

Figure S1 The workflow chart in this study.

Table S1 Analysis of differences in clinical features between mRNAsi related subtypes^a

Characteristic	C1 (n=476)	C2 (n=497)	χ^2	P
Age (years)			1.31	0.25
≤50	158 (46.2)	147 (41.5)		
>50	318 (53.8)	350 (58.5)		
T			36.57	<0.001
T1	160 (33.6)	106 (21.3)		
T2	239 (50.2)	316 (63.6)		
T3	71 (14.9)	48 (9.7)		
T4	6 (1.3)	24 (4.8)		
N			5.84	0.12
N0	227 (47.7)	232 (46.7)		
N1	162 (34.0)	166 (33.4)		
N2	41 (8.6)	61 (12.3)		
N3	40 (8.4)	28 (5.6)		
M			-	0.40
M0	379 (79.6)	436 (87.7)		
M1	4 (0.8)	9 (1.8)		
Stage			15.02	0.002
Stage I	105 (22.1)	69 (13.9)		
Stage II	247 (51.9)	304 (61.2)		
Stage III	110 (23.1)	104 (20.9)		
Stage IV	4 (0.8)	9 (1.8)		
PR			98.42	<0.001
Negative	77 (16.2)	223 (44.9)		
Positive	383 (80.5)	248 (49.9)		
ER			109.67	<0.001
Negative	38 (8.0)	176 (35.4)		
Positive	423 (88.9)	296 (59.6)		
HER2			10.13	<0.001
Negative	265 (55.7)	233 (46.9)		
Positive	43 (9.0)	100 (20.1)		
Subtype			94.45	<0.001
Luminal	292 (61.3)	210 (42.3)		
HER2+	3(0.6)	31 (6.2)		
TNBC	12 (2.5)	89 (17.9)		

Values are expressed as number (percentage). ^a, 3 TX, 16 NX, 145 MX, 21 stage X, 42 with indeterminate PR status, 40 with indeterminate ER status, 332 with indeterminate HER2 status and 336 with indeterminate molecular subtypes in the TCGA-BRCA cohort. T, tumor size; N, lymph node status; M, distant metastasis; ER, estrogen receptor; PR, progesterone receptor; HER2, human epidermal growth factor receptor 2.

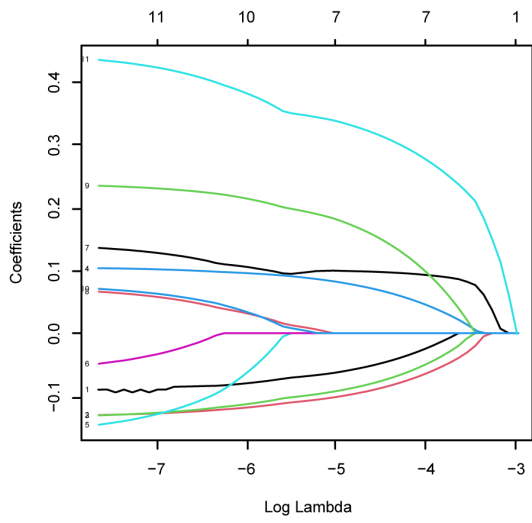
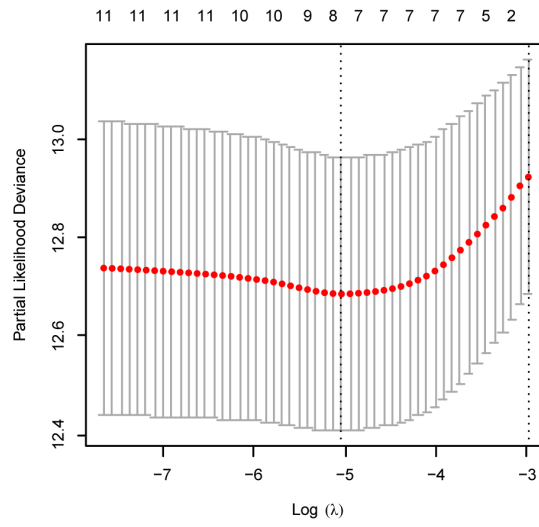
A**B**

Figure S2 LASSO algorithm process. (A) Regression coefficient path diagram. (B) Cross-curve graph. LASSO, least absolute shrinkage and selection operator.