frontiers in Genetics

REVIEW
published: 31 January 2020
doi: 10.3389/fgene.2019.01346

Probing lncRNA–Protein Interactions: Data Repositories, Models, and Algorithms

Lihong Peng[1†], Fuxing Liu[1†], Jialiang Yang[2], Xiaojun Liu[1], Yajie Meng[3], Xiaojun Deng[1], Cheng Peng[1], Geng Tian[2] and Liqian Zhou[1*]*

[1] School of Computer Science, Hunan University of Technology, Zhuzhou, China, [2] Department of Sciences, Genesis (Beijing) Co. Ltd., Beijing, China, [3] College of Computer Science and Electronic Engineering, Hunan University, Changsha, China

Identifying lncRNA–protein interactions (LPIs) is vital to understanding various key biological processes. Wet experiments found a few LPIs, but experimental methods are costly and time-consuming. Therefore, computational methods are increasingly exploited to capture LPI candidates. We introduced relevant data repositories, focused on two types of LPI prediction models: network-based methods and machine learning-based methods. Machine learning-based methods contain matrix factorization-based techniques and ensemble learning-based techniques. To detect the performance of computational methods, we compared parts of LPI prediction models on Leave-One-Out cross-validation (LOOCV) and fivefold cross-validation. The results show that SFPEL-LPI obtained the best performance of AUC. Although computational models have efficiently unraveled some LPI candidates, there are many limitations involved. We discussed future directions to further boost LPI predictive performance.

Keywords: lncRNA–protein interaction, computational method, network-based method, machine learning-based method, data repositories

OPEN ACCESS

Edited by:
Quan Zou,
University of Electronic Science and Technology of China, China

Reviewed by:
Guohua Huang,
Shaoyang University, China
Guoxian Yu,
Southwest University, China

****Correspondence:***
Geng Tian
tiang@genesis.com
Liqian Zhou
zhoulq11@163.com

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Bioinformatics and Computational Biology, a section of the journal Frontiers in Genetics

Received: *27 September 2019*
Accepted: *09 December 2019*
Published: *31 January 2020*

Citation:
Peng L, Liu F, Yang J, Liu X, Meng Y, Deng X, Peng C, Tian G and Zhou L (2020) Probing lncRNA–Protein Interactions: Data Repositories, Models, and Algorithms. Front. Genet. 10:1346. doi: 10.3389/fgene.2019.01346

INTRODUCTION

Long non-coding RNAs (lncRNAs) are transcripts with greater than 200 nucleotides but lack protein coding capacity (Sanchez Calle et al., 2018). lncRNAs are closely associated with various key biological processes, such as cell cycle regulation, immune response, and embryonic stem cell pluripotency (Liu et al., 2018; Agirre et al., 2019; Li et al., 2019b). More importantly, lncRNAs play an important role in understanding pathogenesis of various diseases, especially tumors (Chen et al., 2016a; Fu et al., 2017; Jiang et al., 2018; He et al., 2018a; Dallner et al., 2019). Although lncRNAs play a spectrum of regulatory roles across different cellular pathways, understanding about their regulatory mechanisms is very limited (Munschauer et al., 2018).

Recently, one broad theme is that lncRNAs can drive the assembly of RNA–protein complexes by facilitating the regulation of gene expression (Rinn and Chang, 2012; Chen and Yan, 2013; Hentze et al., 2018; Munschauer et al., 2018; Nozawa and Gilbert, 2019). lncRNAs achieve their specific functions by interacting with multiple proteins and thus regulating multiple cellular processes (Zhang et al., 2018c; Pyfrom et al., 2019). Studies reported that lncRNAs can activate post-transcriptional gene regulation, splicing, and translation by binding to proteins (Zhang et al,. 2018c; Li et al., 2019a) Therefore, identifying possible lncRNA–protein interactions (LPIs) is

essential for unraveling lncRNA-related activities (Qian et al., 2018; Zhang et al., 2018c; Zhao et al., 2018c). Wet experiments validated parts of LPIs, but experimental methods remain costly and time-consuming. Therefore, different computational models are explored to infer potential LPIs (Pan et al., 2016; Cheng et al., 2018; Zhang et al., 2018c; Zhao et al., 2018c). There exist numerous unexplored lncRNAs and proteins in public databases, which makes it possible to efficiently identify their underlying associations.

In this study, we introduced relevant repositories, summarized computational models and algorithms for LPI prediction, discussed their advantages and weaknesses by comparison, and presented further directions for boosting LPI prediction performance. We focused on two categories of computational models: network-based methods and machine learning-based methods. The machine learning-based methods contain matrix factorization-based methods and ensemble learning-based methods.

RELEVANT REPOSITORIES

There are abundant repositories related to LPI prediction. These repositories provide diverse information for efficiently uncovering potential LPIs.

Noncode

The NONCODE database (Zhao et al., 2015) (http://www.noncode.org/) is an interactive database aiming to collect the most complete annotation for ncRNAs, especially lncRNAs. The latest NONCODE database (current version v5.0) contains lncRNA information from 17 species including human, mouse, cow, rat, chimp, gorilla, orangutan, rhesus, opossum, platypus, chicken, zebrafish, fruit fly, *Caenorhabditis elegans*, yeast, Arabidopsis, and pig. There are 548,640 lncRNAs in the latest version. There are 172,216 and 131,697 lncRNAs from human and mouse, respectively. More importantly, NONCODE has introduced some important features including conservation annotation, lncRNA–disease associations, and an interface to select credible datasets.

NPInter

The NPInter database (Hao et al., 2016) (http://www.bioinfo.org.cn/NPInter/contact.htm) provides abundant association data that are experimentally verified. For example, the database contains information on interactions between noncoding RNAs (ncRNAs) and biomolecules including proteins, mRNAs, miRNAs, and genomic DNAs. The database contains 491,416 interactions in 188 tissues/cell lines from 68 types of experimental technology.

RAID

The RAID database (Yi et al., 2016) (http://www.rna-society.org/raid/) includes more than 40,668 lncRNA-associated RNA–protein interactions and more than 34,790 lncRNA-associated RNA–RNA interactions.

starBase

The starBase database (Li et al., 2013) (http://starbase.sysu.edu.cn/) contains more than 1,100,000 miRNA–ncRNA (CLIP) interactions, 117,000 RNA-binding protein (RBP)–ncRNA interactions, and 32,000 miRNA–ncRNA interactions. In addition, it provides more than 10,800 RNA-seq data and 10,500 miRNA-seq data from 32 cancer types and 3,236,000 mutations from 366 disease types.

VirBase

The ViRBase database (Li et al., 2014) (http://www.rna-society.org/virbase) integrates experimental and predictive association information from manual literature curation and other resources based on one common framework from 119 species, especially ncRNA-associated virus–virus, host–host, host–virus, and virus–host interactions.

POSTAR2

The POSTAR2 database (Zhu et al., 2018) (http://lulab.life.tsinghua.edu.cn/postar2/index.php) provides various post-transcriptional regulation data based on CLIP-seq, Ribo-seq, RNA-seq, and other high-throughput sequencing information from six species: yeast, Arabidopsis, fly, worm, mouse, and human. It hosts about 40 million RBP binding sites validated by CLIP-seq experiments. It provides three modules: the "RBP" module, "RNA" module, and "Translatome" module. The "RBP" module contains RBP binding sites and their annotations and functions. The "RNA" module is composed of a few sub-modules, including "disease," "variation," "crosstalk," and "binding sites," and is applied to annotate the RBP binding sites.

ChIPBase

The ChIPBase database (Zhou et al., 2016) (http://rna.sysu.edu.cn/chipbase/) is used to identify transcription factor binding sites and motifs, and decode transcriptional regulatory networks of miRNA, lncRNAs, and other ncRNAs from ChIP-seq data. It provides about 10,200 curated peak datasets from 10 species: human, mouse, fruit fly, worm, *Arabidopsis thaliana*, yeast, rat, zebrafish, *Xenopus tropicalis*, and chicken.

LNCipedia

The LNCipedia database (Volders et al., 2018) (https://lncipedia.org) is a comprehensive database. Its central work is to merge redundant transcripts from different data sources and group the transcripts into genes, thus producing a highly consistent database. The latest update of lncRNA (LNCipedia 5) contains information about annotation and sequence for 1,555 human lncRNAs from 2,482 lncRNA publications. This information originates from Ensembl (Cunningham et al., 2018), RefSeq (Rajput et al., 2018), and FANTOM CAT (Hon et al., 2017).

lncRNA2target

The lncRNA2Target database (Cheng et al., 2018) (http://123.59.132.21/lncrna2target) contains a comprehensive repository of lncRNA target genes to provide information about target genes regulated by lncRNAs. The latest version provides a special web

interface in which users can search the targets for a particular lncRNA or the lncRNAs for a particular gene.

lncRNAdb

The lncRNAdb database (Quek et al., 2014s) (http://lncrnadb.org) is a comprehensive database in compliance with the International Nucleotide Sequence Database Collaboration. It provides 287 eukaryotic lncRNAs and an interface enabling users to access sequence data, expression information, and the literature. The latest update of lncRNAdb integrated nucleotide sequence information, Illumina Body Atlas expression profiles, and a BLAST search tool.

lncRNASNP2

The lncRNASNP2 database (Miao et al., 2017) (http://bioinfo.life.hust.edu.cn/lncRNASNP2) provides 7,260,238 single nucleotide polymorphisms (SNPs) on 141,353 human lncRNA transcripts, and 3,921,448 SNPs on 117,405 mouse lncRNA transcripts. More importantly, it contains abundant information about mutations in lncRNAs and their impacts on lncRNA structure and function. It also provides online tools for analyzing new variants in lncRNA.

LbcRNAwiki

The lncRNAWiki database (Ma et al., 2014) (http://lncrna.big.ac.cn) integrated various human lncRNAs from different resources. It makes existing lncRNAs able to be updated, edited, and curated by diverse users. More importantly, any user can add newly uncovered lncRNAs.

Lnc2Cancer

The Lnc2Cancer database (Gao et al., 2018) (http://www.bio-bigdata.net/lnc2cancer) provides lncRNA–cancer associations supported by experiments. It contains 4,989 associations between 165 human cancer subtypes and 1,614 human lncRNAs, 366 experimentally validated circulating-related lncRNA-cancer associations, 593 drug-resistance-related lncRNA-cancer associations, and 1,928 prognosis-related lncRNA–cancer associations, and abundant lncRNA regulatory mechanisms in cancers including 211, 1139, 225, and 319 lncRNAs regulated by variant, miRNA, transcription factor, and methylation, respectively.

LncRNAdisease

The lncRNADisease database (Bao et al., 2018) (http://www.rnanut.net/lncrnadisease/) integrated experimentally validated circular RNA–disease associations, and regulatory mechanisms among mRNA, miRNA, and ncRNA. Particularly, it contains more than 200, 000 lncRNA–disease associations. In addition, it gives confidence scores for all ncRNA–disease associations and maps each disease to disease ontology and medical subject headings.

