

# TransLnc: a comprehensive resource for translatable lncRNAs extends immunopeptidome

Dezhong Lv<sup>1,†</sup>, Zhenghong Chang<sup>1,†</sup>, Yangyang Cai<sup>1,†</sup>, Junyi Li<sup>1</sup>, Liping Wang<sup>1</sup>,  
Qiushuang Jiang<sup>1</sup>, Kang Xu<sup>1</sup>, Na Ding<sup>1</sup>, Xia Li<sup>1,2,\*</sup>, Juan Xu<sup>1,2,\*</sup> and Yongsheng Li<sup>1,2,\*</sup>

<sup>1</sup>College of Bioinformatics Science and Technology, Harbin Medical University, Harbin, Heilongjiang, 150081, China and <sup>2</sup>Key Laboratory of Tropical Translational Medicine of Ministry of Education, College of Biomedical Information and Engineering, Hainan Women and Children's Medical Center, Hainan Medical University, Haikou, 571199, China

Received August 08, 2021; Revised September 05, 2021; Editorial Decision September 10, 2021; Accepted September 10, 2021

## ABSTRACT

**lncRNAs are not only well-known as non-coding elements, but also serve as templates for peptide translation, playing important roles in fundamental cellular processes and diseases. Here, we describe a database, TransLnc (<http://bio-bigdata.hrbmu.edu.cn/TransLnc/>), which aims to provide comprehensive experimentally supported and predicted lncRNA peptides in multiple species. TransLnc currently documents approximate 583 840 peptides encoded by 33 094 lncRNAs. Six types of direct and indirect evidences supporting the coding potential of lncRNAs were integrated, and 65.28% peptides entries were with at least one type of evidence. Considering the strong tissue-specific expression of lncRNAs, TransLnc allows users to access lncRNA peptides in any of the 34 tissues involved in. In addition, both the unique characteristic and homology relationship were also predicted and provided. Importantly, TransLnc provides computationally predicted tumour neoantigens from peptides encoded by lncRNAs, which would provide novel insights into cancer immunotherapy. There were 220 791 and 237 915 candidate neoantigens binding by major histocompatibility complex (MHC) class I or II molecules, respectively. Several flexible tools were developed to aid retrieve and analyse, particularly lncRNAs tissue expression patterns, clinical relevance across cancer types. TransLnc will serve as a valuable resource for investigating the translation capacity of lncRNAs and greatly extends the cancer immunopeptidome.**

## INTRODUCTION

Long non-coding RNAs (lncRNAs) have recently been demonstrated as an abundant class of RNA molecules that play important roles in various types of cellular processes (1,2). Advances in transcriptome have identified a large number of lncRNAs, and mis-regulation of lncRNA expression programs has been found to cause a broad range of human diseases (3–5). Increasing lncRNAs have been identified as oncogenic drivers or tumour suppressors in a variety of cancers (6–8).

lncRNAs are not only well-known as ‘non-coding’ elements, but also serve as templates for protein translation. Functional peptides encoded by lncRNAs have drawn increasing attention from researchers due to their emerging important roles in fundamental cellular processes and diseases (9,10). The existence of functional peptide products of lncRNAs emphasizes the need to thoroughly separate the RNA functions and peptide functions of lncRNAs. Several mechanisms and evidences for lncRNA translation have been developed. For example, many short sequences, such as *N*-6-methyladenosine (m6A) sites, have been reported as internal ribosome entry sites (IRESs)-like elements to drive ncRNAs translation (11,12). Rapid developments of genome-wide translation profiling, ribosome profiling and mass spectrometry further provide a chance that has never existed before for genome-wide identification of potential lncRNA peptides (13). Advances in ribosome profiling aid to the systematically identification of small open reading frames (sORFs) from lncRNAs (14). In addition, it is possible to identify translatable lncRNAs based on proteomic data (15). Currently, large number of lncRNAs have been identified based on transcriptome sequencing, and several databases have been developed for describing their genomic annotation (16), tissue-specific expression (17), regulation (5) or diseases (18). However, a specialized and comprehensive resource for lncRNA peptides is still lacking.

\*To whom correspondence should be addressed. Tel: +86 451 86615922; Fax: +86 451 86615922; Email: liyongsheng@hainmc.edu.cn  
Correspondence may also be addressed to Juan Xu. Email: xujuanbiocc@ems.hrbmu.edu.cn  
Correspondence may also be addressed to Xia Li. Email: lixia@hrbmu.edu.cn

†The authors wish it to be known that, in their opinion, the first three authors should be regarded as joint First Authors.

Moreover, peptides that are presented on surface-bound MHC proteins provide valuable targets for cancer immunotherapy (19). Neoantigens have been identified extensively based on the detection of cancer-specific mutations (20,21). Besides genetic mutations, several lines of evidences have suggested other potential sources of neoantigens, such as alternative splicing, RNA editing and peptides generated from noncoding regions (22–24). All these peptides thus expand the pool of MHC immunopeptidomes, targetable by immunotherapies. In particular, lncRNAs have been demonstrated to play an important role in immune regulation (25,26), and their encoded peptides provided a rich resource for tumour neoantigen. However, the potential neoantigen from lncRNAs have yet to be systematically predicted.

