

Supplementary materials:

CellCircLoc: Deep Neural Network for Predicting and Explaining Cell line-specific CircRNA Subcellular Localization

Min Zeng¹, Jingwei Lu¹, Yiming Li¹, Chengqian Lu², Shichao Kan¹, Fei Guo¹, Min Li^{1,*}

¹School of Computer Science and Engineering, Central South University, Changsha, Hunan, 410083, China.

²School of Computer Science, Key Laboratory of Intelligent Computing and Information Processing, Xiangtan University, Xiangtan, Hunan, 411105, China.

Contact: limin@mail.csu.edu.cn

This supplementary file includes:

1. Supplementary Figures

Fig. S1. The prediction performance of CellCircLoc and the baseline models. (A) HepG2 (B) Hela-S3 (C) HUVEC (D) Keratinocyte (E) GM12878 (F) H1-hESC

Fig. S2. Confusion matrices of CellCircLoc with RNALight on the test set. (A) K562 (B) HepG2 (C) Hela-S3 (D) HUVEC (E) Keratinocyte (F) GM12878 (G) H1-hESC. Each row represents the true class while each column represents the predicted class.

Fig. S3. ROC curve (receiver operating characteristic curve) of CellCircLoc with RNALight on the test set.

Fig. S4. The prediction performance of CellCircLoc and RNALight . (A)

HepG2 (B) Hela-S3 (C) HUVEC (D) Keratinocyte (E) GM12878 (F) H1-hESC

Fig. S5. Prediction of SRSF1 binding motifs of hsa_circBIRC6_008 using RBPmap. Position 62 and position 86 get the two highest Z-scores

Fig. S6. User interface of CellCircLoc web server for cell line-specific circRNA subcellular localization prediction.

