

## Supplementary

**Table S1** Information for the patients enrolled in this study

Patient ID	Gender	Pathological diagnosis	Captured CTCs	Tumor size (cm)	EGFR mutation	TNM stage (AJCC 8th)
1	Female	Lung adenocarcinoma	13	1.0	19del	IA
2	Male	Lung squamous cell carcinoma	7	2.5	None	IA
3	Male	Lung adenocarcinoma	40	2.0	None	IA
4	Female	Lung adenocarcinoma	12	1.5	None	IA
5	Male	Lung squamous cell carcinoma	10	3.0	None	IA
6	Female	Lung adenocarcinoma	5	2.5	19del	IA

CTC, circulating tumor cell; EGFR, epidermal growth factor receptor; AJCC, American Joint Committee on Cancer; TNM, tumor node metastasis.

**Table S2** Confusion table of training cohort

Training	LUAD	LUSC	Normal lung	Total
LUAD	310	7	0	317
LUSC	5	225	0	230
Normal lung	3	3	60	66
Total	318	235	60	613
Correct	310	225	60	595
Correct (%)	97.48428	95.74468	100	97.06362

LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma.

**Table S3** Confusion table of validation cohort 1

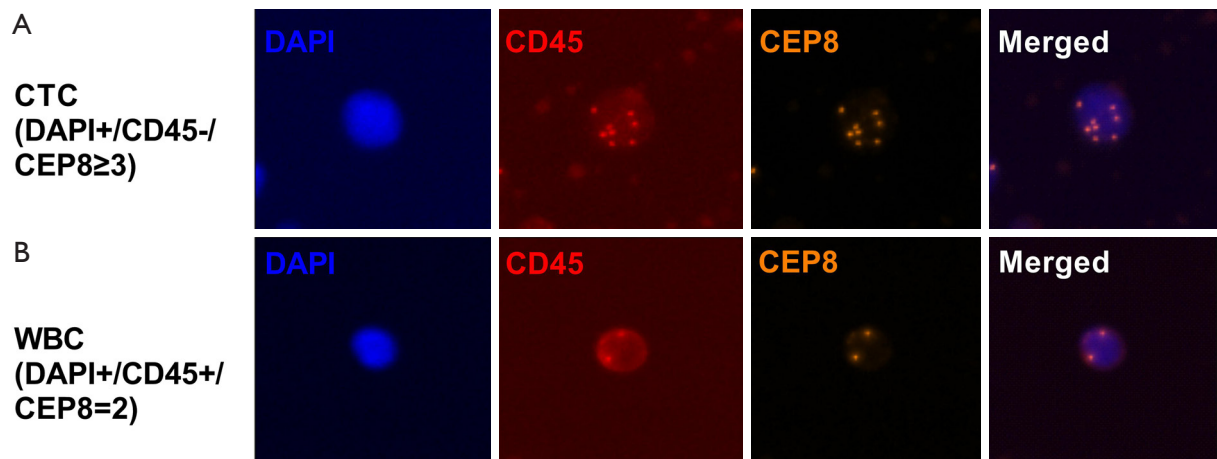
Validation 1	LUAD	LUSC	Normal lung	Total
LUAD	151	5	0	156
LUSC	2	127	0	129
Normal lung	1	2	14	17
Total	154	134	14	302
Correct	151	127	14	292
Correct (%)	98.05195	94.77612	100	96.68874

LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma.

**Table S4** Confusion table of validation cohort 2

Validation 2	LUAD	LUSC	Normal lung	Total
LUAD	25	1	1	27
LUSC	2	7	0	9
Normal lung	2	0	73	75
Total	29	8	74	111
Correct	25	7	73	105
Correct (%)	86.2069	87.5	98.64865	94.59459

LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma.



**Figure S1** *In situ* phenotypic and karyotypic characterization of aneuploid CTCs. CTCs are DAPI+ (blue)/CD45-/FISH+ (aneuploid chromosome 8, orange)  $\geq 3$ . WBCs are DAPI+ (blue)/CD45+/FISH- (aneuploid chromosome 8, orange) =2. The picture was magnified with 40x under the fluorescence microscope.

