COX7A1
IQSEC3	SGCG	COL10A1	TRAIP	CMTM2
CLEC3B	MSR1	GPR19	RASAL1	AC006329.1
BUB1	BARX1	LINC01996	ANKRD29	OIP5
SULF1	AC024560.2	ATOH8	AATK	DTL
FAM124B	AC099850.3	UBE2S	MIR27A	NECTIN4
AL136452.1	AC236972.3	PEBP4	FOXF1	MYCT1
COX4I2	ATP2A1-AS1	PHACTR1	CCNB2	PECAM1
TYMSOS	TPSAB1	PIF1	E2F2	STXBP6
SULT1C4	AP000769.1	LHFPL6	CENPA	PLAC9
ACADL	ZNF695	GATA2	COL11A1	MT1M
APOLD1	AC011511.5	ADM2	MMRN2	AC010976.2
SLC44A5	NCKAP5	AL136369.1	GAPDH	FAM83A
MT1A	RETN	CST1	FMO2	ANGPTL1
FGR	F12	MASP1	CNTN6	BUB1B
LINC00968	MEX3A	CHIAP2	DNASE2B	DLC1
MS4A2	GLDN	SGO1	AC131649.2	E2F8
CXCL2	SLC46A2	ZNF385B	ARHGEF15	LINC00511
AP000251.1	SLCO2A1	CD52	ADGRE1	EPN3
EFNA3	FXYD1	HSD17B6	VEGFD	PLOD2
CASS4	MMP11	ATP6V0D2	ESCO2	CDIPTOSP
KNTC1	FAM167A	AUNIP	ADGRE3	UHRF1
AC027277.2	SASH1	ASF1B	AC080037.1	FBP1
SFTPD	AC093110.1	TYMS	CDH5	AC084880.1
RAD51	CENPW	TMEM100	RAD54L	PCAT19
RAC3	FOSB	DES	LGI4	FBXO32

Table S3 (*continued*)

Table S3 (*continued*)

AL132712.2	LANCL1-AS1	LMOD1	SPARCL1	CAV1
RPL13AP17	ARHGAP6	CENPI	FAM72B	LDB2
SH3GL2	SIGLEC17P	AL606469.1	DPYSL2	KLF2
LRRN3	KIFC1	LRRK2	PYCR1	DUXAP8
PI16	LHFPL3	CCL14	CSF3	EMP2
CEP55	LILRA5	AOC3	GPM6A	ROBO4
ADRB2	XRCC2	CKAP2L	MGAT3	MCEMP1
TCF21	KL	SKA3	DEPDC1	PRC1
CDT1	CD300LG	SIX1	ROR1	HIST3H2A
C1QTNF6	ATAD2	NLRC4	DNASE1L3	CAVIN2
NUSAP1	TCEAL2	CD300C	ABCA9	SLC6A4
AP000866.2	JAM2	MRC1	C8orf34-AS1	FAM189A2
FSD1	LTBP4	OLFML1	RASL12	AC005856.1
MS4A15	NXPH3	CHRNA5	C1orf162	AP001453.2
MMP12	TFR2	ERCC6L	CCM2L	CASQ2
MYZAP	TK1	MYOZ1	PGM5P4	ITIH5
AC026369.3	TEK	EME1	FOXM1	SFTPC
ADAMTS1	HIGD1B	NCAPG	AQP4	GALNT14
DOK2	TRGJP2	CYP27A1	SFTPA1	CGNL1
CD93	DKK2	C5orf34	NECAB1	CASP17P
GIMAP1	PRSS35	EGR1	TIMP3	AC112722.1
CGREF1	COL1A1	AC104984.4	AC116407.1	FPR1
SMAD6	ADGRD1	GIMAP5	SFRP5	CARD14
GPIHBP1	MYO16-AS1	ACSS3	OLR1	CKS1B
CAV2	CAMP	KCNK3	CDCA4	MCM4
PROM2	ITM2A	CCDC85A	TROAP	GIMAP6
STX11	ADAM12	LEFTY2	USHBP1	PRR19
AP001528.3	TRPV2	GPR4	GIMAP8	JCAD
CRABP2	ECT2	PODXL2	RMI2	F11
ACOXL	AC091057.1	RBBP8NL	LPL	CCNE1
TTK	HSPA12B	ATP5MC1P4	GINS2	LHFPL3-AS2
GPX3	MARCO	HHIP	CRTAC1	CORO2B
ASPM	B3GNT4	SCN7A	GYPC	C14orf132
SEC14L6	AGMAT	BLM	CTGF	HYAL1
CPAMD8	CEACAM21	AL162511.1	AC011944.1	CENPM
SIRPB1	SERTM1	LRRC32	ITLN2	AL109741.1
MYBL2	TNXB	NUF2	ESAM	AP001189.3
ARHGEF26	CENPK	SEMA3G	FAM111B	SLIT3
ECSCR	RND1	FAM180A	B3GNT3	BOP1
PDK4	GADD45B	RPL39L	PFKP	HOXC9
HOXC-AS1	COL4A3	EXO1	SH3GL3	CA4
HBA2	TONSL	CAMK2N2	CCNF	SSTR1
ACVRL1	STRA6	IGF2BP1	PTCRA	STARD8
PLK4	ARC	TACC3	AP001189.1	TLR4
ABCB1	KIF23	NOVA2	IQANK1	SPAAR
CHAF1B	CHRDL1	GYPE	MELK	PPP1R14BP3
CCL24	STX1A	PRR11	C11orf96	TAL1
SLC25A10	CPED1	AC091133.4	SYNDIG1L	VIPR1
CDCA7	CHI3L2	HOXB7	MCM2	RSPO4
GPT2	GTSE1	SPP1	MIR30C2	AK4
AC007743.1	SHMT2	CENPU	C17orf53	LMNB1