MNDR

The MNDR database (Cui et al., 2017) (http://www.rna-society.org/mndr/) integrates more than 260,000 ncRNA–disease associations. These associations are supported by 10 experiments and 4 predictive algorithms. The experimental repositories include Lnc2Cancer (Gao et al., 2018), dbDEMC (Yang et al,. 2016), LncRNADisease (Bao et al., 2018), MNDR (Wang et al., 2013), HMDD (Huang et al., 2018b), NSDNA (Wang et al., 2016a), LincSNP (Ning et al., 2016), miRCancer (Xie et al., 2013), PhenomiR (Ruepp et al., 2012), and miR2Disease (Jiang et al., 2008). The four prediction algorithms are LDAP (Lan et al., 2016), miRDP (Mørk et al., 2013) LncDisease (Wang et al., 2016b), and PBMDA (You et al., 2017). It provides 8,824 experimental lncRNA–disease, 70,381 experimental miRNA–disease, 118 experimental piRNA–disease, and 67 experimental snoRNA–disease associations across 6 mammalsix (*Homo sapiens*, *Macaca mulatta*, *Mus musculus*, *Pan troglodyte*, *Rattus norvegicus*, and *Sus scrofa*). In addition, it provides 153,508 predicted lncRNA–disease associations and 28,144 predicted miRNA–disease associations for *H. sapiens*. MNDR contains 19,575, 110, 4,150, and 23 non-redundant lncRNA–disease, piRNA–disease, miRNA–disease, and snoRNA–disease interactions, respectively, associated with 1,416 disease.

UniProt

The UniProt database (Consortium et al., 2018) (http://www.uniprot.org/) is an important database providing protein sequences and annotations. It provides 80 million sequences and is a useful tool. Users can calculate a new proteome identifier to find a particular assembly for a species or subspecies. It also provides an effective measurement for computing an annotation score for all entries.

METHODS

Most computational methods contain two procedures: data extraction and model selection. In the first part, computational methods usually extract LPIs related to human lncRNA, lncRNA sequences, and protein sequences from NPInter (Hao et al., 2016), NONCODE (Zhao et al., 2015), and UniProt (Consortium et al., 2018), respectively. Computational methods filter LPIs by removing lncRNAs/proteins only interacting with one protein/lncRNA. In the second procedure, computational methods design various models to uncover potential LPIs. These models can be roughly classified into two categories: network-based methods and machine learning-based methods.

Data Representation

Computational methods utilize an lncRNA set $l = \{l_1, l_2, l_3 \ldots. l_n\}$, a protein set, $P = \{p_1, p_2, p_3 \ldots. p_m\}$, and an LPI matrix $Y_{n\times m}$, where $y_{ij} = 1$ if there is an association between an lncRNA l_i and a protein p_j; otherwise, $y_{ij} = 0$.

Network-Based Methods

Network-based methods obtain better performance by effectively integrating related biological information and network propagation algorithms into a unified framework.

LPIHN

Li et al. (2015) developed an LPI prediction method combing a heterogeneous network model and random walk with restart, LPIHN. LPIHN can be broken down into four steps:

Step 1 Extracting known ncRNA–protein associations from the Npinter 2.0 database (Hao et al., 2016) and filtering the ncRNAs and their associated proteins based on organism and type of ncRNAs. LPIHN then selects lncRNAs from filtered ncRNAs based on the human lncRNA dataset provided by the NONCODE database (Zhao et al., 2015).

Step 2 Obtaining lncRNA expression profiles from the NONOCODE 4.0 database (Zhao et al., 2015). Given the expression profiles of two lncRNAs E_1 and E_2, LPIHN calculates lncRNA expression similarity based on the Pearson correlation coefficient:

$$SL(i,j) = \left| \frac{\mathrm{cov}(E_1, E_2)}{\sigma_{e_1} \sigma_{e_2}} \right| \tag{1}$$

where $cov(E_1, E_2)$ is the covariance of E_1 and E_2, and σ_{e_1} and σ_{e_2} are the standard deviations of E_1 and E_2, respectively.

Step 3 Extracting protein–protein interactions (PPIs) from STRING 9.1 (Szklarczyk et al., 2016) and obtaining 804 PPIs and the corresponding score matrix *SP*. *SP* is normalized as follows:

$$SP_{ij}^{\star} = \frac{SP_{ij}}{\sqrt{M(i,i)M(j,j)}} \tag{2}$$

where *M* is a diagonal matrix, and $M(i, i)$ is the sum of row *i* in *SP*.

Step 4 Propagating the random walk to score for unknown lncRNA–protein pairs based on the following iterative equation:

$$Y_{t+1} = (1-\delta) W^T Y_t + \delta\, Y_0 \tag{3}$$

The details are shown as **Figure 1**.

LPLNP

Zhang et al. (2018b) proposed a linear neighborhood propagation-based method, LPLNP, to probe potential LPIs. LPLNP found novel LPIs through the following steps.

Step 1 Extracting 4,158 LPIs between 27 proteins and 990 lncRNAs from NPInter (Hao et al., 2016) and NONCODE (Zhao et al., 2015) by filtering unreliable lncRNA sequences and removing lncRNAs/proteins only interacting with one protein/lncRNA.

Step 2 Obtaining three types of features for lncRNAs (interaction profile, expression profile, and sequence composition) and two types of features for proteins [interaction profile and CTD (composition, transition, and destruction)].

Step 3 Computing linear neighborhood similarity and regularized linear neighborhood similarity between lncRNA/proteins by Eqs. (4) and (5), respectively:

$$\begin{aligned} &\epsilon_i = \| X_i - \sum_{i_j : X_{i_j} \in N(X_i)} w_{ii_j} X_{i_j} \|^2 \\ &s.t. \sum_{i_j : X_{i_j} \in N(X_i)} w_{ii_j} = 1, w_{ii_j} \geq 0 \end{aligned} \tag{4}$$

where X_i denoted the feature vector of the *i*th lncRNA, and $N(X_i)$ is *K* nearest neighbors of X_i.

$$\begin{aligned} &\epsilon_i = w_i^T (G^i + \lambda I) w_i \\ &s.t. \sum_{i_j : X_{i_j} \in N(X_i)} w_{ii_j} = 1, w_{ii_j} \geq 0 \end{aligned} \tag{5}$$

where $G_{i_j i_k} = (X_i - X_{i_j})^T (X_i - X_{i_j})$.

Step 4 Computing the interaction probabilities for unobserved lncRNA–protein pairs:

$$Y = (1-\alpha)(I - \alpha W)^{-1} Y^0 \tag{6}$$

The details are shown in **Figure 2**.

LPI-BNPRA

Zhao et al. (2018a) developed a novel LPI prediction model based on a bipartite network projection recommended technique, LPI-BNPRA. LPI-BNPRA can be broken down into five steps.

Step 1 Extracting 4,158 high-confidence LPIs between 990 lncRNAs and 27 proteins from NPInter (Hao et al., 2016) and NONCODE (Zhao et al., 2015) by filtering unreliable lncRNA sequences and removing lncRNAs/proteins only associated with one protein/lncRNA.

Step 2 Calculating lncRNA–lncRNA similarity based on the Smith–Waterman technique:

$$LSM(l_i, l_j) = \frac{sw(l_i, l_j)}{\max\ (sw(l_i, l_i), sw(l_j, l_j))} \tag{7}$$

where $sw(l_i, l_j)$ denotes the Smith–Waterman score between two lncRNAs l_i and l_j.

Step 3 Calculating the protein–protein similarity matrix based on the Smith–Waterman technique:

$$PSM(p_i, p_j) = \frac{sw(p_i, p_j)}{\max\left(sw(p_i, p_i), sw(p_j, p_j)\right)} \tag{8}$$

where $sw(p_i, p_j)$ denotes the Smith–Waterman score between two proteins p_i and p_j.

Step 4 For a given lncRNA l_j, computing its bias ratings of lncRNAs for a protein p_i with the agglomerative hierarchical clustering and associated measurement of minimum variance method:

$$r(p_i, l_j) = \frac{n_{cr}}{T(p_i)} \tag{9}$$

where n_{cr} is the number of lncRNAs in the cluster *cr* including l_j, and Tp_i is the number of all lncRNAs interacting with p_i.

Step 5 Finding LPI candidates based on the recommended bipartite network projection technique and bias ratings of every lncRNA for proteins:

$$R_{fin}(l_j) = \sum_{i=1}^{n} R_{fin}(p_i, l_j) \tag{10}$$

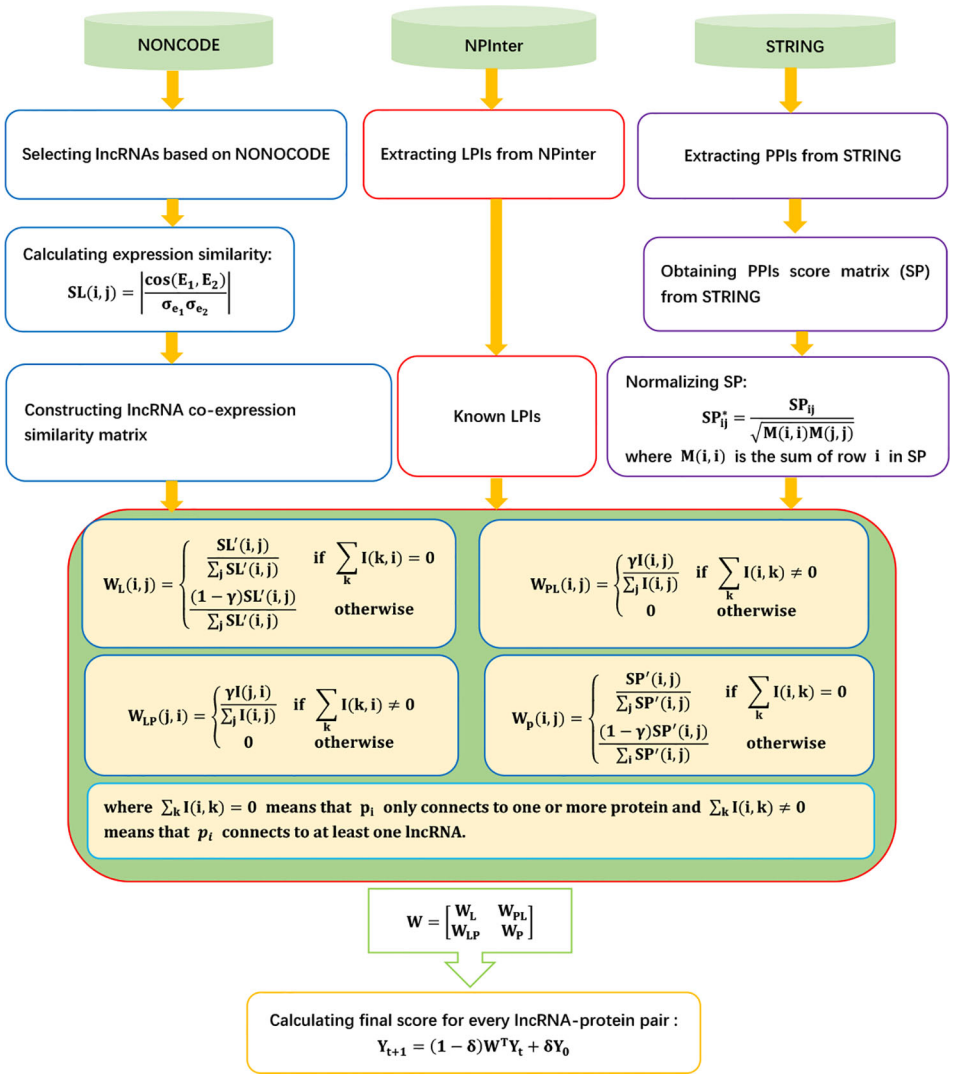

FIGURE 1 | Flowchart of LPI prediction method based on heterogeneous network model and random walk with restart.

where

$$R_{fin}(p_i,l_j)=\frac{r(p_i,l_j)}{\sum_{k=1}^{n}r(p_k,l_j)}\times R(p_i) \tag{11}$$

$$R(p_i)=\sum_{j=1}^{m}R(p_i,l_j) \tag{12}$$

$$R(p_i,l_j)=\frac{r_{ini}(p_i,l_j)}{\sum_{k=1}^{n}r_{ini}(p_k,l_j)}\times R_{ini}(l_j) \tag{13}$$

$$R_{ini}(l_j)=r_{ini}(p_i,l_j) \tag{14}$$

$$r_{ini}(p_i,l_j)=\frac{r(p_i,l_j)}{r_{ave}(p_i,l_j)} \tag{15}$$

$$r(p_i,l_j)=\frac{n_{cr}}{T(p_i)} \tag{16}$$

$$r_{ave}(p_i,l_j)=\frac{\sum_{j=1}^{m}r(p_i,l_j)}{T(p_i)} \tag{17}$$

The details are shown in **Figure 3**.