2. Supplementary Tables

Table S1. Sequence length distribution of CircRNAs

1. Supplementary figures

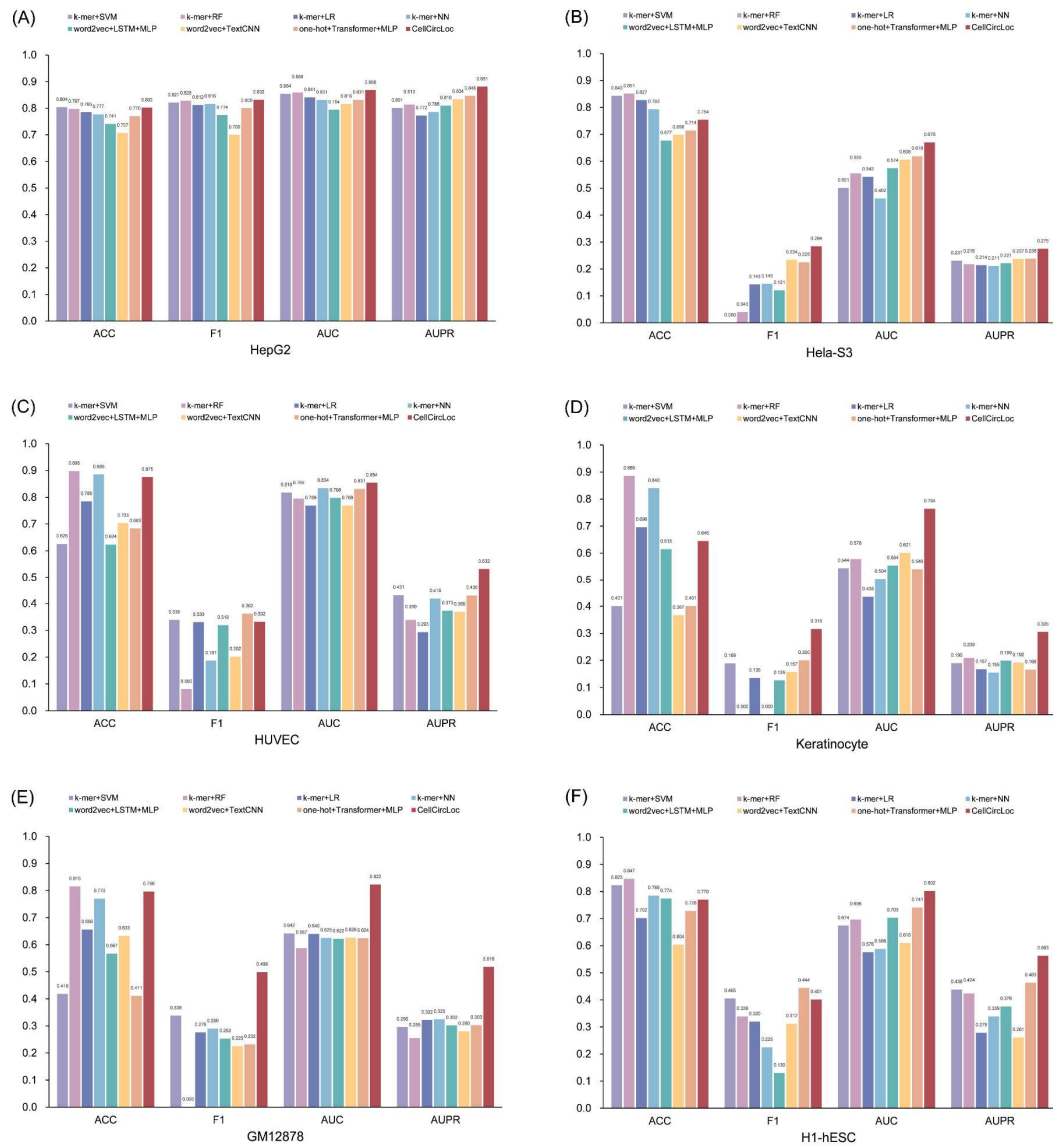


Fig. S1. The prediction performance of CellCircLoc and the baseline models. (A) HepG2 (B) HeLa-S3 (C) HUVEC (D) Keratinocyte (E) GM12878 (F) H1-hESC