Table S3 (*continued*)

Table S3 (*continued*)

|				
EGLN3	PRG4	TBX4	DLGAP5	NXF3
ARHGAP11A	ZNF366	LINC02016	LIMCH1	KIF15
GPRC5A	WWC2	TIE1	SPC25	AC004816.1
ALOX5AP	SKA1	TNNT1	UPK3B	CENPF
AL136162.1	NR4A1	HAS1	KIF11	BIRC5
RAD54B	HPDL	MAD2L1	SEMA3B	TBX5-AS1
FHL5	ZYG11A	PLK1	SFTPA2	SCARA5
EEF1E1P1	PPFIA4	AFF3	PVT1	ACKR4
RBP2	RDM1	HEG1	RACGAP1	INMT
STAC	AGER	CDC6	RCOR2	ST6GALNAC5
SPAG5	WFDC1	PKMYT1	NTM	SH2D3C
AC078778.1	PIMREG	POLQ	PTPN21	CDC25A
ESPL1	RFC4	TNS1	NR4A3	AL355338.1
SRSF12	NOTCH4	SLC1A1	DDX12P	LINC01836
KIAA1324L	HBB	AC008669.1	ESPN	CCNE2
SELP	TSACC	TPX2	EEF1A1P6	PALMD
LIMS2	SHC3	ERG	AC079467.1	LIN7A
BIK	CSRNP1	PRAM1	PGM5	NPR1
STYK1	ANKRD22	PDZD2	TENT5B	GPC2
RHOJ	PKNOX2	FAM83A-AS1	MAMDC2	LRP2BP
ADH1B	PF4	TMEM88	PREX2	DACH1
PLA2G4F	CLEC14A	OSCAR	LINC02321	LINC02471
TMEM139	AGRP	GRIA1	IGFBP3	GINS1
WNT3A	MS4A7	EIF4EBP1	HASPIN	FOLR3
VPS9D1-AS1	FCN1	PTH1R	RSPO1	ZFPM2-AS1
CYP4B1	AC093890.1	RASGRF1	CCNA2	ST6GALNAC3
AC141557.2	PRX	AC084864.1	NOSTRIN	TOP2A
ATP1A2	IL1RL1	CCBE1	CXCR2	SCUBE1
ZWINT	PTPRB	AC073585.1	RXFP1	MIR23A
GNG11	S1PR1	CLIC5	KCNAB1	CDC25C
AC004921.1	IQGAP3	LINC02185	IL23A	GAL
BTNL9	FAM162B	DEPDC1B	COLEC12	FANCB
C8B	SOX17	CX3CR1	LARGE2	CDC20
OR52K3P	MMRN1	STARD9	LEPR	DUSP5P1
DNMT3B	SPN	KPNA2	CFD	PSAT1
NMU	PTX3	RNU5B-4P	CXCL3	DENND3
PTTG1	UBE2C	PLA2G1B	ACE	AC083837.1
KIF4A	PCMTD1P3	AC147067.2	CDKN3	ANLN
TNFRSF18	CTSG	PIP5K1B	S1PR4	SELE
CENPE	ZNF423	CCL18	ZBTB16	AC007128.1
RAMP3	AL357054.4	VSIG2	SHCBP1	RSPO2
LINC01703	UBE2T	PID1	ORC6	CD101
CBLC	MIR4653	CCRL2	ACKR1	ALOX5
RXRG	RGS9	IHH	CST5	F8
SLIT2	MTFR2	GSTM5	OGN	DSCC1
PPARG	RAD51AP1	ALKAL2	ARHGAP31	
MFAP4	MND1	C1QTNF7	SAPCD2	