To meet this need, we proposed TransLnc for providing or visualizing both computationally predicted and experimentally supported lncRNA peptides in multiple species (human, mouse and rat). By integrating several direct and indirect evidences, we performed an integrative analysis to identify functional peptides encoded by lncRNAs. Importantly, we have computationally predicted the neoantigens from the peptides encoded by lncRNAs, which would provide novel insights into the immunotherapy of cancer. Several flexible tools have been developed to aid retrieval and analysis of the data, particularly the tissue expression patterns of lncRNAs across tissues, the clinical relevance of lncRNAs across cancer types, and the BLAST alignment for finding the best matched lncRNAs of the user queried sequences. Collectively, we expect the TransLnc database will be an important resource for providing novel insights into cancer immunotherapy. All the data and tools are freely available at <http://bio-bigdata.hrbmu.edu.cn/TransLnc/>.

## MATERIALS AND METHODS

### Prediction of lncRNAs translation

Both the genomic coordinates and transcript sequences of human and mouse lncRNAs were downloaded from GENCODE (Human V32 and Mouse VM23). The corresponding lncRNA information in rat was downloaded from Ensemble database.

To identify all putative peptides encoded by lncRNAs, we used three-frame translation similar as previous studies (11,15,27). All sequences of lncRNA transcripts were first translated into amino acids by the function ‘translate’ in seqinr package in R program (Figure 1A). ORFs are the nucleic acids sequences between start codons (AUG) and stop codons (UAA, UAG and UGA). Only ORFs that produced peptides longer than 10 amino acids were retained. The selected minimum threshold for ORF lengths was based on the lengths of lncRNAs known to encode peptides in previous studies (27,28). The coding ability of an lncRNAs transcript is calculated as the proportion of RNA sequences that can encode peptides.

### Multi-omics evidence for lncRNAs translation

With the development of high-throughput experimental technologies, multi-omics evidence for lncRNAs translation have been revealed. For each computationally predicted

lncRNA peptide, TransLnc integrates six types of direct and indirect evidences to support the coding potential of lncRNAs (Figure 1A).

*Collection of experimentally verified translated lncRNAs.* We collected the lncRNAs encoded peptides from four manually curated databases, cncRNAdb (29), EVlncRNAs (30), ncEP (31) and FuncPEP (32). In total, 1185, 18, 13 and 41 lncRNAs that can encode peptides were obtained from four databases. The experimentally supported peptides were aligned with predicted peptides. We identified all lncRNAs and their encoded peptides that covered the experimentally supported ones.

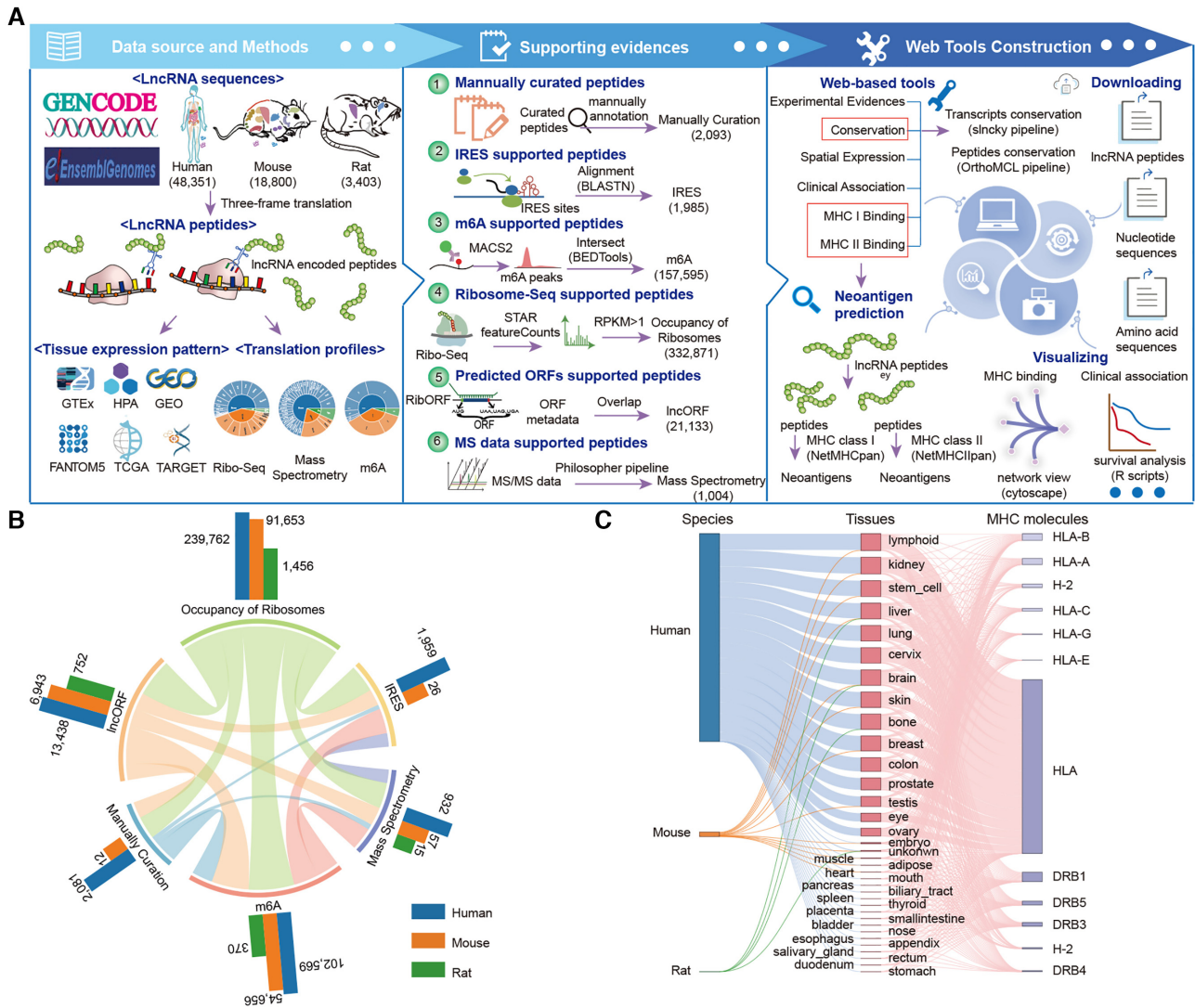
*Ribosome-associated lncRNAs.* We collected publicly available 2115 Ribo-seq samples from human, mouse and rat. To identify ribosome-associated lncRNAs, the Ribo-seq datasets were processed and reanalysed by a pipeline similar as one of previous studies (33). Briefly, SRA-Toolkit was used to convert sra files to fastaq format. Quality control was performed based on FastQC. We next removed adapter and low quality sequences using Cutadapt (34) and Sickle. The sequences from tRNAs and rRNAs were removed and then aligned to the reference genome by STAR (35). The read count and reads per kilobase per million mapped reads (RPKM) were estimated by featureCounts (36). lncRNAs with RPKM >1 in at least one sample were identified.