**Table S5** Summary of the basic sequencing parameters, including the sequencing depth, for all 6 patients

Case No.	Raw_bases	Conversion rate (%)	Map-ability (%)	Duplication rate (%)	Sequence depth	1xCpG coverage
01B	34023049106	99.51	88.62	16.50	7.001361	95.608
02B	31388659920	91.46	29.14	52.24	1.156164	44.667
03B	33872628040	99.77	87.78	18.54	6.717153	94.709
04B	31534218786	99.73	88.77	16.83	6.30206	93.076
05B	34806104906	99.75	87.05	16.96	6.863009	94.18
06B	33407950002	99.72	88.71	22.55	6.349538	93.927
01N	40379812376	99.68	87.23	21.45	7.645562	93.645
02N	1436336462	98.76	81.85	32.54	0.202227	14.748
03N	34455483812	99.69	85.31	17.22	6.717548	92.865
04N	37161722120	99.67	87.90	18.91	7.343653	93.019
05N	33460825068	99.68	87.93	21.77	6.38553	91.777
06N	36764933682	99.68	87.54	19.69	7.145472	93.673
01T	34563878558	99.65	87.21	19.57	6.687883	92.707
02T	31370116818	99.51	76.32	15.80	5.738217	92.61
03T	34458690146	99.72	80.61	18.77	6.189226	90.995
04T	36849359292	99.61	88.02	19.81	7.190256	92.878
05T	34271216700	99.71	86.99	22.66	6.35076	91.395
06T	38044759550	99.68	86.86	19.38	7.373278	93.425
01C	90723626720	98.99	62.06	66.06	3.856598	77.455
02C	91474230506	98.00	59.38	86.49	1.408053	45.289
03C	97534784928	98.98	71.55	88.42	1.777489	45.97
04C	1.00386E+11	98.88	71.49	91.78	1.254073	26.617
05C	1.0637E+11	98.94	71.95	94.55	0.863636	16.98
06C	93647437304	99.20	67.79	79.47	2.988654	71.377

**Table S6** Parameter conditions for the clustering of cancer tissues, normal tissues, WBCs and CTCs

Optimized Condition	CT-similar	CB-diff	CN-diff
Raw <sub>450K</sub>	–	–	–
CT <sub>450K</sub>	P>0.05; diff <10%	–	–
CBT <sub>450K</sub>	P>0.05; diff <10%	P<0.05; diff >0.15	–
CBNT <sub>450K</sub>	P>0.05; diff <10%	P<0.05; diff >0.15	P<0.05; diff >0.15

diff, difference.