DEGs, differentially expressed genes.
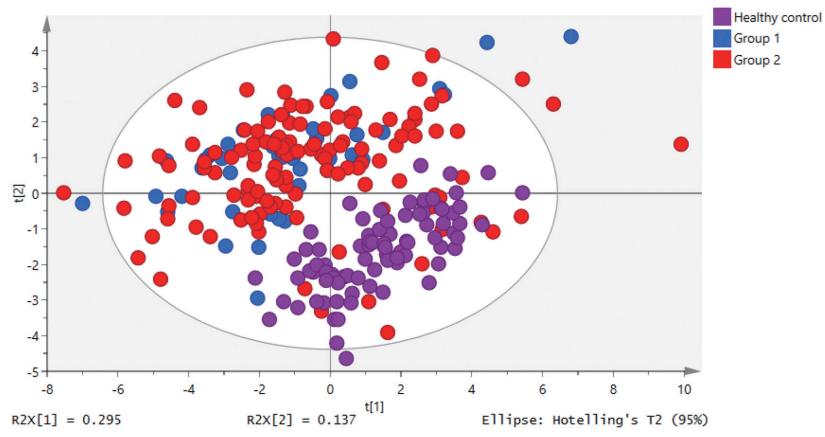
Figure S1 PCA of the grouped data. PCA, principal component analysis.
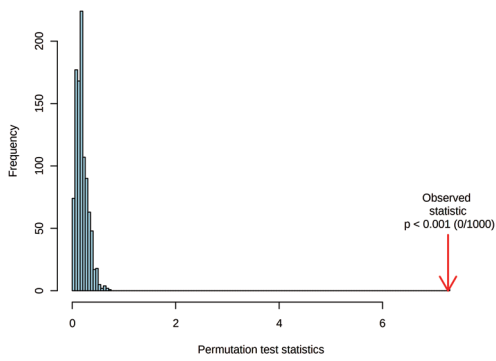

Figure S2 One thousand permutations test of the PLS-DA model. PLS-DA, partial least squares-discriminant analysis.
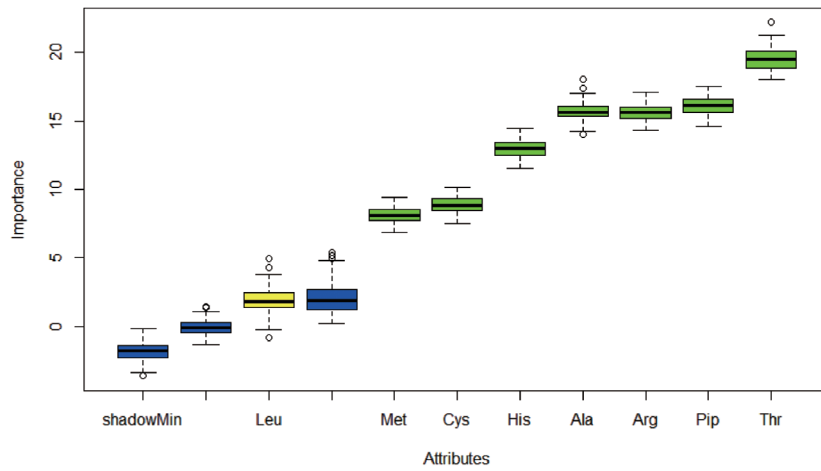