*Prediction of lncORFs.* RibORF was used to systematically detect the actively translated regions in the Ribo-seq datasets (37). Moreover, the predicted ORFs from Meta-mORF (38) and OpenProt (39) were obtained. The ORFs predicted by at least two methods were remained for further analyses. The predicted ORFs were compared with lncRNAs encoded peptides. lncRNA encoded peptides overlapped with the predicted ORFs were defined as supported by lncORFs.

*IRES-supported lncRNAs translation.* To identify IRES elements in lncRNAs, we first extracted experimentally validated IRES sequences from IRESbase (40). BLASTN was used to align the IRES sequences to lncRNAs peptides (41). The parameters ‘-evalue 1e-5 –outfmt 6 –perc\_identity 80 –num\_alignments 30’ were used in our analyses. All lncRNAs encoded peptides with IRES elements were identified.

*m6A sites in lncRNAs.* The m6A modification peaks across cell lines or tissues were extracted from REPIC (42), which is a database for exploring the m6A methylome. Three computational tools, including exomePeak (43), MeTPeak (44) and MACS2 v2.1.1 (45) were used to detect peaks. The m6A peaks were aligned to the genomic regions of lncRNAs encoded peptides by BEDTools (46).

*Detection of lncRNAs-encoded peptides by mass spectrometry.* Mass spectrometry-based proteomic data from were used for detection peptides derived from lncRNAs. We first extracted the peptides from PepTransDB resource across nine cancer types in human (15). In addition, we detected the peptides from collected additional MS data of 31 human



**Figure 1.** Schematic of overall design of TransLnc. (A) Data content, supporting evidences and web tools construction of TransLnc. (B) Numbers of lncRNAs-encoded peptides with diverse types of supporting evidences. (C) Numbers of potential neoantigens encoded by lncRNAs across different tissues in three species.

normal tissues, as well as mouse and rat tissues. All the MS datasets were obtained from the PRIDE proteomics identifications database (47). There were 41 samples of 31 tissues for human, 9 samples of 3 tissues for mouse and 3 samples of 1 tissue for rat. The similar pipeline in PepTransDB was used to search lncRNA peptides, where MS/MS spectra search was performed using the MSFragger search engine (48) and PSMs were processed using PeptideProphet as implemented in the Philosopher pipeline (49).

### Cross-species conservation of translatable lncRNAs

The conservation of both lncRNA transcripts and peptides were calculated. For lncRNA transcripts, we used *slinky* pipeline for identifying orthologous lncRNAs between two species (50), which is a commonly used sensitive method for orthologous lncRNAs alignment. The parameters were the same as one previous study (51). On the other hand, we used reciprocal best hits between pairs of species to cluster lncR-

NAs peptides into homologous families with OrthoMCL (52). The parameters were used similar as the study of Sarropoulos *et al.* (2).

### Detection of potential neoantigens encoded by lncRNAs

The binding of antigenic peptides to the major histocompatibility complex (MHC) molecules represents a critical step for cellular immunity. MHC molecules come in two main variants: MHC Class I (MHC-I) and MHC Class II (MHC-II). MHC-I binds peptides from intracellular proteins after these undergo proteasomal degradation while MHC-II binds peptides generated by protease-digestion of extracellular proteins. We thus used NetMHCpan-4.1 and NetMHCIIpan-4.0, two commonly used web servers to predict binding between lncRNA peptides and MHC-I or MHC-II, respectively (53). We used the %Rank values for detecting strong MHC binder (SB) or a weak MHC binder (WB). The %Rank < 0.5% and %Rank < 2% thresholds are



considered for SBs and WBs for class I, and %Rank < 2% and %Rank < 10% for SBs and WBs for class II.

### Tissue-specificity analysis

Besides Ribo-Seq and MS data with known tissue information, the other data were first mapped to corresponding tissues according to the annotations in Cancer Cell Line Encyclopedia (CCLE) (54). Then the remaining cell types were manually mapped to corresponding tissues. The expressions of lncRNAs across various types of human tissues were collected from LncSpA (17), which is a comprehensive resource that provides the spatial expression patterns of lncRNAs. In addition, expression levels of mouse and rat lncRNAs across multiple-tissues were also integrated into TransLnc (55,56).

### Clinical relevance

Differentially expressed lncRNAs in cancer with more than five adjacent tissues were identified. We compared the expression of lncRNAs between cancer and normal samples by *t*-test and Fisher's exact test. lncRNAs that were with *P*-value < 0.05 were defined as differentially expressed ones. In addition, we used univariate and multivariate cox regression to evaluate the association between survival and expression level of each lncRNA. Patients were ranked based on expression of lncRNA and the difference in survival between low- and high-expression groups was evaluated by the Kaplan-Meier method.

