

LETTER TO EDITOR

A long non-coding RNA signature for diagnostic prediction of sepsis upon ICU admission

Dear Editor,

Sepsis, the highest mortality disease in critically ill patients, is clinically diagnosed through the dysregulated systemic inflammatory response of patients to infection in the presence of organ dysfunction.^{1–3} No effective biomarkers and approved molecular therapies have been developed for sepsis to diagnose and treat the immune response state of the patients, leading to the management of these critically ill patients only relies on early recognition by experience and supportive care.^{4,5} Long noncoding RNAs (lncRNAs) are implicated in a wide variety of biological processes and accumulative studies have demonstrated that several dysregulated lncRNAs play important roles in tumorigenesis and tumor progression.^{6–8} However, the lncRNA signature has not been studied for the rapid diagnosis of sepsis, due to the limitation of data sources and lack of RNA-seq datasets.³ Hence, we analyzed three whole blood transcriptome cohorts of critically ill adult patients and identified a 28-lncRNA signature for sepsis diagnosis, which imputes a score to assess the risk of sepsis.

The expression profiling of 3745 lncRNAs in three cohorts, GSE95233, GSE28750, and GSE57065, were normalized and reannotated for the investigation^{6,9} (Table S1). The largest cohort GSE95233 was set as the discovery dataset, while the other two independent cohorts were set as the validation datasets. To select lncRNAs for the predictive signature, we first determined 84 differentially expressed (DE) lncRNAs between sepsis patients and healthy individuals based on the discovery dataset. Then we took advantage of a regression algorithm least absolute shrinkage and selection operator (LASSO) to further identify 28 predictive lncRNAs, named SepSig28, which serves as a molecular diagnostic signature to calculate the risk score to predict whether individuals were suffering from sepsis or not. After that, we validated the diagnostic signature in two independent datasets and demonstrated the

high performance of the 28 lncRNAs in the risk prediction of sepsis (Figure 1A).

Risk score = (BOLA3.AS1 × 0.254) + (LINC00354 × 0.1996) + (C5orf27 × 0.1537) + (RP1.187B23.1 × -0.1427) + (MBNL1.AS1 × -0.1419) + (LINC01420 × -0.1140) + (RP13.436F16.1 × 0.1060) + (CTB.31O20.2 × 0.1023) + (LINC01425 × 0.0949) + (C10orf25 × -0.0763) + (RP11.111M22.3 × 0.0743) + (LAMTOR5.AS1 × 0.0739) + (FLJ37453 × 0.0713) + (AX746755 × -0.0690) + (TTTY12 × 0.0678) + (ASMTL.AS1 × -0.0535) + (LOC101928491 × 0.0461) + (RBM26.AS1 × -0.0438) + (ANP32A.IT1 × 0.0437) + (LOC101060691 × 0.0319) + (MSH5 × -0.0311) + (LOC100507221 × 0.0289) + (RP11.1137G4.3 × -0.0245) + (LOC100506457 × 0.0237) + (MIR612 × -0.0189) + (AC114730.11 × 0.0079) + (LOC101927526 × 0.0026) + (LINC01019 × -0.0020). The values following the symbols are the importance weights of the expression abundance of each lncRNA. These lncRNAs are listed in order of decreasing importance.

When tuned in the discovery dataset using fivefold cross-validation, the SepSig28 can perfectly classify the sepsis patient samples and healthy control samples, with all the measures equal 1, including the area under curve (AUC), accuracy, sensitivity, and specificity (Figure 1B,C). To test the randomness of the model, we randomly picked up an equivalent number of lncRNAs 1000 times and evaluated their performance using the same procedure as SepSig28. Our result shows that no random combinations can achieve the score of AUC as high as 1 (Figure 1D). Besides, we constructed all possible 27-lncRNA signatures (28 minus 1) by excluding one lncRNA once a time to evaluate the predictive capability of each lncRNA in the SepSig28 model. For the discovery dataset, two lncRNA members are not necessary for the model, as the model can perform equally well without either of them (Figure 1C). We added these two as supplementary features to make the model more robust.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. *Clinical and Translational Medicine* published by John Wiley & Sons Australia, Ltd on behalf of Shanghai Institute of Clinical Bioinformatics