**Table S7** Confusion table of our NSCLC tissue cohort

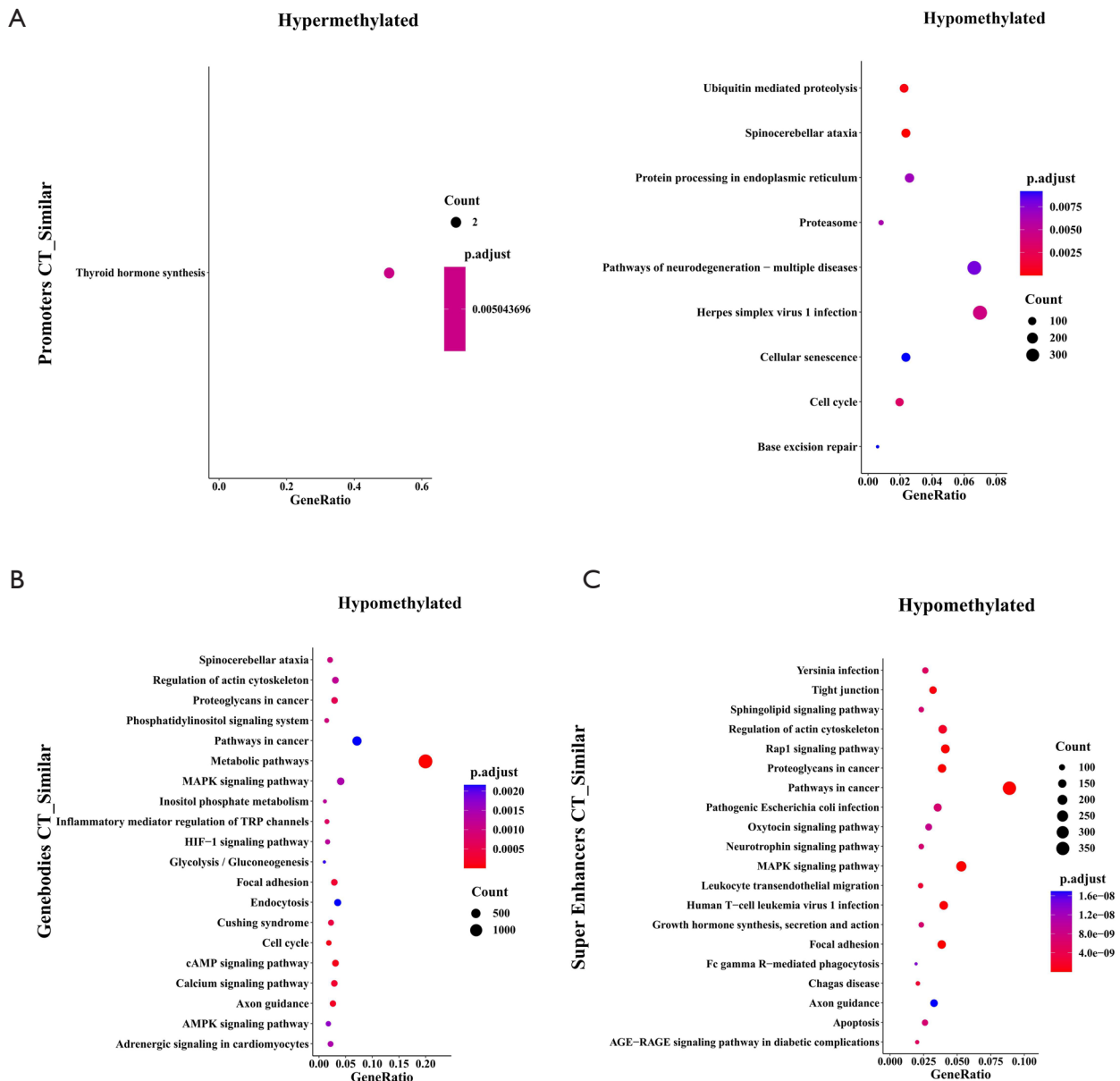
Validation	LUAD	LUSC	Normal Lung	Total
LUAD	4	0	0	
LUSC	0	2	0	
Normal lung	0	0	6	
Total	4	2	6	12
Correct	4	2	6	12
Correct (%)	100	100	100	100

NSCLC, non-small cell lung cancer; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma.