### Database implementation

The backend of TransLnc is powered by MySQL (v5.5.48) and accessed via the Java Server Pages with Tomcat container (v 6.0) as the middleware. The front-end of TransLnc is a multi-page web application built based on HTML, JavaScript, and CSS code that consists of jQuery (v3.3.1), Datatable (1.10.25) and ECharts (V5.5.1) plugin. All statistical analyses were performed using the R framework (V3.6.3). TransLnc has been tested on several popular web browsers, including Google Chrome (preferred), Firefox or Apple Safari browsers.

### DATABASE CONTENT

Based on data collection and integrative analysis, there were 380,630 (33.58%), 153,926 (13.58%) and 15,160 (1.34%) peptides under the 10 aa length in Human, Mouse and Rat, respectively. As number of studies focused on longer peptides (15,37), we thus excluded the peptides < 10 aa in our analysis. In total, the current version of TransLnc contains a total of 583 840 computationally predicted peptides for 33 094 translatable lncRNAs across three different species (Human, Mouse, Rat). In addition, 383 105 lncRNA peptides were supported by at least one of six types of direct or indirect evidences across 34 tissues (Figure 1B). Next, we compared the general characterizations of the predicted peptides versus experimentally confirmed. We found that the predicted peptides were with similar length and chromosome localization (Supplementary Figure S1), but different in amino acid compositions (Supplementary Figure

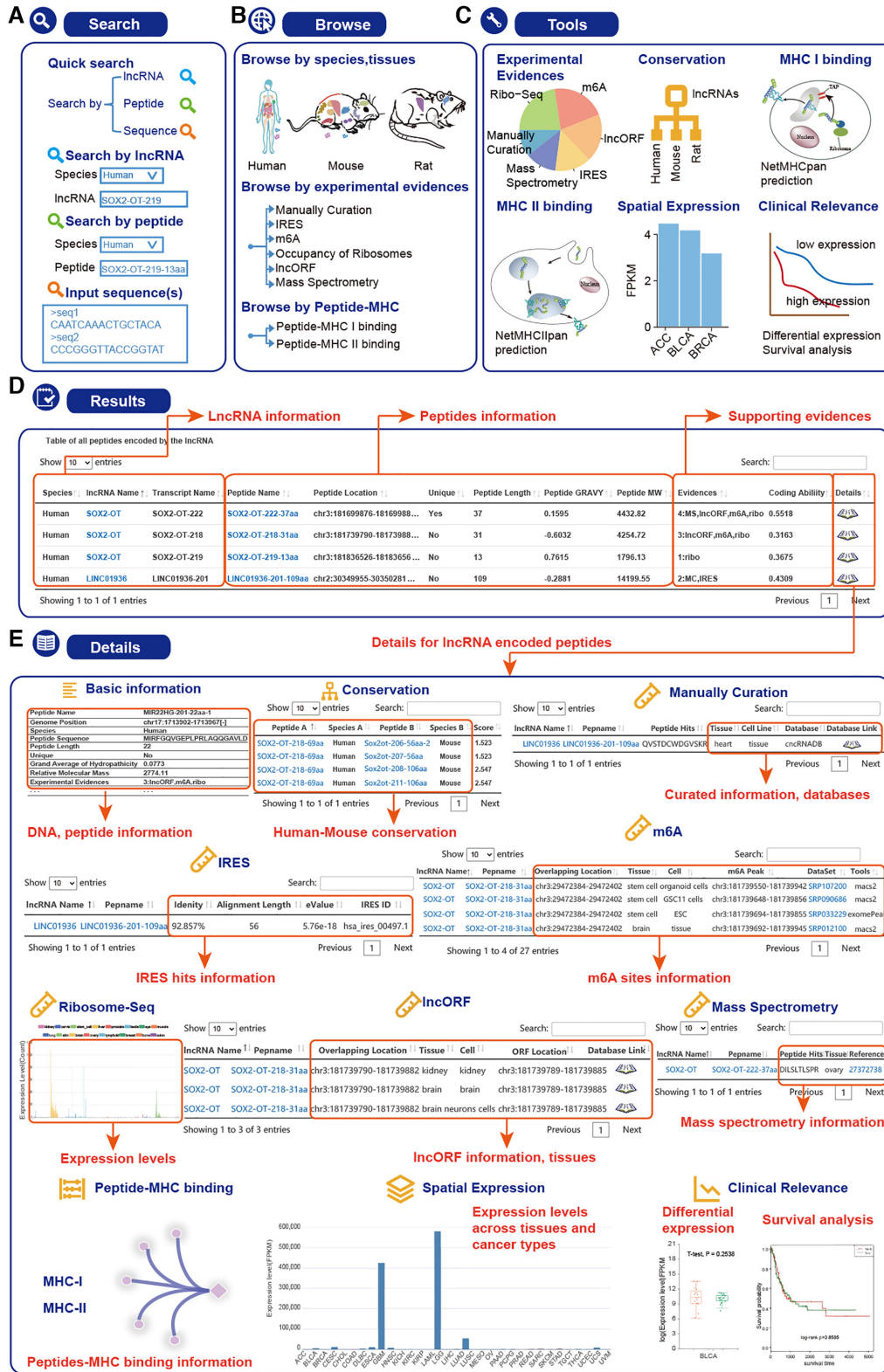
S2), with experimentally confirmed ones. Importantly, there are 220 791 candidate neoantigens binding by MHC class I proteins, and 237 915 candidates binding by MHC II proteins (Figure 1C), extending the cancer immunopeptidome. We found that there were 67,288 peptides can be bound by both molecules, which account for 30.48% and 28.28% of MHC-I and MHC-II binding peptides. Moreover, we compared the characterizations of lncRNAs that encoded peptides binding to MHCs versus no binding. We found that the lncRNAs that encode peptides binding to MHCs were more likely to be implicated in cancer (Supplementary Figure S3). In addition, we obtained the immune-related lncRNAs from ImmLnc (6), and we found that lncRNAs that encode peptides binding to MHCs were more likely to be involved in immune regulation (Supplementary Figure S4).