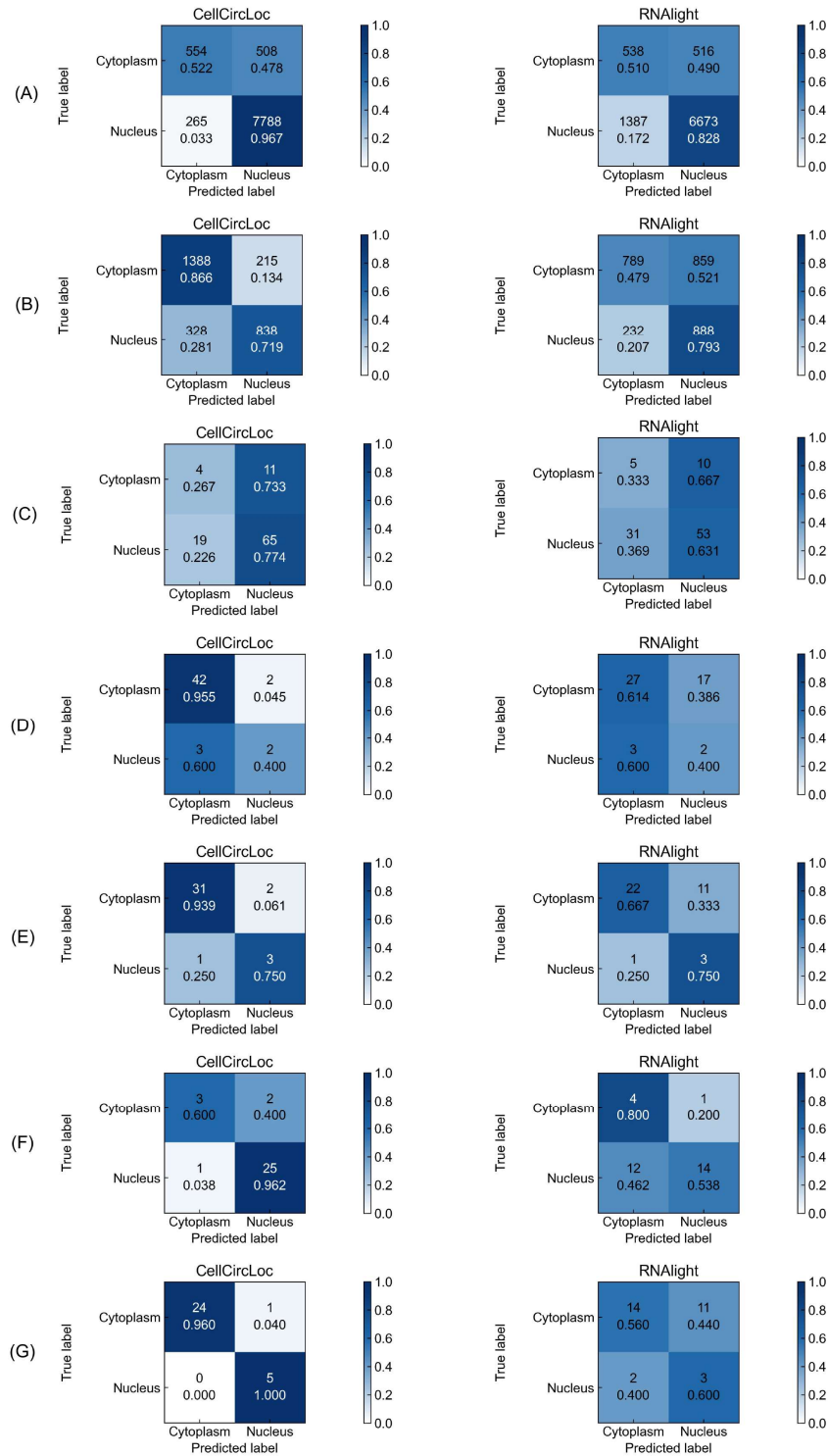


Fig. S2. Confusion matrices of CellCircLoc with RNAlight on the test set. (A) K562 (B) HepG2 (C) HeLa-S3 (D) HUVEC (E) Keratinocyte (F) GM12878 (G) H1-hESC. Each row represents the true class while each column represents the predicted class.

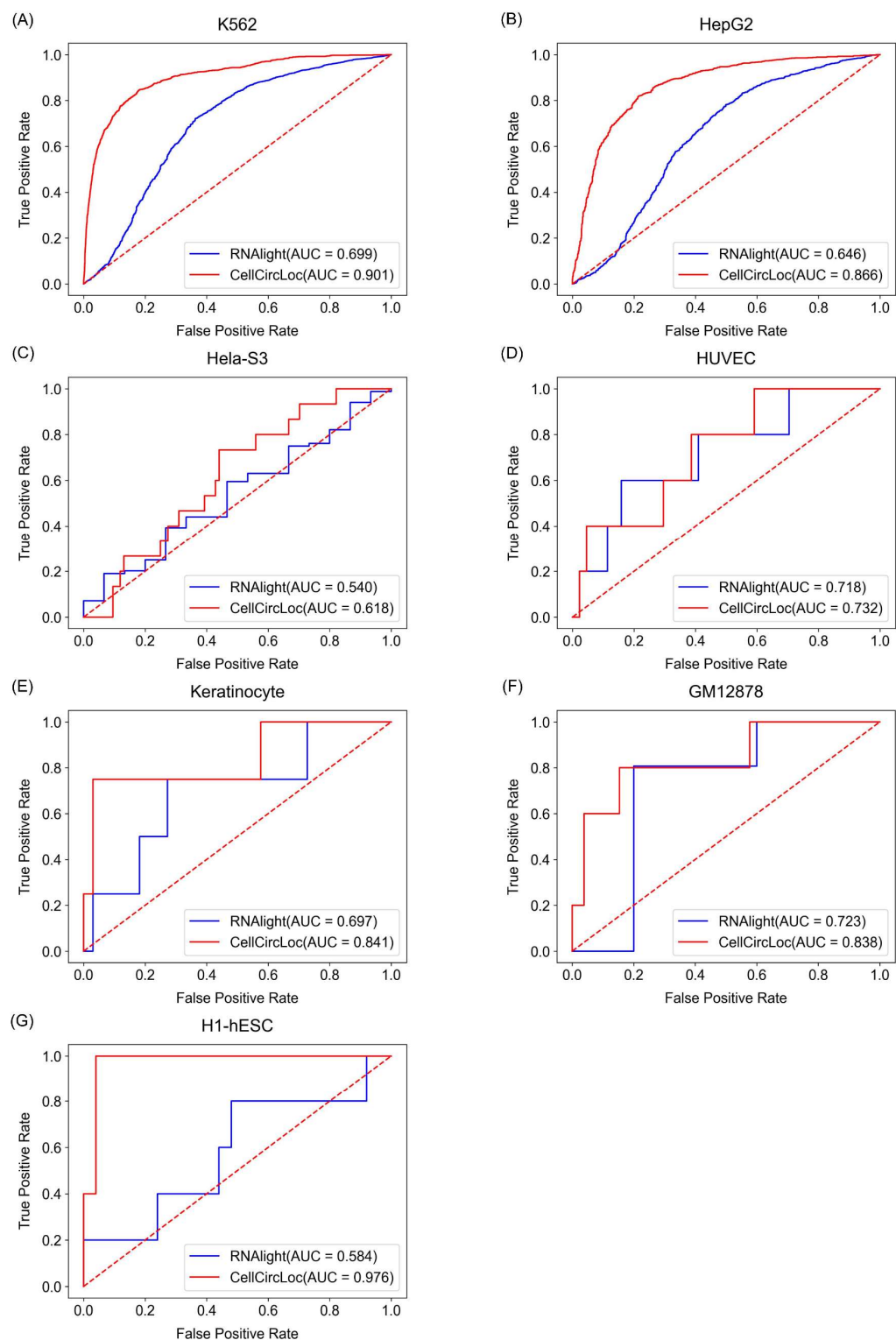


Fig. S3. ROC curves of CellCircLoc with RNAlight on the test set. (A) K562 (B) HepG2 (C) Hela-S3 (D) HUVEC (E) Keratinocyte (F) GM12878 (G) H1-hESC.