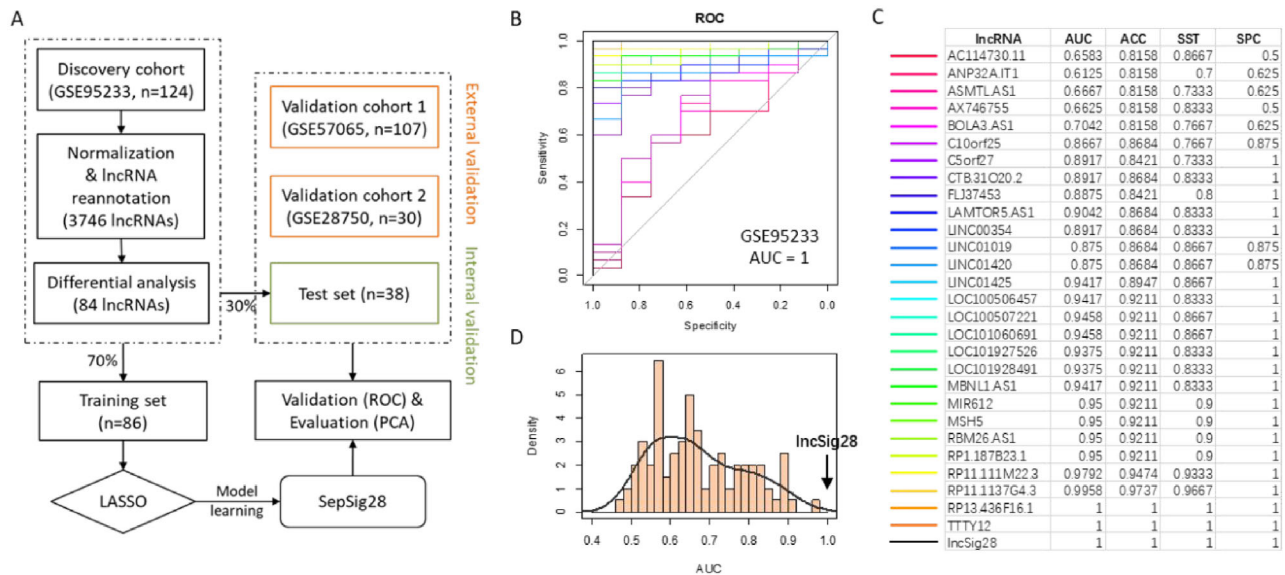


FIGURE 1 Model construction and internal validation. **A**, Workflow to identify the lncRNA signature of sepsis. **B**, ROC curves for the 28-lncRNA signature and other 28-minus-one lncRNA signatures. **C**, AUC, accuracy, sensitivity, and specificity for the 28-lncRNA signature and other 28-minus-one lncRNA signatures. **D**, Distribution of AUCs for the simulated models in which the lncRNAs were randomly picked up. ROC curve, receiver operating characteristic curve; AUC, area under curve

In the independent cohorts GSE28750 and GSE57065, the hierarchical clustering shows altered expression pattern of the SepSig28 lncRNAs cannot well distinguish sepsis patient samples from the normal ones (Figure 2A,B). Using the computed risk scores by weighted sum, however, SepSig28 can achieve the AUC scores as high as 0.9712 for GSE57065 and 0.95 for GSE28750, respectively (Figure 2C,D), which outperforms almost all the other combinations of 27 (28 minus 1) lncRNAs. Overall, SepSig28 has the best classification performance for all three cohorts according to the measures of AUC, accuracy, sensitivity, and specificity (Figure 2E).

To investigate the biological functions the SepSig28 involved, we associated them with their co-expressed genes across the sepsis samples of each cohort. Genes co-expressed with the lncRNAs in all the cohorts (Pearson correlation coefficient > 0.7) were considered to be co-expressed. Gene Ontology (GO) and KEGG pathway enrichment analysis were separately performed for the set of co-expressed genes.¹⁰ GO enrichment analysis showed that the lncRNAs of SepSig28 are mainly involved in three biological processes, including hormone mediated signaling pathway, RNA splicing, and histone modification (Figure S1A). KEGG analysis showed the SepSig28 associated genes are significantly implicated in pathways that are known to be related to sepsis pathogenesis, including Wnt signaling pathway, Th17 cell differentiation, Notch signaling pathway, etc. (Figure S1B). Interestingly, both GO and KEGG enrichment revealed that lncRNAs in SepSig28 tend to participate in hormone signaling related pathways, indi-

cating an underlying association between hormone signaling and sepsis.