To provide an efficient way for users to investigate the function of lncRNAs of interest, TransLnc also provided the useful tools for investigating the function of lncRNAs that encoded peptides. Considering the strong tissue-specificity of lncRNAs, TransLnc allows users to access lncRNA peptides in tissues of interest, multiple visualization results were provided for facilitate understanding their roles in cancer immunity. Several flexible tools have been developed to aid retrieval and analysis of the data, particularly the tissue expression patterns of lncRNAs across tissues, the clinical relevance of lncRNAs across cancer types, and the BLAST alignment for finding the best matched lncRNAs of the user queried sequences.

### WEB INTERFACE

TransLnc allows users to explore the peptides encoded by lncRNAs and identify the potential peptides binding by MHC molecules. The easy-to-use interface allows for querying, browsing, visualising and downloading data (Figure 2). In the 'Search' page, TransLnc allows users to query lncRNAs or peptides of interest. TransLnc provides a direct BLAST tool for customized sequence query that allows identifying relevant lncRNAs with associated peptides (Figure 2A). In the 'Browse' page, the users can browse lncRNA encoded peptides by species, tissues, or by experimental evidences, or by peptides bound by MHC I or II (Figure 2B). In addition, six types of interactive tools (experimental evidences, conservation, MHC I or II binding, spatial expression and clinical relevance) were provided in TransLnc to retrieve detail information of lncRNA encoded peptides (Figure 2C). In the 'Result' page, lncRNA and corresponding encoded peptide information, and supporting evidences were shown (Figure 2D).

Once in the detailed page for each lncRNA peptide, the users can get extensive information of sequences and supporting evidences for lncRNA translation (Figure 2E). First, the basic information for lncRNA encoded peptides was provided in a table. In addition, the conservations of lncRNAs and peptides between species were analysed, including the identified novel homologous families. Detailed information for supporting evidences were also provided in the corresponding pages. Moreover, the binding between peptides and MHC molecules were shown as a network. Considering the strong tissue-specificity of lncRNAs, TransLnc allows users to access lncRNA expressions across



**Figure 2.** User interface and workflow of using TransLnc. (A) Search page for IncRNAs or peptides of interests. (B) Browse page for IncRNAs or peptides. (C) Interactive tools provided in TransLnc. (D) The result list for IncRNAs and peptides. (E) Detailed information for IncRNA encoded peptides with diverse supporting evidences. Conservation, spatial expression and clinical relevance in diverse cancer types are also provided in TransLnc resource.

tissues or diverse cancer types. Multiple visualization results were provided for facilitate understanding their roles in cancer. TransLnc provides the download of predicted and experimentally supported lncRNA peptides in the 'Download' page and a detail tutorial was provided in the 'Help' page for users showing how to use TransLnc.

## CASE STUDY

The lncRNAs HOXB-AS3 is a tumor suppressor and substantially down-regulated in several types of cancer (27). When querying the HOXB-AS3 by 'lncRNA Search' or 'Peptide Search' in TransLnc (Supplementary Figure S5A-B), we obtained 36 entries suggesting that this lncRNAs can encode peptides (Supplementary Figure S5C). In particular, we found that a peptide with 53 aa length was supported by four types of evidence (Supplementary Figure S5C), including manually curation, lncORF, m6A and ribosome-seq. Indeed, previous studies have demonstrated that HOXB-AS3 encoded a conserved 53-aa peptide, than can suppress cancer cell growth, colony formation and migration (27,57). Moreover, based on the expression of lncRNAs across cancer types, we found that high expression of lncRNAs was associated with better survival of cancer patients. This is consistent with the observation that high HOXB-AS3 peptide levels indicated a better prognosis for colorectal cancer patients (57). Another example is the lncRNA LINC00689, which has been found to play important roles in glioma development and progression (58). Moreover, we found that this lncRNAs can encode 73 potential peptides (Supplementary Figure S6). The lncRNA-encoded peptides were supported by multiple lines of evidence. Indeed, previous study has found that linc00689 can encode a micropeptide, STORM (Stress- and TNF- $\alpha$ -activated ORF Micropeptide) (59). Expression analysis indicated that LINC00689 was highly expressed in glioma patients, suggesting the oncogenic roles in cancer (Supplementary Figure S6).

## CONCLUSIONS AND FUTURE DEVELOPMENT

With the development of high-throughput technology and experimental validation methods, increasing evidence have suggested that lncRNA-encoded proteins play a critical role in cancer development and progression. However, identification of translatable lncRNAs and analysis of their clinical relevance is lagging. Here we reported an interactive database where multiple-omics evidences were integrated to identify lncRNA-encoded peptides. The database is easy to navigate with interactive tools were integrated, and all data can be downloaded for additional analysis. Importantly, we have computationally predicted the novel neoantigens from the peptides encoded by lncRNAs, which would provide new insights into the immunotherapy of cancer.