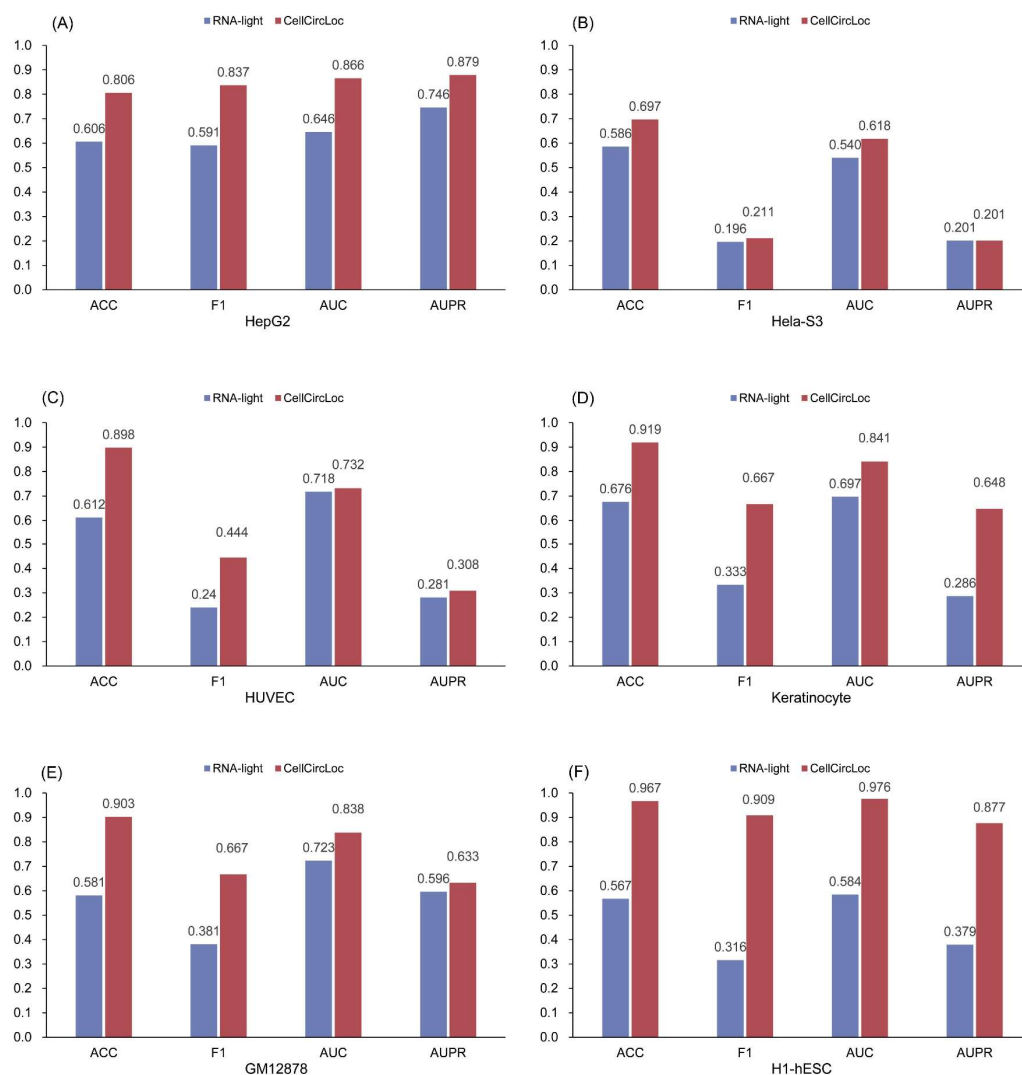

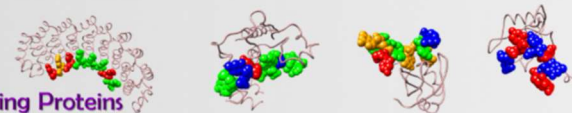


Fig. S4. The prediction performance of CellCircLoc and RNAlight. (A) HepG2 (B) HeLa-S3 (C) HUVEC (D) Keratinocyte (E) GM12878 (F) H1-hESC



version 1.2

Mapping Binding Sites of RNA Binding Proteins

[Home](#)
[Overview](#)
[Manual](#)
[Versions](#)
[Download](#)
[Contact](#)