LPISNFHS

Zheng et al. (2017) presented a new LPI identification method, LPISNFH. LPISNFHS fused multiple protein–protein similarity networks, the similarity network fusion (SNF)

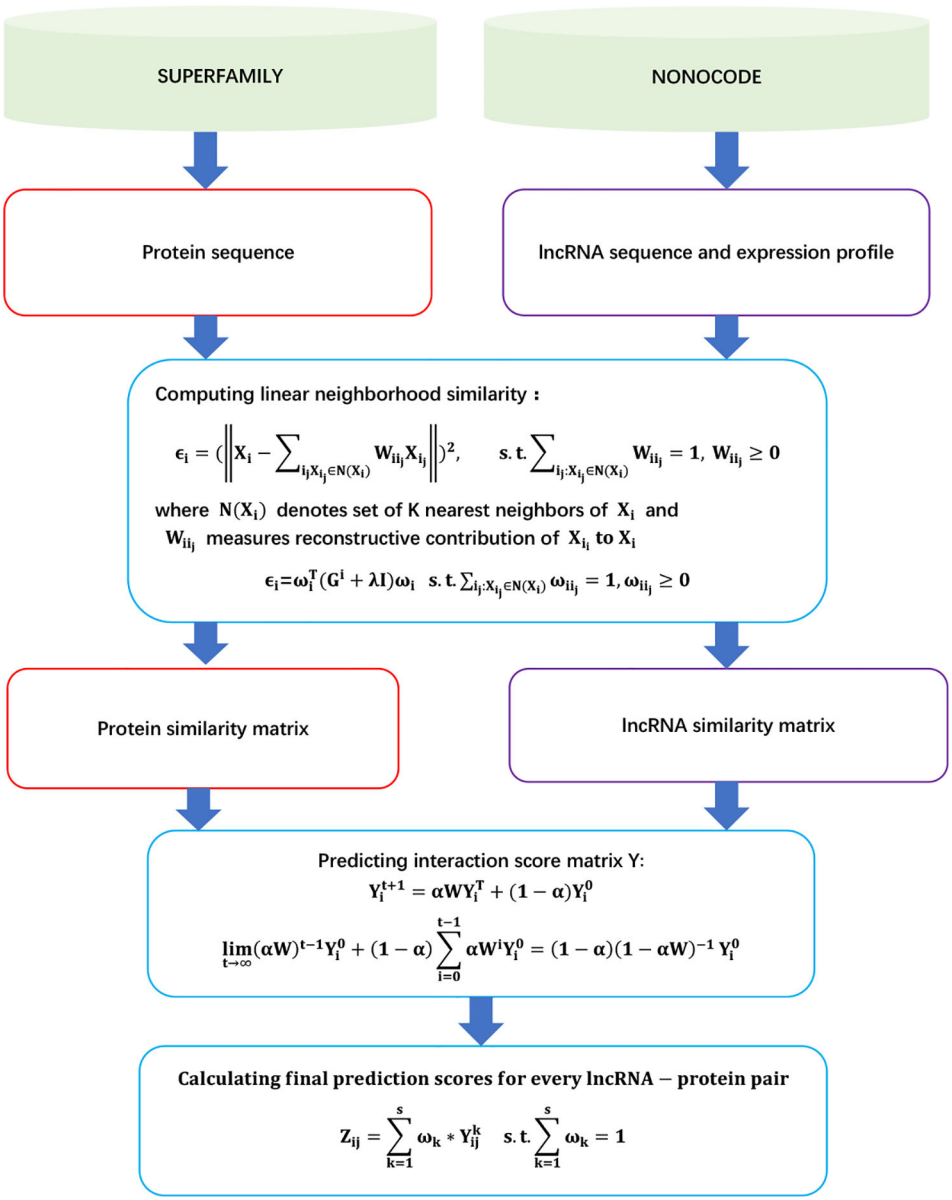

FIGURE 2 | Flowchart of linear neighborhood propagation-based LPI prediction method.

technique, HeteSim algorithm, and known LPI network into a unified framework. LPISNFH can be broken down into three steps.

Step 1 Obtaining 4,467 LPIs between 1,050 unique lncRNAs and 84 unique proteins from NPInter (Hao et al., 2016) and NONCODE (Zhao et al., 2015) by manually filtering LPIs not involving lncRNAs and removing the lncRNAs only associated with one protein.

Step 2 Constructing a protein–protein similarity network. LPISNFHS fused the sequence similarity, functional annotation semantic similarity (Go), domain similarity, and STRING similarity into a unified protein–protein similarity network based on the SNF technique.

Step 3 Inferring novel LPIs by combining the HeteSim algorithm and heterogeneous LPI network.

LPI-IBNRA

Xie et al. (2019) developed a LPI prediction model, LPI-IBNRA. LPI-IBNRA integrated lncRNA–protein interactions, protein–protein interactions, and similarity matrix for proteins and lncRNAs, and improved bipartite network recommender algorithm. LPI-IBNRA can be broken down into seven steps.

Step 1 Obtaining 4,796 LPIs between 1,105 lncRNAs and 26 proteins from NPInter (Hao et al., 2016) and NONCODE (Zhao et al., 2015) after filtering lncRNAs and proteins that have only one association.

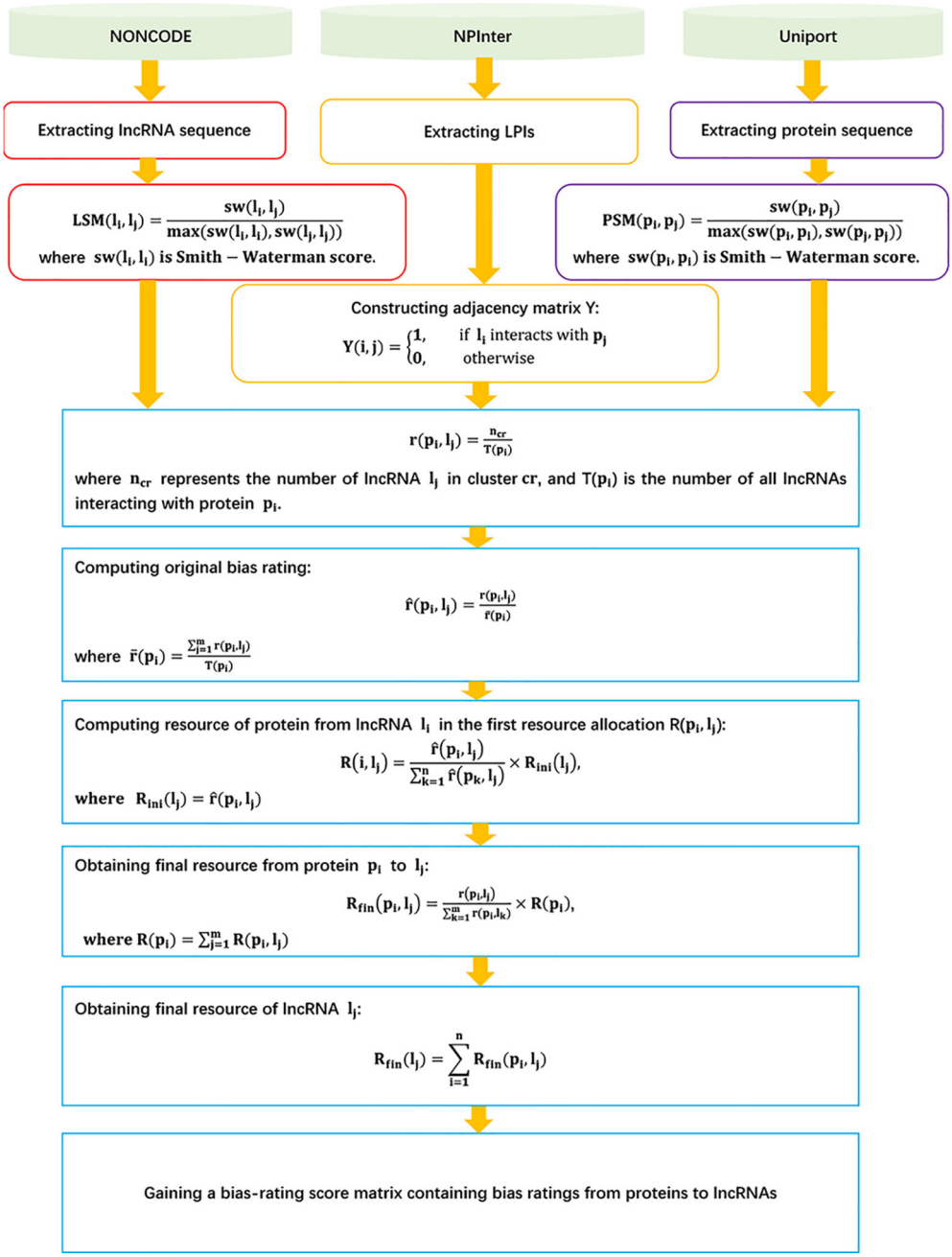

FIGURE 3 | Flowchart of LPI prediction model based on the recommended bipartite network projection technique.

Step 2 Computing lncRNA similarity matrix sim^L based on lncRNA expression similarity and Gaussian interaction profile (GIP) kernel similarity, and protein similarity matrix sim^P based on protein interaction similarity and GIP kernel similarity.

Step 3 Computing the score between protein p_i and lncRNA l_j based on protein similarity and lncRNA similarity by Eqs. (18) and (19), respectively.

$$S^P(p_i, l_j) = \begin{cases} \dfrac{\sum_{k=1}^{np} sim^P(p_i, p_k) I(p_k, l_j)}{\sum_{k=1}^{np} sim^P(p_i, p_k)} & if \quad I(p_i, l_j) = 1 \\ 0 & otherwise \end{cases} \tag{18}$$

$$S^L(p_i, l_j) = \begin{cases} \dfrac{\sum_{k=1}^{nl} I_{(p_i,l_k)} sim^L(l_k, l_j)}{\sum_{k=1}^{np} sim^L(l_k, l_j)} & if \quad I(p_i, l_j) = 1 \\ 0 & otherwise \end{cases} \tag{19}$$

Step 4 Obtaining the initialized association score matrix as follows:

$$S_{ini} = \gamma S^P + (1-\gamma) S^L \tag{20}$$

Step 5 Computing the first-round scores of the lncRNA l_k over all proteins:

$$s_1(l_k) = \sum_{j=1}^{np} \frac{S_{ini}(p_j, l_k) s_0(p_j)}{d(p_j)} \tag{21}$$

Step 6 Computing the second-round scores of the protein p_i over all lncRNAs:

$$s_2(p_i) = \sum_{k=1}^{nl} \frac{S_{ini}(p_i, l_k)}{d(l_k)} \sum_{k=1}^{np} \frac{S_{ini}(p_j, l_k) s_0(p_j)}{d(p_j)} \tag{22}$$

Step 7 Computing the final association score matrix:

$$S'_{fin} = W' S_{ini} \tag{23}$$

where $W' = W + \alpha W^2$ and $\alpha \in (-1,0)$.