Although several database have been developed for describing lncRNAs genomic annotation (16), tissue-specific expression (17), regulation (5) or association with diseases (18), TransLnc is the first resource that focus on lncRNAs encoded peptides. Moreover, TransLnc also integrated useful tools for investigating the tissue expression patterns and clinical association for lncRNAs of interest. Thus,

TransLnc provides one-stop service for users to explore the function of lncRNAs in cancer. Moreover, there is still room for improvement of TransLnc. In the future, we will continue to update and integrate data content in TransLnc by: (i) continuous literature mining to add more experimentally validated lncRNA peptides, (ii) expanding the available Ribo-seq, m6A and IRES datasets to support the peptides, (iii) extending collection of proteomics data to further enhance the identification of ORF-encoded peptides. These additions are anticipated to enhance the efficiency of TransLnc and it will be an important resource for investigating the translation capacity of lncRNAs and the potential lncRNA encoded peptides. More rigorous investigation of lncRNAs and the small peptides hidden within them would lead to a profound understanding and provide new insights into cancer immunotherapy.

## DATA AVAILABILITY

TransLnc is an open comprehensive database for translatable lncRNAs, which is freely available at <http://bio-bigdata.hrbmu.edu.cn/TransLnc/>.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## FUNDING

National Natural Science Foundation of China [31871338, 31970646, 61873075, 32060152, 32070673, 32170676]; project supported by Hainan Province Clinical Medical Center, Hainan Provincial Natural Science Foundation of China [820MS053]; Key R&D projects in Hainan Province [ZDYF2021SHFZ051]; Major Science and Technology Program of Hainan Province [ZDKJ202003]; National Key R&D Program of China [2018YFC2000100]; National Science Foundation for Distinguished Young Scholars of Heilongjiang Province [JQ2019C004]; Heilongjiang Touyan Innovation Team Program. Funding for open access charge: National Natural Science Foundation of China.

*Conflict of interest statement.* None declared.

## REFERENCES

1. Fatica, A. and Bozzoni, I. (2014) Long non-coding RNAs: new players in cell differentiation and development. *Nat. Rev. Genet.*, **15**, 7–21.
2. Sarropoulos, I., Marin, R., Cardoso-Moreira, M. and Kaessmann, H. (2019) Developmental dynamics of lncRNAs across mammalian organs and species. *Nature*, **571**, 510–514.
3. Wang, K.C. and Chang, H.Y. (2011) Molecular mechanisms of long noncoding RNAs. *Mol. Cell*, **43**, 904–914.
4. Lin, X., Jiang, T., Bai, J., Li, J., Wang, T., Xiao, J., Tian, Y., Jin, X., Shao, T., Xu, J. *et al.* (2018) Characterization of transcriptome transition associates long noncoding RNAs with glioma progression. *Mol. Ther. Nucleic Acids*, **13**, 620–632.
5. Li, Y., Li, L., Wang, Z., Pan, T., Sahni, N., Jin, X., Wang, G., Li, J., Zheng, X., Zhang, Y. *et al.* (2018) lncMAP: Pan-cancer atlas of long noncoding RNA-mediated transcriptional network perturbations. *Nucleic Acids Res.*, **46**, 1113–1123.
6. Li, Y., Jiang, T., Zhou, W., Li, J., Li, X., Wang, Q., Jin, X., Yin, J., Chen, L., Zhang, Y. *et al.* (2020) Pan-cancer characterization of immune-related lncRNAs identifies potential oncogenic biomarkers. *Nat. Commun.*, **11**, 1000.