RBPmap Results for job: 1691397463
 Job status: **Finished**
[General calculation parameters](#)
 Genome: Human (hg38)
 Selected motifs: SRSF1(Hs/Mm):gragga
 Stringency level: Default
 Conservation filter: Off

Results for sequence: sequence1
 Genomic position: N/A (BLAT could not find at least 95% identity match for the sequence).
☐ [View binding sites predictions summary](#)



Protein: SRSF1(Hs/Mm)

Position	Motif	Occurrence	Z-score	P-value
62	gragga	cugcuguggaauaaguuuuuuug <u>ggaug</u> auuugcaguaggguaggaaggac	3.290	5.01e-04
75	gragga	aguuaauuuuguggaugauuauugc <u>gaagg</u> uguaggaaggaccuuauuggaauuc	2.731	3.16e-03
82	gragga	uuuguggaugauuauugcaguagggu <u>gaagg</u> aaggaccuuauuggaauucuuuguu	3.151	8.14e-04
85	gragga	guggaugauuauugcaguaggguag <u>gaagg</u> accuuauuggaauucuuuguuuaga	2.688	3.59e-03
86	gragga	uggaugauuauugcaguaggguag <u>gaagg</u> accuuauuggaauucuuuguuuagac	3.280	5.19e-04

Predictions summary: [Download predictions summary as text file](#)

Note: the results will be kept on our server for 7 days.

If you use RBPmap please cite: Paz I, Kosti I, Ares M Jr, Cline M, Mandel-Gutfreund Y. (2014) RBPmap: a web server for mapping binding sites of RNA-binding proteins. Nucleic Acids Res., 2014. [\[PDF\]](#)


 Yael Mandel Gutfreund's Lab
 Computational Molecular Biology
 
 Technion – Israel Institute of Technology
 Faculty of Biology

This web site is supported by United States - Israel Binational Science Foundation

Fig. S5. Prediction of SRSF1 binding motifs of hsa_circBIRC6_008 using RBPmap. Position 62 and position 86 get the two highest Z-scores

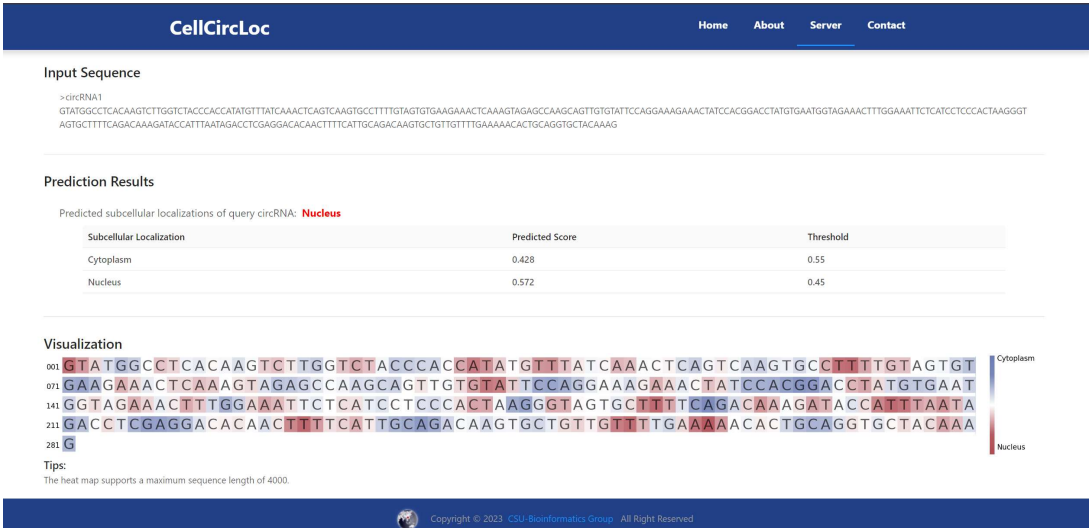


Fig. S6. User interface of CellCircLoc web server for cell line-specific circRNA subcellular localization prediction.

2. Supplementary Tables

Table S1. Sequence length distribution of CircRNAs

Sequence length	0-200	200-1000	1000-2000	2000-3000	3000-4000	4000-5000	5000+
Number	34613	62142	13435	4198	2065	1284	3529