The details are shown in **Figure 4**.

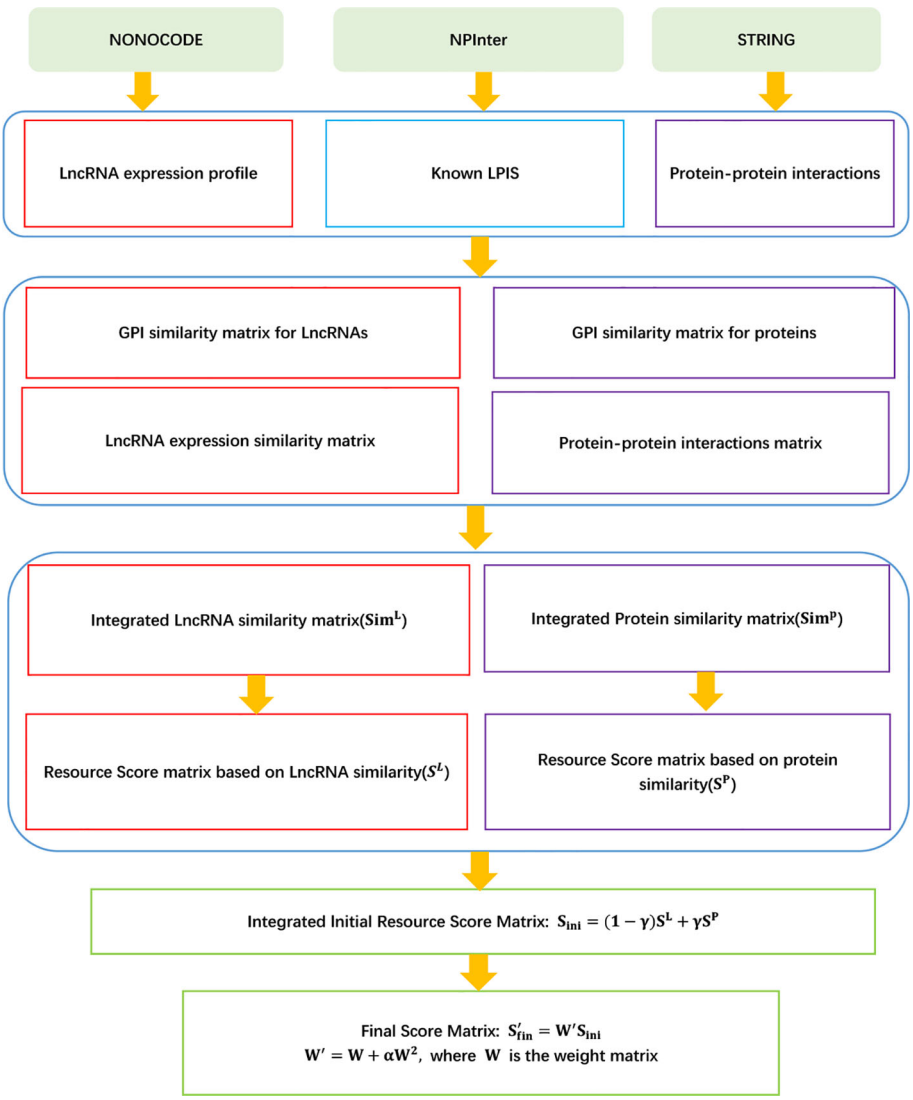

FIGURE 4 | Flowchart of LPI prediction method based on improved bipartite network recommender algorithm.

LPBNI

Ge et al. (2016) proposed an lncRNA–protein bipartite network inference method, LPBNI, to find potential LPIs. LPBNI can be broken down into five steps.

Step 1 Extracting data. LPBNI first downloads 7,576 ncRNA–protein associations from NPInter 2.0 (Hao et al., 2016) with the restricted type of "NONCODE" and organism "*Homo sapiens*." LPBNI then selects 2,380 lncRNAs based on a human lncRNA dataset provided by the NONCODE database (Zhao et al., 2015). Finally, LPBNI extracts 4,870 LPIs between 2,380 lncRNAs and 106 proteins.

Step 2 Utilizing the LPI network to construct a bipartite graph G (L, P, Y).

Step 3 Propagating known biological information in G. For a lncRNA l_j, S_L (l_j) denotes the score on l_j after the first step of propagation:

$$S_L(l_j) = \sum_{i=1}^{m} \frac{a_{ij} S_0(i)}{d(p_i)}, j \in \{1, 2, 3 \ldots . n\} \tag{24}$$

where S_0 $(i) = s_{ij}$, $i \in \{i, 2, \ldots, m\}$ denotes the original information of P for a given lncRNA l_j. $s_{ij} = 1$ if p_i associates with l_j; otherwise, $s_{ij} = 0$. $d(p_i) = \sum_{j=1}^{n} a_{ij}$ denotes the number of lncRNAs associated with p_i.

Step 4 Propagating all information in L back to P. $S_F(p_i)$ represents the final information on protein p_i to denote the associated score between p_i and l_j:

$$S_F(i) = \sum_{j=1}^{n} \frac{a_{ij} S_L(l_j)}{d(l_j)} = \sum_{j=1}^{n} \frac{a_{ij}}{d(l_j)} \sum_{k=1}^{m} \frac{a_{kj} S_0(k)}{d(p_k)} \tag{25}$$

where $d(l_i) = \sum_{i=0}^{m} a_{ij}$ is the number of proteins interacting with l_j.

Step 5 Computing the final associated score S_F after the above two-step information propagation yields

$$\vec{S}_F = W\overrightarrow{S_0} \tag{26}$$

where $\vec{S}_0$ denotes the column vector of S_0, $S_F(i) = \sum_{k=1}^{m} w_{ik} S_0(k)$, where $wij = \frac{1}{d(pi)} \sum_{j=1}^{n} \frac{a_{ij} a_{kj}}{d(lj)}$.

The details are shown in **Figure 5**.

ACCBN

Zhu et al. (2019) exploited an ant-colony-clustering-based bipartite network method for revealing potential LPIs, ACCBN. The model can be roughly broken down into three steps.

Step 1 Describing lncRNA interaction profiles and protein interaction profiles as row vectors and column vectors based on the LPI network, respectively.

Step 2 Calculating the probability that two entities x_i and x_j belong to the same cluster based on the ant colony clustering method:

$$p_{ij}(t) = \frac{\left[T_{ij}(t)\right]^{\alpha} \left[\eta_{ij}(t)\right]^{\beta}}{\sum_{j=1}^{k} \left[T_{ij}(t)\right]^{\alpha} \left[\eta_{ij}(t)\right]^{\beta}} \tag{27}$$

where

$$\eta_{ij} = \frac{1}{d_{ij}} \tag{28}$$

$$d_{ij} = \left(\sum_{k=1}^{m} |x_{ik} - x_{jk}|^2\right)^{\frac{1}{2}} \tag{29}$$

$$T_{ij}(t+1) = (1-\rho) T_{ij}(t) + \Delta T_{ij}(t) \tag{30}$$

$$T_{ij}(t) = \begin{cases} 1 & d_{ij} \le r \\ 0 & d_{ij} > r \end{cases} \tag{31}$$

$$\Delta T_{ij}(t) = \frac{Q}{d(x_i, c_j)} \tag{32}$$

where r is the cluster radius, c_j is the cluster center of the jth cluster, and $\alpha \in (0, 5)$, $\beta \in (0, 5)$, $\rho \in (0.1, 0.99)$, and $Q \in (1, 10000)$.

Step 3 Applying lncRNA–protein bipartite network to identify LPI candidates. Given a protein p_k, its association scores with all lncRNAs at the tth iteration P_k^t can be computed as follows:

$$P_k^t = \rho W P_k^{t-1} + (1-\rho) M(:, k) \tag{33}$$

where W is a similarity matrix.

The association scores for all proteins $\{p_1, p_2, \ldots, pm\}$ can be represented as follows:

$$P^t = \rho W P^{t-1} + (1-\rho) M \tag{34}$$

Machine Learning-Based Methods

Machine learning-based LPI prediction methods utilize machine learning-based models and algorithms to uncover potential LPIs. This type of method can be roughly classified into two categories: matrix factorization-based methods and ensemble learning-based methods.

Matrix Factorization-Based Models

Matrix factorization is exploited in recommendation systems and has been widely applied to bioinformatics (Shi et al., 2018; Zhang et al., 2018a; Zhao et al., 2018b; Cantini et al., 2019). Matrix factorization-based LPI prediction techniques transformed the problem of LPI identification into a recommender task, and adopted the matrix factorization model to capture unobserved LPIs. Given an LPI matrix Y and two nonnegative matrices $W \in \mathfrak{R}^{k \times n}$ and $H \in \mathfrak{R}^{k \times m}$ the problem of predicting LPIs can be formulated as the following objective function:

$$\min_{W,H} \| Y - W^T H \|_F^2 \quad \text{s.t.} \quad W \ge 0, H \ge 0 \tag{35}$$

A few LPI identification methods have been designed based on matrix factorization method.

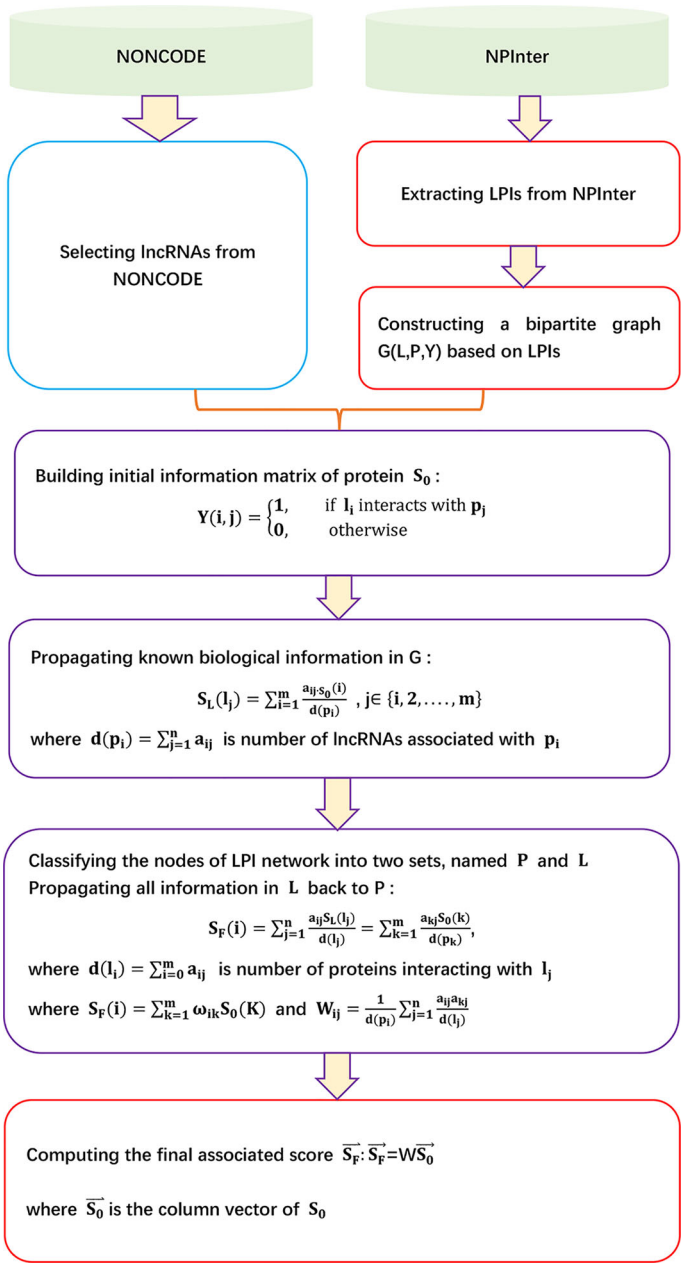

FIGURE 5 | Flowchart of lncRNA–protein bipartite network inference method.