7. Xu, J., Shao, T., Song, M., Xie, Y., Zhou, J., Yin, J., Ding, N., Zou, H., Li, Y. and Zhang, J. (2020) MIR22HG acts as a tumor suppressor via TGFbeta/SMAD signaling and facilitates immunotherapy in colorectal cancer. *Mol. Cancer*, **19**, 51.
8. Zhang, J., Li, S., Zhang, L., Xu, J., Song, M., Shao, T., Huang, Z. and Li, Y. (2020) RBP EIF2S2 promotes tumorigenesis and progression by regulating MYC-mediated inhibition via FHIT-related enhancers. *Mol. Ther.*, **28**, 1105–1118.
9. Wang, H., Wang, Y., Xie, S., Liu, Y. and Xie, Z. (2017) Global and cell-type specific properties of lincRNAs with ribosome occupancy. *Nucleic Acids Res.*, **45**, 2786–2796.
10. Lu, S., Zhang, J., Lian, X., Sun, L., Meng, K., Chen, Y., Sun, Z., Yin, X., Li, Y., Zhao, J. et al. (2019) A hidden human proteome encoded by 'non-coding' genes. *Nucleic Acids Res.*, **47**, 8111–8125.
11. Kong, S., Tao, M., Shen, X. and Ju, S. (2020) Translatable circRNAs and lincRNAs: driving mechanisms and functions of their translation products. *Cancer Lett.*, **483**, 59–65.
12. Xu, K., Cai, Y., Zhang, M., Zou, H., Chang, Z., Li, D., Bai, J., Xu, J. and Li, Y. (2021) Pan-cancer characterization of expression and clinical relevance of m(6)A-related tissue-elevated long non-coding RNAs. *Mol. Cancer*, **20**, 31.
13. Bazin, J., Baerenfaller, K., Gosai, S.J., Gregory, B.D., Crespi, M. and Bailey-Serres, J. (2017) Global analysis of ribosome-associated noncoding RNAs unveils new modes of translational regulation. *PNAS*, **114**, E10018–E10027.
14. Plaza, S., Menschaert, G. and Payre, F. (2017) In search of lost small peptides. *Annu. Rev. Cell Dev. Biol.*, **33**, 391–416.
15. Othoum, G., Coonrod, E., Zhao, S., Dang, H.X. and Maher, C.A. (2020) Pan-cancer proteogenomic analysis reveals long and circular noncoding RNAs encoding peptides. *NAR cancer*, **2**, zcaa015.
16. Frankish, A., Diekhans, M., Ferreira, A.M., Johnson, R., Jungreis, I., Loveland, J., Mudge, J.M., Sisu, C., Wright, J., Armstrong, J. et al. (2019) GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res.*, **47**, D766–D773.
17. Lv, D., Xu, K., Jin, X., Li, J., Shi, Y., Zhang, M., Jin, X., Li, Y., Xu, J. and Li, X. (2020) LncSpA: LncRNA spatial atlas of expression across normal and cancer tissues. *Cancer Res.*, **80**, 2067–2071.
18. Gao, Y., Shang, S., Guo, S., Li, X., Zhou, H., Liu, H., Sun, Y., Wang, J., Wang, P., Zhi, H. et al. (2021) Lnc2Cancer 3.0: an updated resource for experimentally supported lincRNA/circRNA cancer associations and web tools based on RNA-seq and scRNA-seq data. *Nucleic Acids Res.*, **49**, D1251–D1258.
19. Hu, Z., Ott, P.A. and Wu, C.J. (2018) Towards personalized, tumour-specific, therapeutic vaccines for cancer. *Nat. Rev. Immunol.*, **18**, 168–182.
20. Hilf, N., Kutttruff-Coqui, S., Frenzel, K., Bukur, V., Stevanovic, S., Gouttefangeas, C., Platten, M., Tabatabai, G., Dutoit, V., van der Burg, S.H. et al. (2019) Actively personalized vaccination trial for newly diagnosed glioblastoma. *Nature*, **565**, 240–245.
21. Keskin, D.B., Anandappa, A.J., Sun, J., Tirosh, I., Mathewson, N.D., Li, S., Oliveira, G., Giobbie-Hurder, A., Felt, K., Gjini, E. et al. (2019) Neoantigen vaccine generates intratumoral T cell responses in phase Ib glioblastoma trial. *Nature*, **565**, 234–239.
22. Oka, M., Xu, L., Suzuki, T., Yoshikawa, T., Sakamoto, H., Uemura, H., Yoshizawa, A.C., Suzuki, Y., Nakatsura, T., Ishihama, Y. et al. (2021) Aberrant splicing isoforms detected by full-length transcriptome sequencing as transcripts of potential neoantigens in non-small cell lung cancer. *Genome Biol.*, **22**, 9.
23. Zhou, C., Wei, Z., Zhang, L., Yang, Z. and Liu, Q. (2020) Systematically characterizing A-to-I RNA editing neoantigens in cancer. *Front. Oncol.*, **10**, 593989.
24. Martinez, T.F., Chu, Q., Donaldson, C., Tan, D., Shokhirev, M.N. and Saghatelian, A. (2020) Accurate annotation of human protein-coding small open reading frames. *Nat. Chem. Biol.*, **16**, 458–468.
25. Chen, Y.G., Satpathy, A.T. and Chang, H.Y. (2017) Gene regulation in the immune system by long noncoding RNAs. *Nat. Immunol.*, **18**, 962–972.
26. Statello, L., Guo, C.J., Chen, L.L. and Huarte, M. (2021) Gene regulation by long non-coding RNAs and its biological functions. *Nat. Rev. Mol. Cell Biol.*, **22**, 96–118.
27. Wu, P., Mo, Y., Peng, M., Tang, T., Zhong, Y., Deng, X., Xiong, F., Guo, C., Wu, X., Li, Y. et al. (2020) Emerging role of tumor-related functional peptides encoded by lincRNA and circRNA. *Mol. Cancer*, **19**, 22.
28. Jackson, R., Kroehling, L., Khitun, A., Bailis, W., Jarret, A., York, A.G., Khan, O.M., Brewer, J.R., Skadon, M.H., Duizer, C. et al. (2018) The translation of non-canonical open reading frames controls mucosal immunity. *Nature*, **564**, 434–438.
29. Huang, Y., Wang, J., Zhao, Y., Wang, H., Liu, T., Li, Y., Cui, T., Li, W., Feng, Y., Luo, J. et al. (2021) cncRNAdb: a manually curated resource of experimentally supported RNAs with both protein-coding and noncoding function. *Nucleic Acids Res.*, **49**, D65–D70.
30. Zhou, B., Ji, B., Liu, K., Hu, G., Wang, F., Chen, Q., Yu, R., Huang, P., Ren, J., Guo, C. et al. (2021) EVLncRNAs 2.0: an updated database of manually curated functional long non-coding RNAs validated by low-throughput experiments. *Nucleic Acids Res.*, **49**, D86–D91.
31. Liu, H., Zhou, X., Yuan, M., Zhou, S., Huang, Y.E., Hou, F., Song, X., Wang, L. and Jiang, W. (2020) ncNEP: a manually curated database for experimentally validated ncRNA-encoded proteins or peptides. *J. Mol. Biol.*, **432**, 3364–3368.
32. Dragomir, M.P., Manyam, G.C., Ott, L.F., Berland, L., Knutsen, E., Ivan, C., Lipovich, L., Broom, B.M. and Calin, G.A. (2020) FuncPEP: a database of functional peptides encoded by non-coding RNAs. *Non-coding RNA*, **6**, 41.
33. Wang, H., Yang, L., Wang, Y., Chen, L., Li, H. and Xie, Z. (2019) RPFdb v2.0: an updated database for genome-wide information of translated mRNA generated from ribosome profiling. *Nucleic Acids Res.*, **47**, D230–D234.
34. Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet journal*, **17**, <https://doi.org/10.14806/ej.17.1.200>.
35. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. and Gingeras, T.R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15–21.
36. Liao, Y., Smyth, G.K. and Shi, W. (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, **30**, 923–930.
37. Ji, Z., Song, R., Regev, A. and Struhl, K. (2015) Many lincRNAs, 5'UTRs, and pseudogenes are translated and some are likely to express functional proteins. *eLife*, **4**, e08890.
38. Choteau, S.A., Wagner, A., Pierre, P., Spinelli, L. and Brun, C. (2021) MetamORF: a repository of unique short open reading frames identified by both experimental and computational approaches for gene and metagenome analyses. *Dataverse*, **2021**, baab032.
39. Brunet, M.A., Lucier, J.F., Levesque, M., Leblanc, S., Jacques, J.F., Al-Saedi, H.R.H., Guillo, N., Grenier, F., Avino, M., Fournier, I. et al. (2021) OpenProt 2021: deeper functional annotation of the coding potential of eukaryotic genomes. *Nucleic Acids Res.*, **49**, D380–D388.
40. Zhao, J., Li, Y., Wang, C., Zhang, H., Zhang, H., Jiang, B., Guo, X. and Song, X. (2020) IRESbase: a comprehensive database of experimentally validated internal ribosome entry sites. *Genomics Proteomics Bioinformatics*, **18**, 129–139.
41. Johnson, M., Zaretskaya, I., Raytselis, Y., Merezuk, Y., McGinnis, S. and Madden, T.L. (2008) NCBI BLAST: a better web interface. *Nucleic Acids Res.*, **36**, W5–W9.
42. Liu, S., Zhu, A., He, C. and Chen, M. (2020) REPIC: a database for exploring the N(6)-methyladenosine methylome. *Genome Biol.*, **21**, 100.
43. Meng, J., Cui, X., Rao, M.K., Chen, Y. and Huang, Y. (2013) Exome-based analysis for RNA epigenome sequencing data. *Bioinformatics*, **29**, 1565–1567.
44. Cui, X., Meng, J., Zhang, S., Chen, Y. and Huang, Y. (2016) A novel algorithm for calling mRNA m6A peaks by modeling biological variances in MeRIP-seq data. *Bioinformatics*, **32**, i378–i385.
45. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoutte, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W. et al. (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol.*, **9**, R137.
46. Quinlan, A.R. and Hall, I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.
47. Jones, P. and Cote, R. (2008) The PRIDE proteomics identifications database: data submission, query, and dataset comparison. *Methods Mol. Biol.*, **484**, 287–303.
48. Kong, A.T., Leprevost, F.V., Avtonomov, D.M., Mellacheruvu, D. and Nesvizhskii, A.I. (2017) MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. *Nat. Methods*, **14**, 513–520.

