1. Supplementary Tables

Table S1. The performance evaluation results of 7 feature encoding using XGBoost.

XGBoost			
Feature	Accuracy	F1-score	AUC
PSTNPSSMC	0.645	0.735	0.895
RCKmer	0.468	0.303	0.627
CKSNAP	0.439	0.280	0.600
NAC	0.403	0.261	0.573
ENAC	0.422	0.254	0.577
NCP	0.417	0.235	0.586
ANF	0.318	0.236	0.524

Table S2. The performance evaluation results of 7 feature encoding using SVM.

SVM			
Feature	Accuracy	F1-score	AUC
PSTNPSSMC	0.602	0.700	0.882
RCKmer	0.472	0.232	0.647
CKSNAP	0.457	0.215	0.611
NAC	0.453	0.205	0.592
ENAC	0.432	0.205	0.622
NCP	0.453	0.211	0.614
ANF	0.352	0.151	0.466

Table S3. The performance evaluation results of 7 feature encoding using Random Forest.

Random Forest			
Feature	Accuracy	F1-score	AUC
PSTNPSSMC	0.632	0.752	0.890
RCKmer	0.479	0.280	0.634
CKSNAP	0.486	0.286	0.612
NAC	0.426	0.272	0.603
ENAC	0.454	0.245	0.591
NCP	0.442	0.239	0.580
ANF	0.400	0.220	0.521

Table S4. The performance evaluation results of 7 feature encoding using Logical regression.

Logical regression			
Feature	Accuracy	F1-score	AUC

PSTNPSSMC	0.395	0.242	0.839
RCKmer	0.478	0.215	0.645
CKSNAP	0.468	0.212	0.643
NAC	0.452	0.205	0.634
ENAC	0.376	0.261	0.575
NCP	0.376	0.247	0.547
ANF	0.353	0.169	0.487

Table S5. The performance evaluation results of 7 feature encoding using Multilayer perceptron.

Multilayer perceptron			
Feature	Accuracy	F1-score	AUC
PSTNPSSMC	0.557	0.664	0.854
RCKmer	0.481	0.219	0.655
CKSNAP	0.465	0.231	0.650
NAC	0.463	0.211	0.638
ENAC	0.388	0.266	0.579
NCP	0.393	0.244	0.558
ANF	0.350	0.177	0.493

Table S6. The performance evaluation results of 5 optimal candidate base classifiers with their best feature combination.

Methods	ACC	AUC	Macro F-measure
XGBoost(PSTNPSSMC+RCKmer)	0.687	0.910	0.773
SVM(PSTNPSSMC+RCKmer)	0.632	0.896	0.711
Random Forest(PSTNPSSMC+RCKmer)	0.655	0.901	0.752
Logical regression(RCKmer+CKSNAP)	0.477	0.646	0.215
Multilayer perceptron(PSTNPSSMC+RCKmer)	0.595	0.873	0.687

Table S7.Hyper-parameter values of the optimal candidate base classifiers.

XGBoost	Feature type	PSTNPSSMC+RCKmer
	learning_rate	0.11
	n_estimators	100
	max_depth	3
	subsample	1
	colsample_bytree	0.8
	gamma	0
Random Forest	Feature type	PSTNPSSMC+RCKmer

	bootstrap	False
	n_jobs	1
	n_estimators	400
SVM	Feature type	PSTNPSSMC+RCKmer
	C	15.0
	gamma	1.0
	probability	True
Logical Regression	Feature type	RCKmer+CKSNAP
	C	1e5
	penalty	L2
	tol	0.0001
Multilayer perceptron	Feature type	PSTNPSSMC+RCKmer
	n_epochs	200
	Learning_rate	0.016
	n_hidden	32
	Activation	relu
	Optimizer	adam

Table S8. The performance of 5 fold cross-validation under 10 different random seeds

KFold random_state	Accuracy	F1-score	AUC
1001	0.710	0.802	0.904
2001	0.706	0.794	0.905
2012	0.706	0.786	0.904
3001	0.689	0.782	0.895
4001	0.688	0.778	0.896
5001	0.687	0.775	0.900
6001	0.703	0.788	0.904
7001	0.690	0.785	0.897
8001	0.715	0.808	0.907
9001	0.687	0.778	0.897
Average	0.698	0.788	0.901