LPGNMF

Zhang et al. (2018a) designed a graph regularized nonnegative matrix factorization-based (NMF) method to predict potential LPIs, LPGNMF. LPGNMF consists of three steps.

Step 1 Extracting LPI information based on data provided by NONCODE (Zhao et al., 2015), NPInter (Hao et al., 2016), and UniProt (Consortium et al., 2018). Obtaining 9,484 LPIs between 50 proteins and 2,190 lncRNAs after filtering and removing lncRNAs/proteins only interacting with one protein/lncRNA.

Step 2 Computing lncRNA similarity and protein similarity.

LPGNMF computes the lncRNA expression profile similarity $S^l(i, j)$:

Given the expression profiles of two lncRNAs E_1 and E_2, LPIHN calculates lncRNA expression similarity based on the Pearson correlation coefficient:

$$S^l(i,j)=\left|\frac{\mathrm{cov}(E_1,E_2)}{\sigma_{e_1}\sigma_{e_2}}\right| \tag{36}$$

where $cov(E_1, E_2)$ is the covariance of E_1 and E_2, and σ_{e_1} and σ_{e_2} are the standard deviations of E_1 and E_2, respectively.

LPGNMF computes the weight matrix based on lncRNA similarity:

$$M_{ij}^{l} = \begin{cases} 1 & i \in N(l_j) \quad \& \quad j \in N(l_i) \\ 0 & i \notin N(l_j) \quad \& \quad j \notin N(l_i) \\ & 0.5 \quad otherwise \end{cases} \tag{37}$$

Here, $N(l_i)$ and $N(l_j)$ denote the p nearest neighbors of l_i and l_j.

LPGNMF then calculates the sparse similarity matrix of lncRNAs $S^{l\star}$:

$$S_{ij}^{l\star} = M_{ij}^{l} S_{ij}^{l} \tag{38}$$

Similarly, LPGNMF calculates the sparse similarity matrix of proteins $S^{p\star}$.

Step 3 Building the following optimization model based on the graph regularized nonnegative matrix factorization method:

$$\begin{aligned} &\min_{W,H} \| Y - W^T H \|_F^2 + \lambda_p \sum_{i,j=1}^{n} \| w_i - w_j \|^2 S_{ij}^{p\star} \\ &+ \lambda_l \sum_{i,j=1}^{m} \| h_i - h_j \|^2 S_{ij}^{l\star} + \beta_1 \sum_{i,j=1}^{n} \| W(:,i) \|_1^2 \\ &+ \beta_2 \sum_{i,j=1}^{m} \| H(:,i) \|_1^2 s.t. \quad W \\ &\geq 0, H \geq 0 \end{aligned} \tag{39}$$

The details are shown in **Figure 6**.

LPI-NRLMF

Liu et al. (2017) designed a novel LPI identification model based on neighborhood regularized logistic matrix factorization, LPI-NRLMF. LPI-NRLMF can be roughly broken down into three steps.

Step 1 Extracting the lncRNA sequence, protein sequence, and LPIs based on data provided by NONCODE (Zhao et al., 2015), NPInter (Hao et al., 2016), and UniProt (Consortium et al., 2018); and obtaining 4,158 LPIs between 27 proteins and 990 lncRNAs.

Step 2 Computing lncRNA sequence similarity matrix *LSM* and protein sequence similarity matrix *PSM* based on the Smith–Waterman algorithm:

$$LSM(l_i, l_j) = \frac{sw(l_i, l_j)}{max \ (sw(l_i, l_i), sw(l_j, l_j))} \tag{40}$$

$$PSM(p_i, p_j) = \frac{sw(p_i, p_j)}{max \ (sw(p_i, p_i), sw(p_j, p_j))} \tag{41}$$

Step 3 Defining neighborhood information for lncRNAs and obtaining the adjacency matrix A of lncRNAs:

$$a_{iu} = \begin{cases} s_{iu}^{l} & if \quad l_u \in N(l_i) \\ 0 & otherwise \end{cases} \tag{42}$$

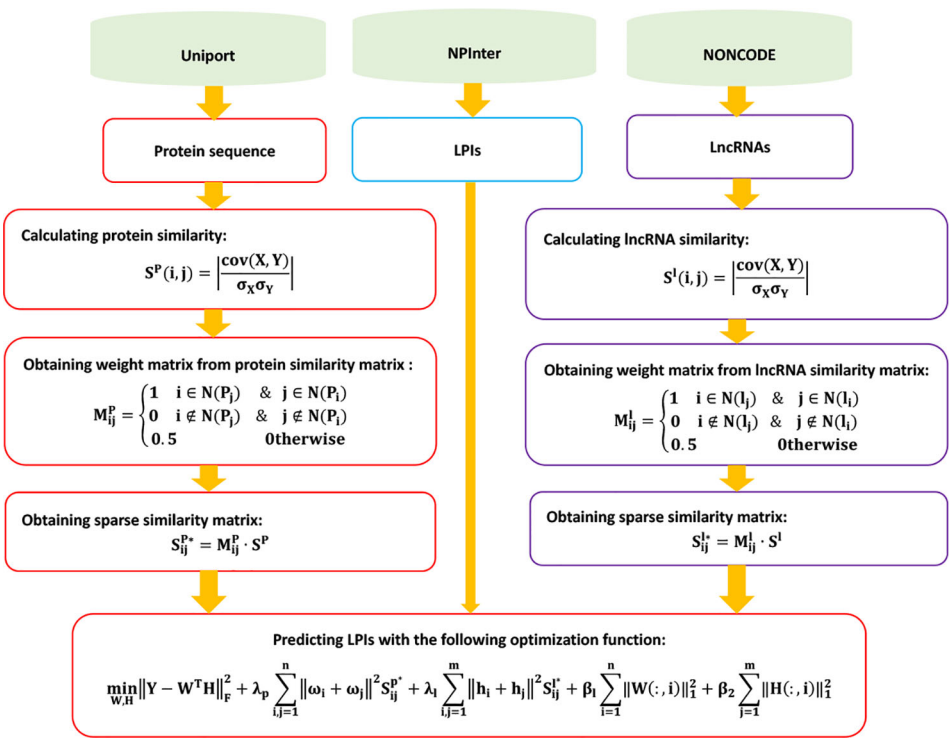

FIGURE 6 | Flowchart of LPI prediction method based on graph regularized nonnegative matrix factorization.

Similarly, LPI-NRLMF computes the adjacency matrix B of proteins.

Step 4 Computing associated scores S_N for unknown lncRNA–protein pairs based on the neighborhood regularized logistic matrix factorization model:

$$p_{ij} = \frac{exp\ (u_i v_j^T)}{1 + exp\ (u_i v_j^T)} \tag{43}$$

Here, $u_i \in \Re^{1xr}$ and $v_j \in \Re^{1xr}$ can be computed by the following neighborhood regularized logistic matrix factorization model:

$$\min_{U,V} \sum_{i=1}^{m}\sum_{j=1}^{n}(1 + cy_{ij} - y_{ij}) \ln[1 + \exp(u_i v_j^T)] - cy_{ij} u_i v_j^T + \frac{1}{2} tr[U^T(\lambda_l I + \alpha L^l)U] + \frac{1}{2} tr[V^T(\lambda_p I + \beta L^p)V] \tag{44}$$

where $L^l = (D_i^l + D_u^l) - (A + A^T)$, $D_i^l = \sum_{u=1}^{m} a_{iu}, D_u^l = \sum_{i=1}^{m} a_{iu}$. Similarly, L_P can be computed. $U \in \Re^{m\ x\ r}$ and $V \in \Re^{1\ x\ r}$ can be calculated by dividing L.

The details are shown in **Figure 7**.

IRWNRLPI

Zhao et al. (2018b) fused the random walk into LPI-NRLMF and exploited a novel LPI prediction model based on LPI-NRLMF, IRWNRLPI. IRWNRLPI is a semi-supervised learning-based model and does not require negative samples. IRWNRLPI contains the following five steps.

Step 1 Extracting the lncRNA sequence, protein sequence, and LPIs from NONCODE (Zhao et al., 2015), NPInter (Hao et al., 2016), and UniProt (Consortium et al., 2018); and obtaining 4,158 LPIs between 27 proteins and 990 lncRNAs.

Step 2 Computing the lncRNA sequence similarity matrix LS and protein sequence similarity matrix PS based on the Smith–Waterman algorithm:

$$LS(l_i, l_j) = \frac{sw(l_i, l_j)}{max\ (sw(l_i, l_i), sw(l_j, l_j))} \tag{45}$$

$$PS(p_i, p_j) = \frac{sw(p_i, p_j)}{max\ (sw(p_i, p_i), sw(p_j, p_j))} \tag{46}$$

Step 3 Building a random walk model to compute associated scores S_R for unknown lncRNA–protein pairs:

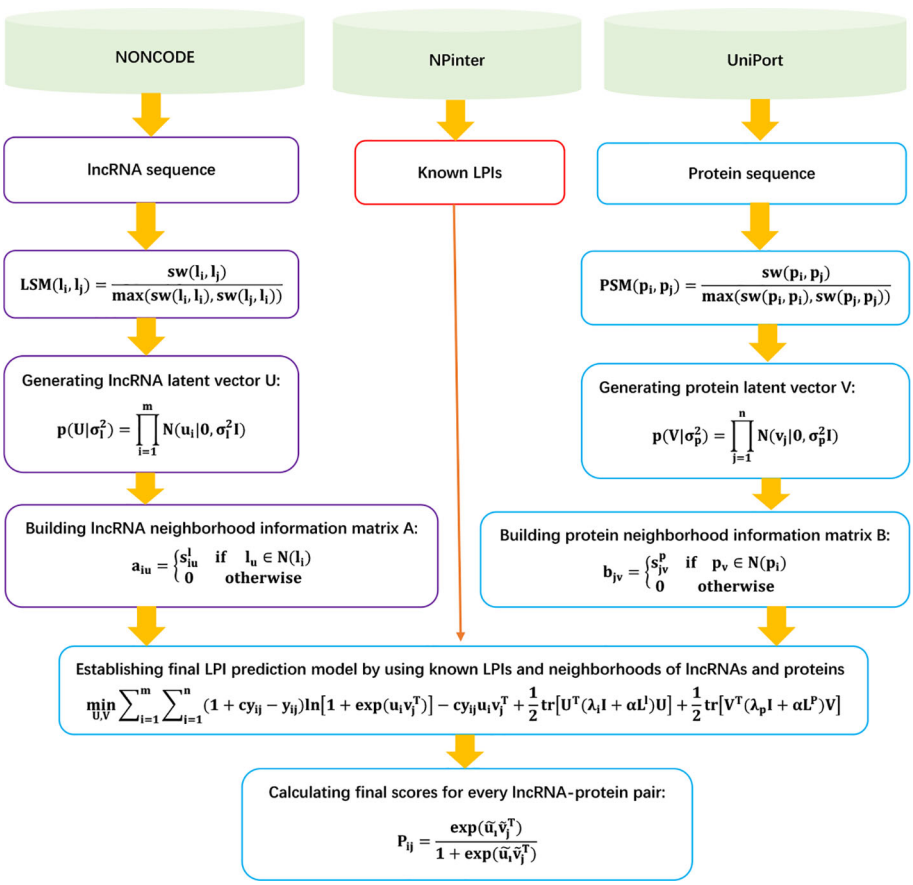

FIGURE 7 | Flowchart of LPI prediction model based on neighborhood regularized logistic matrix factorization.

$$S(t+1) = r_Q L_Q^T S(t) + p_Q(1-r_Q)X + r_U L_U^T S(t) + p_U(1-r_U)X \tag{47}$$

where r_{ij} represents the extent of association between a neighbor v_j and a protein p for a given node v_i. $L(l_{ij})_{M\ x\ M}$ is computed by $l_{ij} = r_{ij}/\sum_{j=1}^{N} r_{ij}$. IRWNRLPI divides L into two arrays of L_U and L_Q.