In conclusion, we identified and validated the first non-coding signature consisting of 28 lncRNAs that can well distinguish sepsis patients from healthy controls for adults. Despite limitations such as the limited number of lncRNA features and the small sample size, we provided evidence that lncRNAs could be adopted as markers for the diagnosis of critical diseases. The proposed model could be used as an alternative or complementary diagnostic metric for sepsis.

AUTHOR CONTRIBUTIONS

LC conceived the idea and drafted the manuscript. LC performed data analysis. XL, XZ, JW, NZ, and RW performed data management and analysis. XL, KL, and XY helped interpret the results and give suggestions. All authors read and approved the final manuscript.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

Xueyan Liu^{1,*}
 Xubin Zheng^{2,3,*}
 Jun Wang^{1,*}
 Ning Zhang¹
 Kwong-Sak Leung³
 Xiufeng Ye²
 Lixin Cheng^{1,2} 

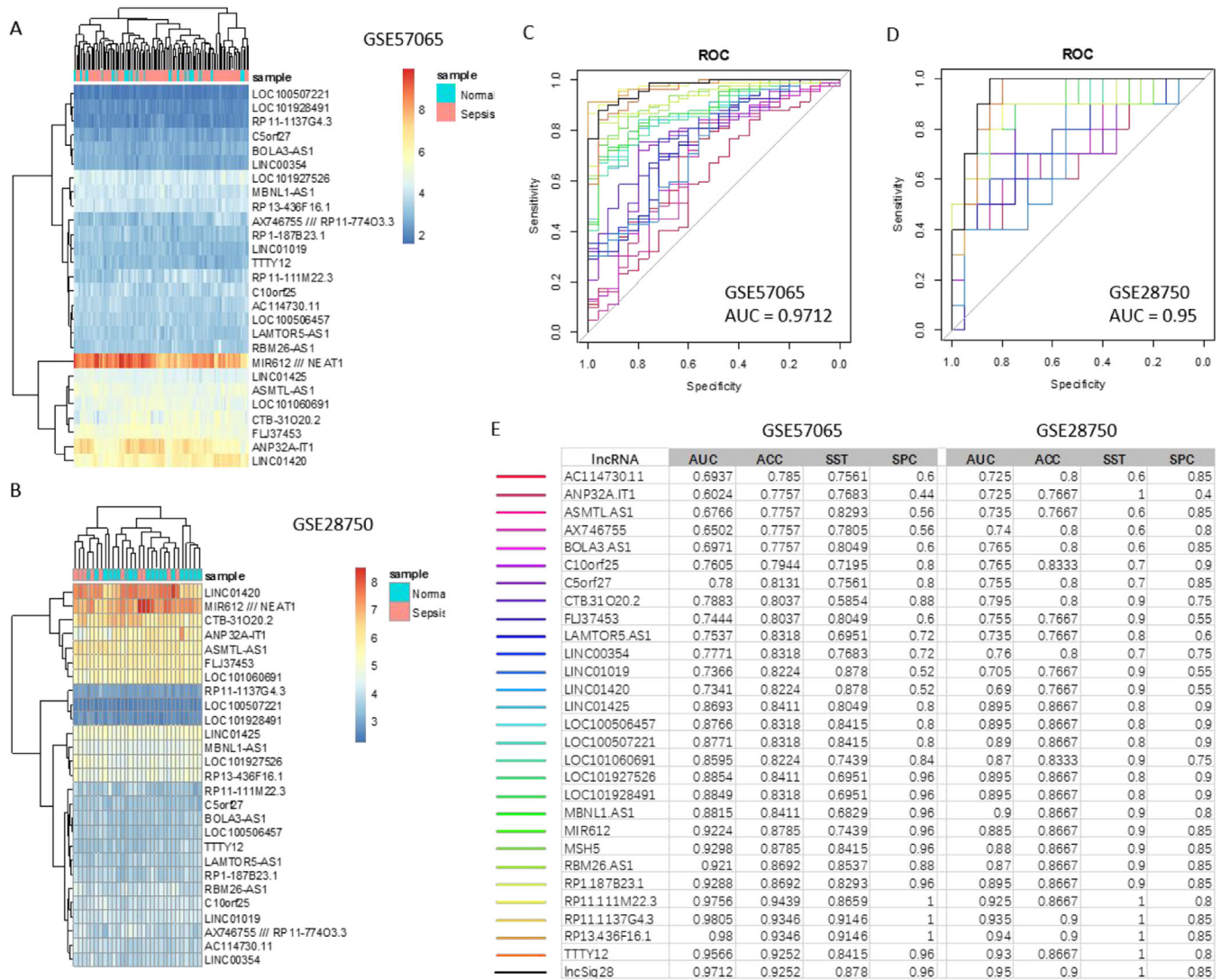


FIGURE 2 External validation of SepSig28. Hierarchical clustering of the expression samples based on the 28-lncRNA signature in dataset GSE57065 (A) and 28750 (B), respectively. ROC curves for the 28-lncRNA signature and other 28-minus-one lncRNA signatures in dataset GSE57065 (C) and 28750 (D). E, AUC, accuracy, sensitivity, and specificity for the 28-lncRNA signature and other 28-minus-one lncRNA signatures in the two validation cohorts

¹ Department of Critical Care Medicine, Shenzhen People's Hospital, First Affiliated Hospital of Southern University of Science and Technology, Shenzhen, China

² Shenzhen People's Hospital, First Affiliated Hospital of Southern University of Science and Technology, Shenzhen, China

³ Department of Computer Science and Engineering, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong

Xiufeng Ye, Shenzhen People's Hospital, First Affiliated Hospital of Southern University of Science and Technology, Shenzhen, China.

Email: easonlcheng@gmail.com; szrosexiu@126.com

*Both authors contributed equally to this work.