Figure S3 Screening of potential biomarkers in NSCLC. The differential metabolites corresponding to the 7 green box plots were selected as potential biomarkers for NSCLC. The yellow box plot corresponds to the excluded differential metabolite; the blue box plot represents the indicator. NSCLC, non-small cell lung cancer.
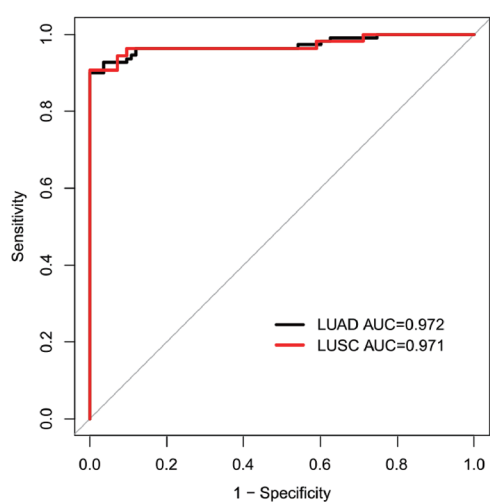
Figure S4 The diagnostic performance of this panel for NSCLC subtypes. This figure was made using the "pROC" R package. LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma; AUC, area under the ROC curve; ROC, receiver operating characteristic; NSCLC, non-small cell lung cancer.
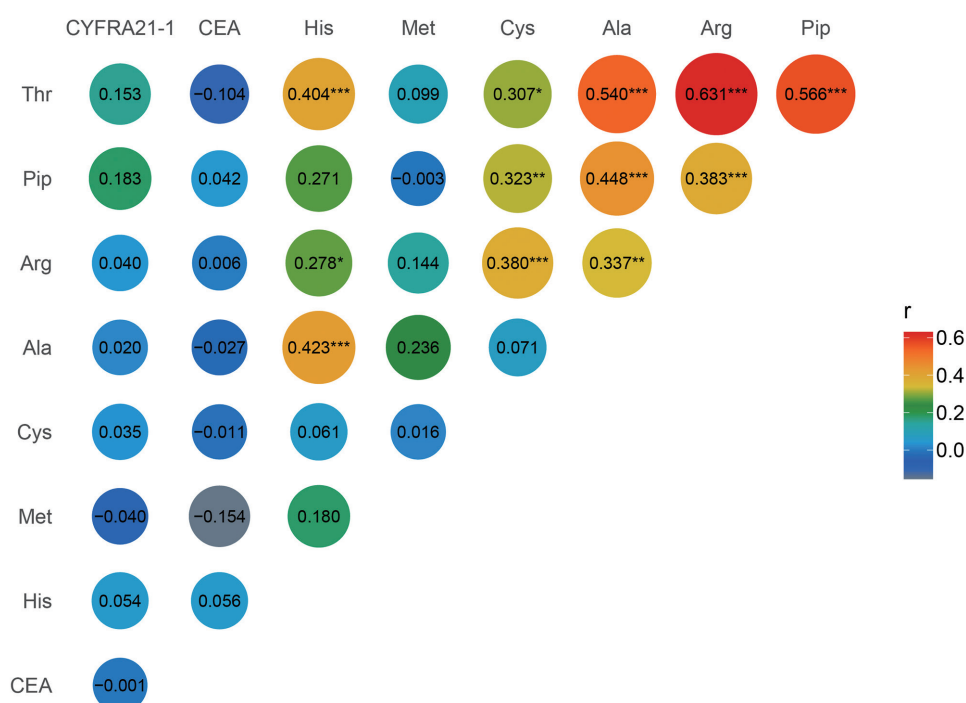

Figure S5 Correlation of 7 metabolic biomarkers with CEA and Cyfra21-1. P<0.05 was considered statistically significant. * represents P<0.05, ** represents P<0.01, and *** represents P<0.001. Cyfra21-1, cytokeratin 19 fragment; CEA, carcinoembryonic antigen.
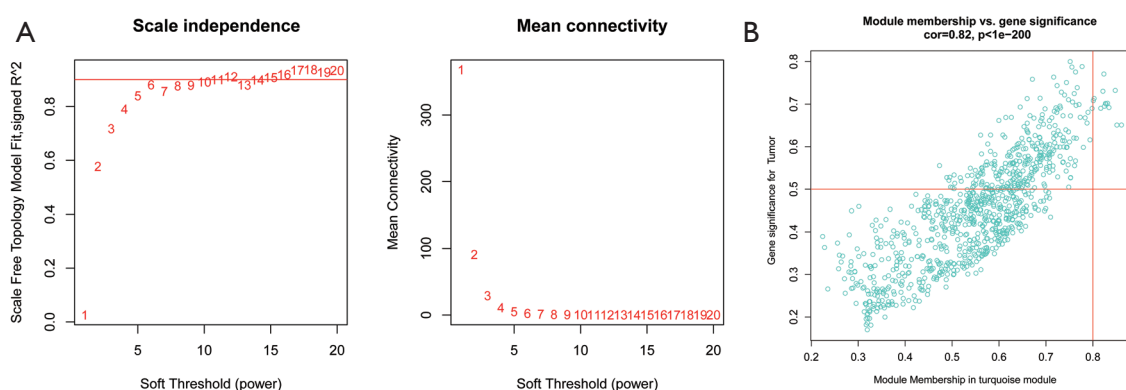