Step 4 Computing associated scores S_N for unknown lncRNA–protein pairs based on the neighborhood regularized logistic matrix factorization model:

$$p_{ij} = \frac{exp\ \ (u_i v_j^T)}{1 + exp\ \ (u_i v_j^T)} \tag{48}$$

$u_i \in \Re^{1\ x\ r}$ and $v_j \in \Re^{1\ x\ r}$ can be computed by the following neighborhood regularized logistic matrix factorization model:

$$\min_{U,V} \sum_{i=1}^{m}\sum_{j=1}^{n}(1 + cy_{ij} - y_{ij}) \ln[1 + \exp(u_i v_j^T)] - cy_{ij} u_i v_j^T + \frac{1}{2} tr[U^T(\lambda_l I + \alpha L^l)U] + \frac{1}{2} tr[V^T(\lambda_p I + \beta L^p)V] \tag{49}$$

where $U \in \Re^{m\ x\ r}$ and $V \in \Re^{n\ x\ r}$.

Step 5 Computing the final associated scores for unknown lncRNA–protein pairs:

$$S = \frac{S_R + S_N}{2} \tag{50}$$

The details are shown in **Figure 8**.

LPI-KTASLP

Shen et al. (2019) designed a kernel target alignment-based semi-supervised model, LPI-KTASLP, to find novel LPIs. LPI-KTASLP utilizes matrix factorization and an approximation technique. LPI-KTASLP can be roughly broken down into three steps.

Step 1 Computing lncRNA kernels and protein kernels from four levels.

Level 1 GIP kernel:

The GIP kernels between two lncRNAs and two proteins are defined as follows, respectively:

$$K_{GIP}^{lnc}(l_i, l_j) = \exp(-\gamma_{lnc} \parallel Y_{l_i} - Y_{l_j} \parallel^2) \tag{51}$$

$$K_{GIP}^{pro}(p_i, p_j) = \exp(-\gamma_{pro} \parallel Y_{p_i} - Y_{p_j} \parallel^2) \tag{52}$$

Level 2 Sequence kernel:

The sequence kernels of two lncRNAs and two proteins are defined as follows, respectively:

$$K_{SW}^{lnc}(l_i, l_k) = \frac{SW(S_{l_i}, S_{l_k})}{\sqrt{SW(S_{l_i}, S_{l_i})}\sqrt{SW(S_{l_k}, S_{l_k})}} \tag{53}$$

$$K_{SW}^{pro}(P_i, P_k) = \frac{SW(S_{p_i}, S_{p_k})}{\sqrt{SW(S_{p_i}, S_{p_i})}\sqrt{SW(S_{p_k}, S_{p_k})}} \tag{54}$$

where $SW(.,.)$ is the Smith–Waterman score, and S represents the sequence information of a lncRNA/protein.

Level 3 Sequence feature kernel:

Constructing radial basis function kernels K_{SF}^{lnc} and K_{SF}^{pro} for lncRNAs and proteins based on the conjoint triad and pseudo position-specific score matrix, respectively.

Level 4 lncRNA expression kernel:

Calculating the expression kernel of lncRNA K_{EXP}^{lnc} based on the expression profiles of lncRNAs provided by the NONCODE database (Zhao et al., 2015).

Step 2 Fusing the above kernels to generate the optimal kernel based on kernel target alignment:

$$K_{lnc}^{\star} = \sum_{a=1}^{4} w_a^{lnc} K_a^{lnc}, K_a^{lnc} \in \Re^{n\times n} \tag{55}$$

$$K_{pro}^{\star} = \sum_{a=1}^{3} w_a^{pro} K_a^{pro}, K_a^{pro} \in \Re^{m\times m} \tag{56}$$

Step 3 Constructing the following model to compute interaction probabilities for unobserved lncRNA–protein pairs based on matrix factorization, low-rank approximation, and eigen decomposition:

$$Y^{\star} = \frac{1}{1+3\delta} Y + \frac{1}{1+3\delta^2} V_{lnc}\left(D \odot \left(V_{lnc}^T F V_{pro}\right)\right) V_{pro}^T \tag{57}$$

The details are shown in **Figure 9**.

Ensemble-Based Methods

Ensemble learning methods are widely applied to LPI prediction. HLPI-Ensemble (Hu et al., 2018) and SFPEL-LPI (Zhang et al., 2018c) are two state-of-the-art ensemble-based LPI prediction methods.

HLPI-Ensemble

Hu et al. (2018) developed the HLPI-Ensemble method for human LPI identification. HLPI-Ensemble consists of two major processes: benchmark dataset construction and HLPI-Ensemble model construction.

In the first process, HLPI-Ensemble downloads lncRNA sequences, protein sequences, and LPIs from NONCODE (Zhao et al., 2015), UniProt (Consortium et al., 2018), and NPinter (Hao et al., 2016). HLPI-Ensemble then extracts 82 features of lncRNAs and 1,516 features of proteins based on Kmer, DAC, and PC-PseDNC-General.

In the second process, HLPI-Ensemble utilizes the ensemble technique and generates three ensemble learning frameworks, HLPI-SVM, HLPI-XGB, and HLPI-RF. These three frameworks are based on support vector machines (SVMs), extreme gradient boosting (XGB), and random forests (RFs), respectively. The details are shown in **Figure 10**.

SFPEL-LPI

Zhang et al. (2018c) exploited a sequence-based feature projection ensemble learning framework, SFPEL-LPI, to uncover novel LPIs. SFPEL-LPI integrated $\ell_{1,2}$-norm regularization, ensemble graph Laplacian regularization, and

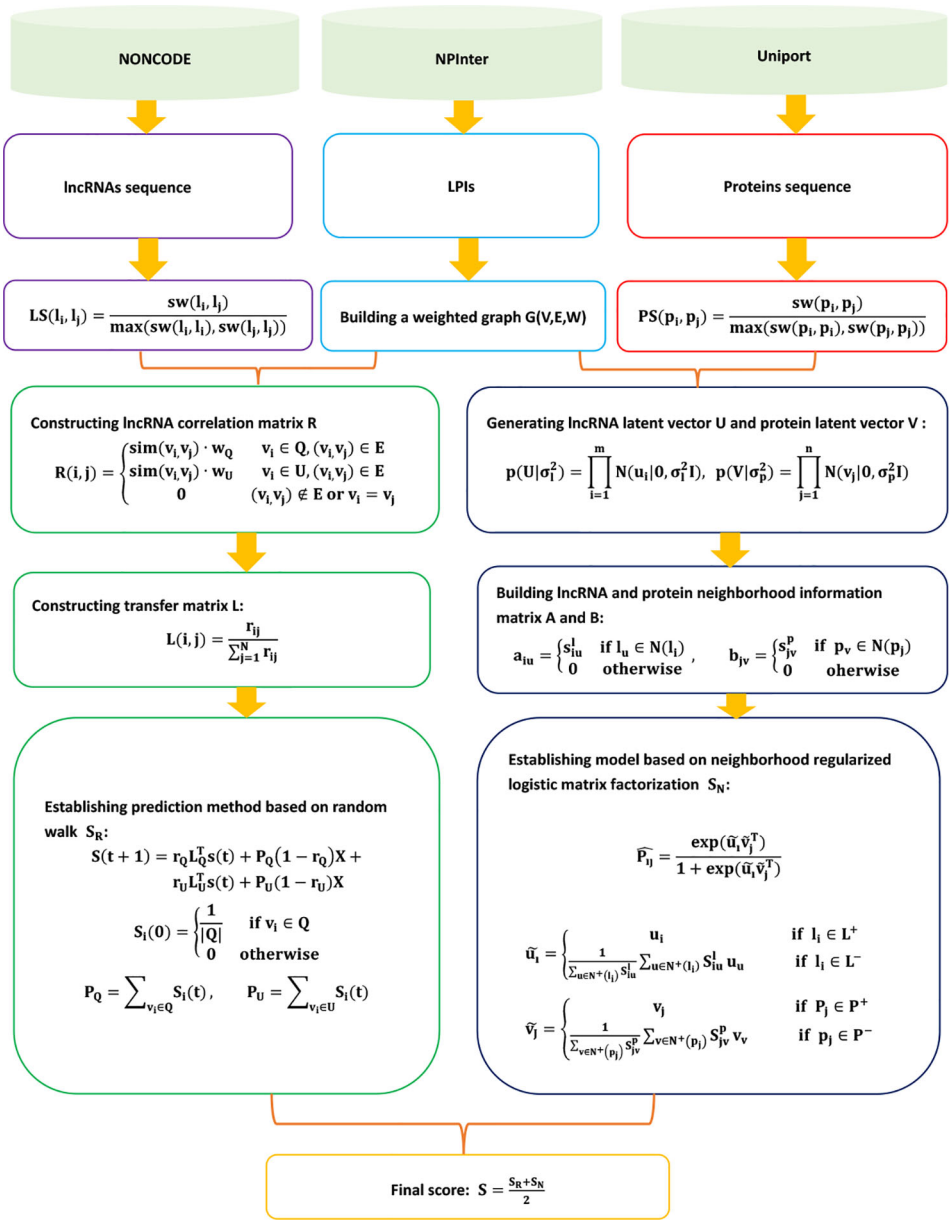

FIGURE 8 | Flowchart of LPI prediction model based on the random walk and neighborhood regularized logistic matrix factorization.

various biological information into a unified framework. It can be roughly broken down into five steps.

Step 1 Downloading LPIs, lncRNA sequences, and protein sequences from NPInter (Hao et al., 2016), NONCODE (Zhao et al., 2015), and SUMPERFAMILY (Pandurangan et al., 2018), respectively.

Step 2 Describing lncRNA and protein features based on sequence information and known LPIs.

SFPEL-LPI describes lncRNA features based on parallel correlation pseudo dinucleotide composition (PSEDNC). Given the occurrence frequency of different dinucleotides and the physicochemical properties of every dinucleotide, the PseDNC feature vector for an RNA sequence L can be represented as

$$L = [d_1, d_2, \ldots\ldots d_{16}, d_{16+1}, \ldots d_{16+\tau}] \tag{58}$$

where

$$d_k = \begin{cases} \frac{f_k}{\sum_{i=1}^{16} f_i + w \sum_{j=1}^{\tau} \theta_j} & 1 \le k \le 16 \\ \frac{w\theta_{k-16}}{\sum_{i=1}^{16} f_i + w \sum_{j=1}^{\tau} \theta_j} & 17 \le k \le 16 + \tau \end{cases} \tag{59}$$

In addition, SFPEL-LPI represents the interaction profile of an lncRNA as a row vector of the LPI matrix Y: $IP_{L_i} = Y(i,:)$.