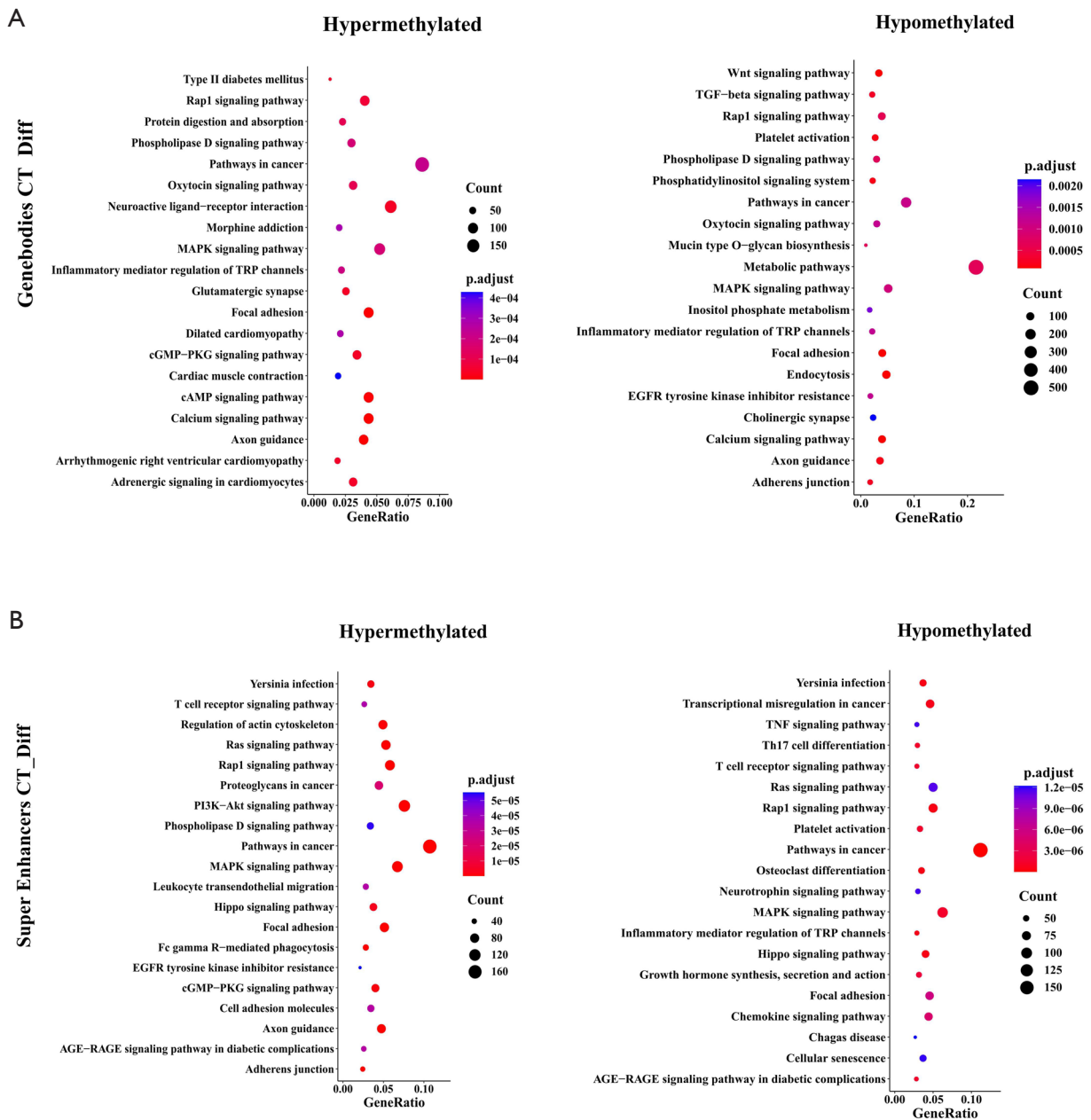
**Table S8** Confusion table of the CTC validation cohort, including both C&B diff (C&B diff >0.1, P<0.05) and 5,426 methylation markers

Validation	LUAD	LUSC	Total
LUAD	4	0	
LUSC	0	2	
Normal lung	0	0	
Total	4	2	12
Correct	4	2	12
Correct (%)	100	100	100

CTC, circulating tumor cell; diff, difference; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma.



**Figure S2** Pathway enrichment analysis of similarly regulated genes between CTCs and matched tumor tissues. The KEGG pathway enrichment of promoter (A) gene bodies (B) and superenhancers (C) displaying a <10% methylation difference (P value >0.05) between CTCs and matched tumor tissues. Gene sets with adjusted P value <0.01 are shown for promoters (A). For hypogene bodies (B) and hyposuperenhancers (C), the top 20 significant gene sets with an adjusted P value <0.01 are shown. For hypergene bodies (B) and hypersuperenhancers (C), no gene had an adjusted P value <0.01.



**Figure S3** Pathway enrichment analysis for differentially regulated genes from CTCs and matched tumor tissues. The KEGG pathway enrichment of gene bodies (A) and superenhancers (B) displaying  $>0.1$  absolute methylation difference ( $P$  value  $<0.05$ ) in CTCs compared to matched tumor tissues among 6 patients. For gene bodies (A) and superenhancers (B), the top 20 significant gene sets with an adjusted  $P$  value  $<0.01$  are shown. For promoters (A), no gene had an adjusted  $P$  value  $<0.01$ .