49. Ma, K., Vitek, O. and Nesvizhskii, A.I. (2012) A statistical model-building perspective to identification of MS/MS spectra with PeptideProphet. *BMC Bioinformatics*, **13**(Suppl. 16), S1.
50. Chen, J., Shishkin, A.A., Zhu, X., Kadri, S., Maza, I., Guttman, M., Hanna, J.H., Regev, A. and Garber, M. (2016) Evolutionary analysis across mammals reveals distinct classes of long non-coding RNAs. *Genome Biol.*, **17**, 19.
51. Bryzghalov, O., Szczesniak, M.W. and Makalowska, I. (2020) SyntDB: defining orthologues of human long noncoding RNAs across primates. *Nucleic Acids Res.*, **48**, D238–D245.
52. Li, L., Stoeckert, C.J. and Roos, D.S. (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.*, **13**, 2178–2189.
53. Reynisson, B., Alvarez, B., Paul, S., Peters, B. and Nielsen, M. (2020) NetMHCpan-4.1 and NetMHCIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Res.*, **48**, W449–W454.
54. Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A.A., Kim, S., Wilson, C.J., Lehar, J., Kryukov, G.V., Sonkin, D. *et al.* (2012) The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*, **483**, 603–607.
55. Huntley, M.A., Lou, M., Goldstein, L.D., Lawrence, M., Dijkgraaf, G.J., Kaminker, J.S. and Gentleman, R. (2016) Complex regulation of ADAR-mediated RNA-editing across tissues. *BMC Genomics*, **17**, 61.
56. Sollner, J.F., Lepar, G., Hildebrandt, T., Klein, H., Thomas, L., Stupka, E. and Simon, E. (2017) An RNA-Seq atlas of gene expression in mouse and rat normal tissues. *Sci Data*, **4**, 170185.
57. Huang, J.Z., Chen, M., Chen, D., Gao, X.C., Zhu, S., Huang, H., Hu, M., Zhu, H. and Yan, G.R. (2017) A peptide encoded by a putative lncRNA HOXB-AS3 suppresses colon cancer growth. *Mol. Cell*, **68**, 171–184.
58. Zhan, W.L., Gao, N., Tu, G.L., Tang, H., Gao, L. and Xia, Y. (2021) LncRNA LINC00689 promotes the tumorigenesis of glioma via mediation of miR-526b-3p/IGF2BP1 axis. *NeuroMol. Med.*, **23**, 383–394.
59. Min, K.W., Davila, S., Zealy, R.W., Lloyd, L.T., Lee, I.Y., Lee, R., Roh, K.H., Jung, A., Jemielity, J., Choi, E.J. *et al.* (2017) eIF4E phosphorylation by MST1 reduces translation of a subset of mRNAs, but increases lncRNA translation. *Biochim. Biophys. Acta Gene Regul. Mech.*, **1860**, 761–772.