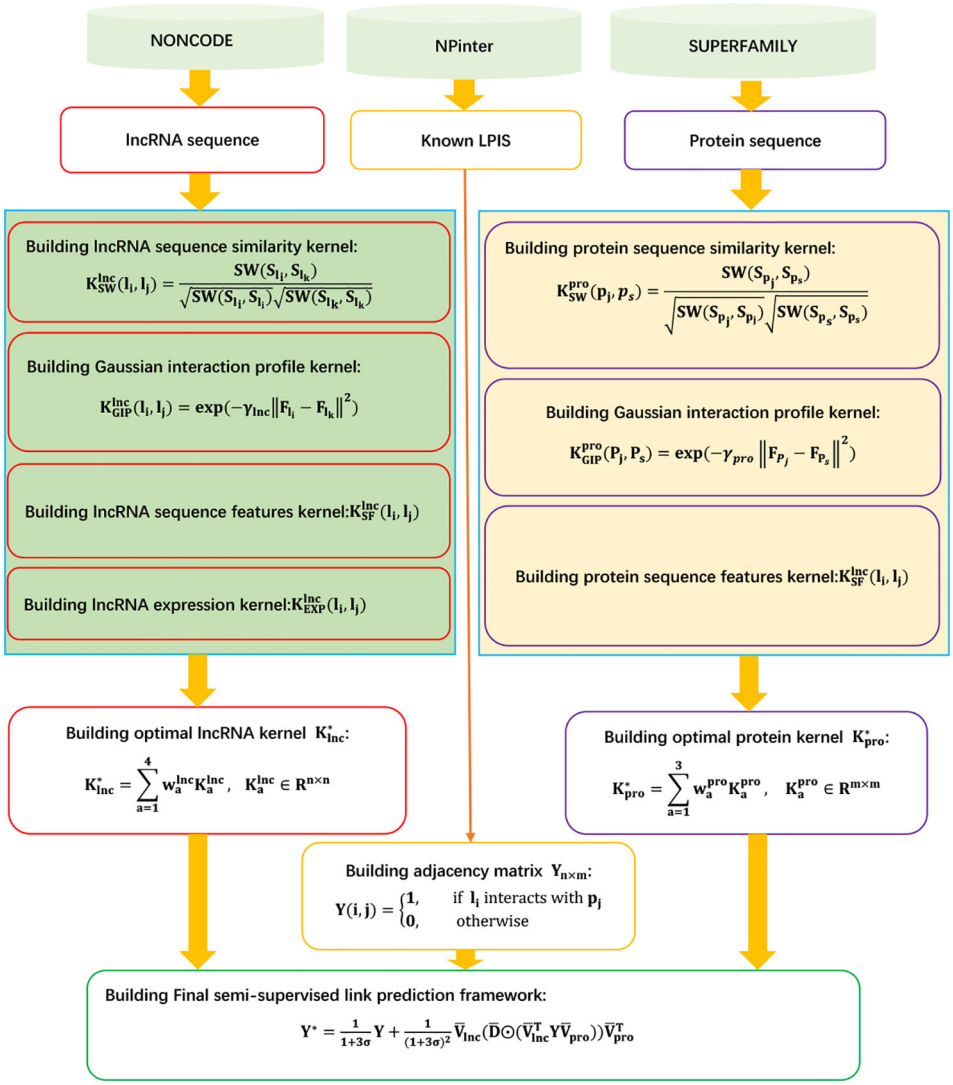

FIGURE 9 | Flowchart of kernel target alignment-based semi-supervised model for LPI prediction.

SFPEL-LPI describes protein features based on the parallel correlation pseudo amino acid composition (PseAAC):

$$P = [c_1, c_2, \ldots\ldots c_{20}, c_{20+1}, \ldots, c_{20+\tau}] \tag{60}$$

where

$$c_k = \begin{cases} \frac{f_k}{\sum_{i=1}^{20} f_i + w\sum_{j=1}^{\tau} \theta_j} & 1 \le k \le 20 \\ \frac{w\theta_{k-20}}{\sum_{i=1}^{20} f_i + w\sum_{j=1}^{\tau} \theta_j} & 20 \le k \le 20+\tau \end{cases} \tag{61}$$

Similarly, the interaction profile of a protein can be defined as a column vector of the LPI matrix Y: $IP_{p_i} = Y(:, i)$.

Therefore, a features for lncRNAs/proteins can be represented as feature matrix: $\{X_i\}_{i=1}^{a}$.

Step 4 Computing lncRNA similarity and protein similarity.

SFPEL-LPI first computes the linear neighborhood similarity of lncRNAs based on PseDNC and IP.

SFPEL-LPI then computes the Smith–Waterman subgraph similarity (SWSS) of lncRNAs:

$$SWSS(L_i, L_j) = \sum_{P_{01}\in A(L_i)} \sum_{P_{02}\in A(L_j)} \frac{SW(P_{o1}, P_{o2})}{n1 \times n2} \tag{62}$$

Similarly, the PseAAC similarity, IP similarity, and SWSS similarity of proteins can be computed.

Therefore, b types of similarities of lncRNAs/proteins can be represented as b similarity matrices $\{W_i\}_{i=1}^{b}$.

Step 5 Computing the association scores for novel lncRNAs/proteins based on Eqs. (63) and (64).

$$R_l = \sum_{i=1}^{u} \theta_{li} X_{li} G_{li}^T \tag{63}$$

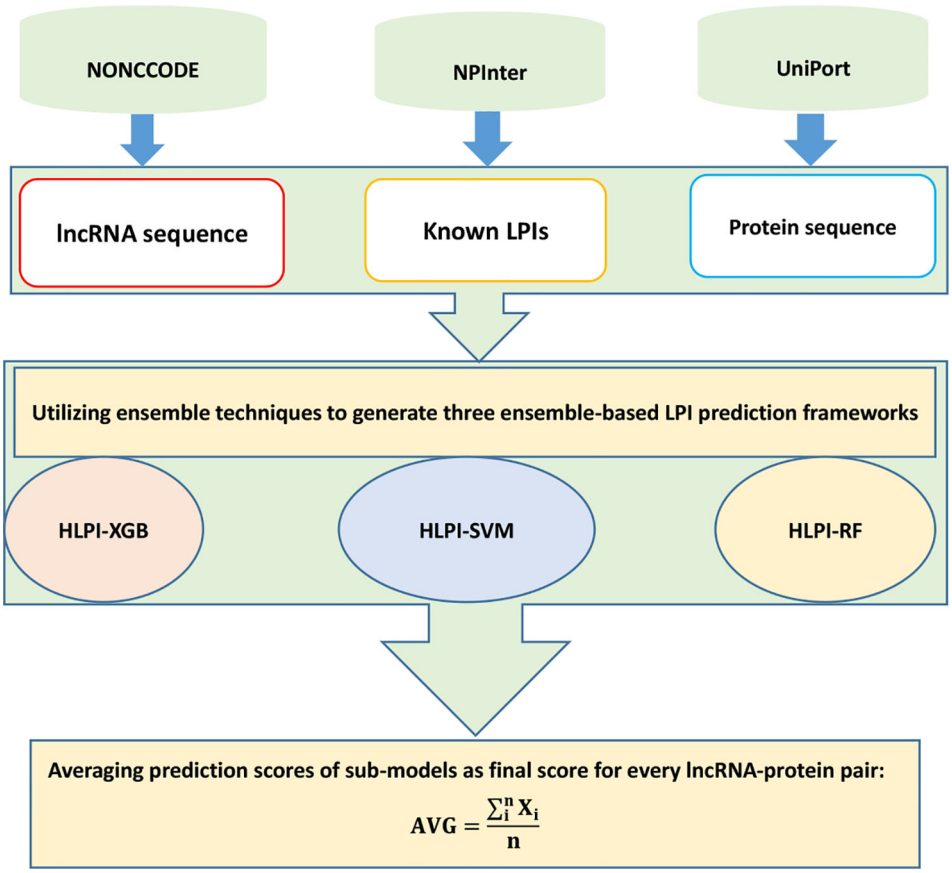

FIGURE 10 | Flowchart of ensemble-based LPI identification method.

$$R_p = \sum_{i=1}^{v} \theta_{pi} X_{pi} G_{pi}^T \tag{64}$$

G_i, R, and θ can be obtained by solving the following optimization model:

$$\begin{aligned} &\min_{G_i,R,\theta} \| R - Y \|_F^2 + \mu \sum_{i=1}^{a} \| X_i G_i^T - R \|_F^2 + \sum_{i=1}^{b} \theta_i^{\eta} tr(R^T (D_i - W_i) R) \\ &+ \lambda \sum_{i=1}^{a} \| G_i \|_{1,2}^2 \\ &s.t. \qquad G_i \geq 0, \sum_{i=1}^{b} \theta_i = 1 \end{aligned} \tag{65}$$

The details are shown in **Figure 11**.

Other Methods

There are several methods used to predict possible LPIs except for matrix factorization-based methods and ensemble learning-based methods, for example, Fisher's linear discriminant-based LPI prediction method (IncPro) (Lu et al., 2013), eigenvalue transformation-based semi-supervised model (LPI-ETSLP) (Hu et al., 2017), and kernel ridge regression model based on fast kernel learning(LPI-FKLKRR) (Shen et al., 2018).

lncPRO

Lu et al. (2013) explored a Fisher's linear discriminant-based LPI prediction method, lncPro. lncPro found new LPI through executing the following four steps.

Step 1 Downloading complexes data from the PDB database.

Step 2 Encoding sequence information into numerical feature vectors for lncRNAs and proteins based on the secondary structure, the Van der Waals' propensities, and the hydrogen-bonding propensities.

Step 3 Transforming the feature vectors to unify the dimension based on the Fourier series:

$$X_k' = \sqrt{\frac{2}{L}} \sum_{i=0}^{L} X_i \cos\left[\frac{\pi}{L}\left(n + \frac{1}{2}\right)\left(k + \frac{1}{2}\right)\right] \tag{66}$$

$$k = 0, 1, \ldots, 9$$

where L is the length of feature vector of lncRNAs/proteins.