Figure S6 WGCNA of DEGs. (A) Determination of soft threshold in scale-free topology network. As shown in the left and right figures, when the soft threshold is 6, the network connectivity of the scale-free topology model is better. Thus, the soft threshold was set to 6; (B) scatter plot of significant genes related to NSCLC in the turquoise module. WGCNA, weighted gene co-expression network analysis; DEGs, differentially expressed genes; NSCLC, non-small cell lung cancer.
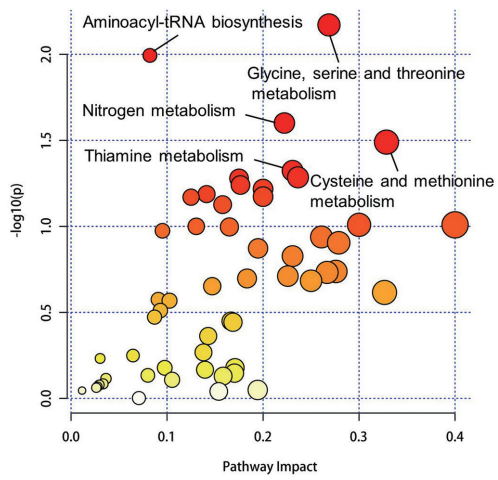
Figure S7 Enrichment analysis of integration pathways of 7 metabolic biomarkers and 893 DEGs. DEGs, differentially expressed genes.
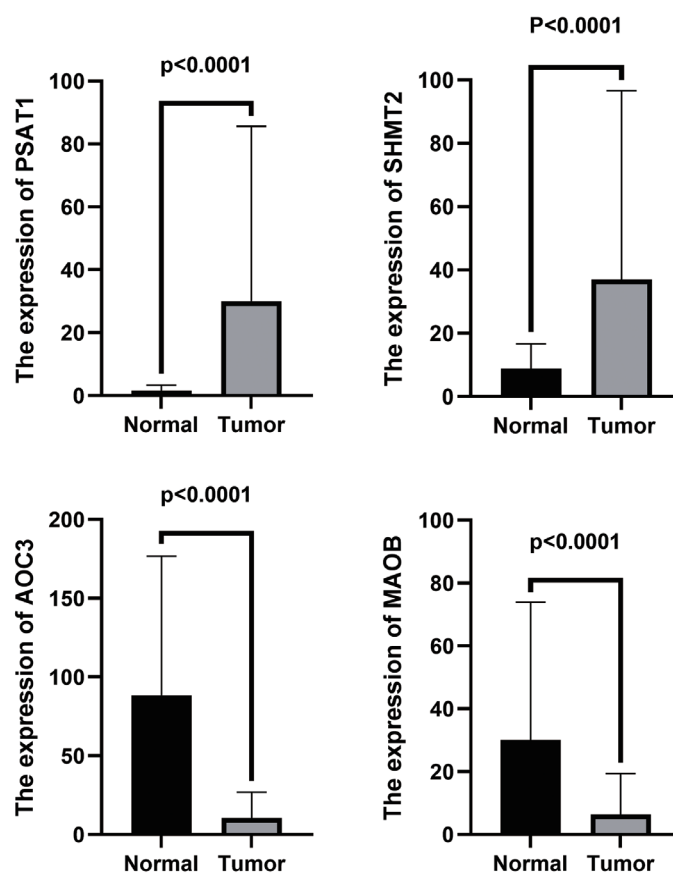

Figure S8 Four significantly changed metabolic genes in the extracted pathways.