2.Supplementary figures

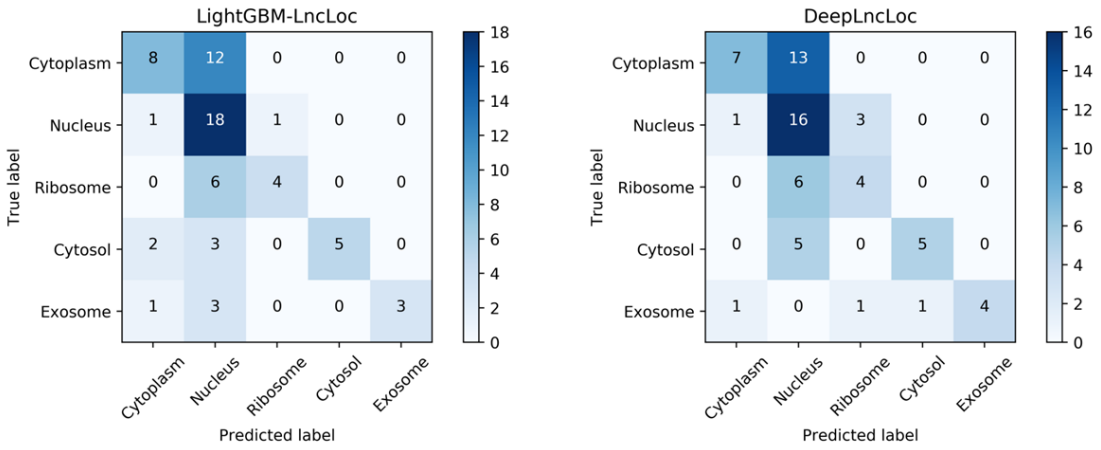

Figure S1. Confusion matrix of LightGBM-Lncloc with DeepLncLoc on the test set. (a) LightGBM-Lncloc, (b) DeepLncLoc. In the confusion matrix, the horizontal coordinates represent the predicted labels and the vertical coordinates represent the true labels. The diagonal line represents the number of correctly predicted samples.

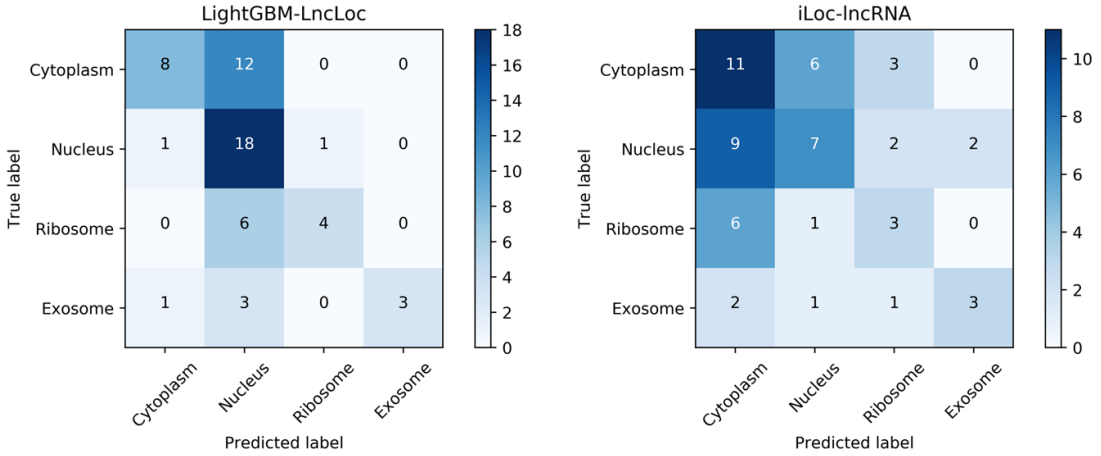

Figure S2. Confusion matrix of LightGBM-Lncloc with iLoc-lncRNA on the test set. (a) LightGBM-Lncloc, (b) iLoc-lncRNA. In the confusion matrix, the horizontal coordinates represent the predicted labels and the vertical coordinates represent the true labels. The diagonal line represents the number of correctly predicted samples.