ORCID

Lixin Cheng <https://orcid.org/0000-0002-9427-383X>

Correspondence

Lixin Cheng, Department of Critical Care Medicine, Shenzhen People's Hospital, First Affiliated Hospital of Southern University of Science and Technology, Shenzhen, China.

REFERENCES

- Angus DC, van der Poll T. Severe sepsis and septic shock. *N Engl J Med.* 2013;369:840-851.

2. Fleischmann C, Scherag A, Adhikari NK, et al. International forum of acute care T: assessment of global incidence and mortality of hospital-treated sepsis. current estimates and limitations. *Am J Respir Crit Care Med.* 2016;193:259-272.
3. Cheng L, Nan C, Kang L, et al. Whole blood transcriptomic investigation identifies long non-coding RNAs as regulators in sepsis. *J Transl Med.* 2020;18:217.
4. Scicluna BP, van Vught LA, Zwinderman AH, et al. Classification of patients with sepsis according to blood genomic endotype: a prospective cohort study. *Lancet Respir Med.* 2017;5: 816-826.
5. Sutherland A, Thomas M, Brandon RA, et al. Development and validation of a novel molecular biomarker diagnostic test for the early detection of sepsis. *Crit Care.* 2011;15:R149.
6. Zhou M, Zhao H, Wang X, Sun J, Su J. Analysis of long non-coding RNAs highlights region-specific altered expression patterns and diagnostic roles in Alzheimer's disease. *Brief Bioinform.* 2019;20:598-608.
7. Liu X, Xu Y, Wang R, et al. A network-based algorithm for the identification of moonlighting noncoding RNAs and its application in sepsis. *Brief Bioinform.* 2020.
8. Cheng L, Leung KS. Identification and characterization of moonlighting long non-coding RNAs based on RNA and protein interactome. *Bioinformatics.* 2018;34:3519-3528.
9. Cheng L, Lo LY, Tang NL, Wang D, Leung KS. CrossNorm: a novel normalization strategy for microarray data in cancers. *Sci Rep.* 2016;6:18898.
10. Cheng L, Leung K-S. Quantification of non-coding RNA target localization diversity and its application in cancers. *J Mol Cell Biol.* 2018;10:130-138.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.