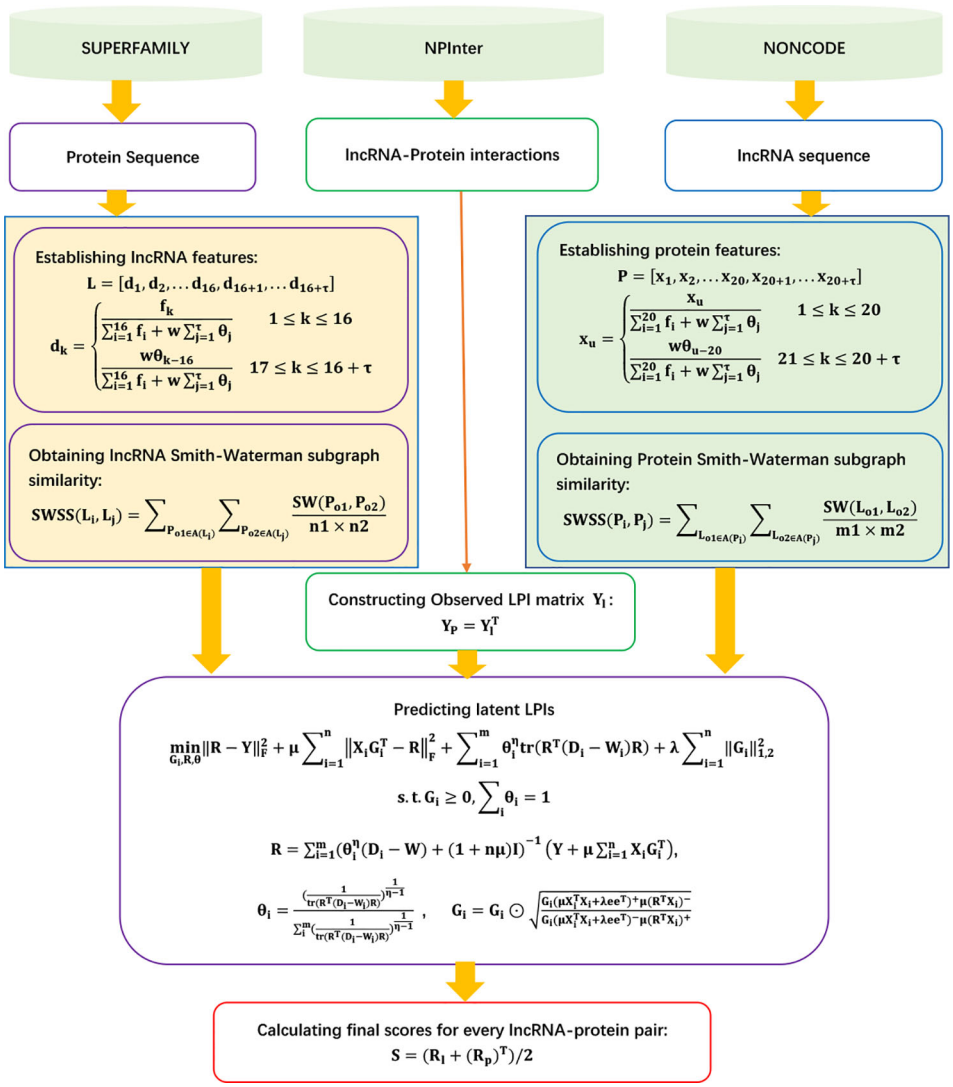

FIGURE 11 | Flowchart of LPI prediction method based on sequence feature projection ensemble learning framework.

Step 4 Calculating the final score matrix $< p|M|r >$ for the RNA feature vector r and a protein feature vector p based on Fisher's linear discriminant method:

$$< p|M|r >= M_1 p_1 r_1 + M_2 p_1 r_2 + M_3 p_2 r_1 + M_4 p_2 r_2 \quad (67)$$

LPI-ETSLP

Hu et al. (2017) presented an eigenvalue transformation-based semi-supervised model, LPI-ESTLP, to uncover the underlying LPIs. LPI-ESTLP can be broken down into three steps.

Step 1 Downloading lncRNA sequences, protein sequences, and LPIs from NONCODE (Zhao et al., 2015), UniProt (Consortium et al., 2018), and NPInter (Hao et al., 2016); and extracting 4,158 LPIs between 27 proteins and 990 lncRNAs after preprocessing.

Step 2 Computing the lncRNA sequence similarity matrix *LSM* and protein sequence similarity matrix *PSM* based on the Smith–Waterman algorithm:

$$LSM(l(i), l(j)) = \frac{sw(l(i), l(j))}{\max(sw(l(i), l(i)), sw(l(j), l(j)))} \quad (68)$$

$$PSM(p(i), p(j)) = \frac{sw(p(i), p(j))}{\max(sw(p(i), p(i)), sw(p(j), p(j)))} \quad (69)$$

Step 3 Calculating the score matrix based on the following objective function:

$$\bar{Y} = \frac{\bar{Y}_l + \bar{Y}_p}{2} \quad (70)$$

where

$$\bar{Y}_l = (\sigma L_l + I)^{-1} Y$$
$$\bar{Y}_p = (\sigma L_p + I)^{-1} Y \quad (71)$$

and $L_l = I - LSM$ and $L_p = I - PSM$ denote the Laplacian matrices of lncRNAs and proteins, respectively.

LPI-ETSLP can obtain the final scores between unobserved lncRNA–protein pairs by integrating eigenvalue transformation into Eq. 70:

$$Y = \frac{1}{2}(V_l U_l V_l^T + V_p^T U_p V_p) \quad (72)$$

where $\bar{U}_l$ is a diagonal matrix with $[\bar{U}_l]_{ii} = (1 + \sigma(1 - \lambda_{l_i}^{\alpha}))^{-1}$. $L_l = I - D_l^{-0.5} K_l D_l^{-0.5}$ and the eigen decomposition of K_l can be expressed as $K_l = V_l U_l V_l$. Similarly, $K_p = V_p U_p V_p$ and U_p can be defined.

The details are shown in **Figure 12**.

LPI-FKLKRR

Shen et al. (2018) developed an LPI prediction algorithm, LPI-FKLKRR, combining a kernel ridge regression model based on fast kernel learning. LPI-FKLKRR can be broken into six steps:

Step 1 Computing lncRNA GIP, sequence feature, sequence similarity, and lncRNA expression kernels K_{GIP}^{lnc}, K_{SW}^{lnc}, K_{SF}^{lnc}, and K_{EXP}^{lnc}.

Step 2 Computing protein GIP, sequence features, protein sequence similarity, and protein GO kernel K_{GIP}^{pro}, K_{SW}^{pro}, K_{SF}^{pro}, K_{GO}^{pro}.

Step 3 Generating the optimal lncRNA and protein kernels with fast kernel learning:

$$K_{lnc} = \sum_{a=1}^{4} w_a^{lnc} K_a^{lnc}, K_a^{lnc} \in \Re^{m \times m} \quad (73)$$

$$K_{pro} = \sum_{a=1}^{4} w_a^{pro} K_a^{pro}, K_a^{pro} \in \Re^{m \times m}$$

where w_a^{lnc} and w_a^{pro} represent each element in w_{lnc} and w_{pro}, respectively; K_a^{lnc} and K_a^{pro} denote the corresponding normalized similarity matrices in lncRNA and protein spaces, respectively.

Step 4 Constructing the optimization model to compute the optimal solution for w^{lnc} or w^{pro}:

$$\min_w w^T(A + \lambda I)w - 2b^T w$$
$$s.t. \sum_a^J w_a = 1 \quad (74)$$
$$A_{u,v} = tr(K_u^T K_v)$$

where w denotes the optimal solution w_{lnc} or w_{pro}, K_u and K_v denote two different kernel matrices, and $tr(\cdot)$ denotes the trace function.

Step 5 Computing lncRNA–protein association score matrix:

$$F^* = K_{lnc}(K_{lnc} + \lambda_\ell I)^{-1} F (K_{pro} + \lambda_p I)^{-1} K_{pro} \quad (75)$$

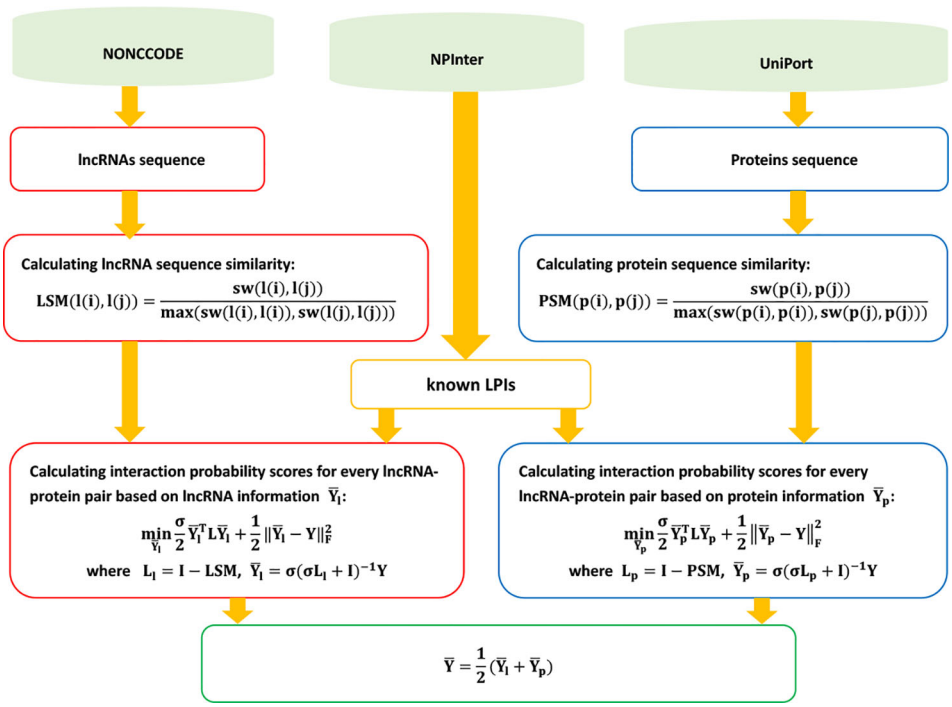

FIGURE 12 | Flowchart of eigenvalue transformation-based semi-supervised model.

Step 6 Producing the optimal F^* by adjusting the parameters λ_ℓ and λ_p.

The details are shown in **Figure 13**.

DISCUSSION

lncRNAs play important regulatory roles in diverse biological processes, such as protein modification, DNA methylation, and chromosome (Weber et al., 2018; Huang et al., 2018a; He et al., 2018b; Zhao et al., 2018c). However, the regulatory mechanism remains unknown (Esteller, 2011; Jiang et al., 2018; Agirre et al., 2019). Studies reported that identifying protein molecules binding specific lncRNAs help to probe the mechanism of lncRNAs (Lu et al., 2013; Ge et al., 2016; Chen et al., 2018). Therefore, identifying possible LPIs has an important role in understanding lncRNA-related activities (Lu et al., 2013; Pan et al., 2016; Peng et al., 2017; Zhang et al., 2018c).

However, experimental methods are expensive and time-consuming. For limited existing knowledge, computational methods become vital as a silver-bullet solution to capture LPIs on a large scale, which contributes to prioritize LPI candidates and deploys further experimental validation (Chen et al., 2018).

In this study, databases involved in LPI identification are summarized. More importantly, the components of state-of-the-art computational models for LPI prediction, such as network-based methods and machine learning-based methods, are introduced. Particularly, machine learning-based models can be broken into matrix factorization-based methods and ensemble learning-based methods. To consider the performance of LPI prediction methods, we compared nine models (IRWNRLPI, LPBNI, LPGNMF, LPI-BNPRA, LPI-ETSLP, LPIHN, LPI-NRLMF, LPLNP, and SFPEL-LPI) on leave-one-out cross-validation (LOOCV). These nine models are conducted on the datasets provided by the corresponding papers. Parameters are set as the values recommended by the corresponding studies. **Table 1** shows the comparison results based on AUC, precision, accuracy, and F1. In **Table 1**, SFPEL-LPI obtained the best performances of AUC and accuracy; LPGNMF obtained the best performances of precision and F1. The results demonstrated that SFPEL-LPI can correctly predict LPIs with a relative high proportion. LPGNMF can better identify potential LPIs when taking into account the proportion of correctly predicted LPIs and successfully predicted LPIs.

To further detect the performance of SFPEL-LPI, we compared it with four representative LPI prediction methods, LPBNI, LPI-ETSLP, LPIHN, and LPLNP, on fivefold cross-validation. The experiments were conducted on the same dataset, i.e., LPIs, lncRNA sequences, and protein sequences are from NPInter (Hao et al., 2016), NONCODE (Zhao et al., 2015), and SUMPERFAMILY (Pandurangan et al., 2018), respectively. The details are shown in **Table 2**. The results demonstrate that SFPEL-LPI obtained the best performance of AUC and can better identify possible LPIs.

In general, network-based methods have become one type of effective tool in possible LPI identification by utilizing LPI network, lncRNA similarity network, and protein similarity matrix. Although network-based methods efficiently discovered unknown LPIs and obtained promising results from the perspective of propagation (Li et al., 2015; Ge et al., 2016; Zheng et al., 2017; Zhao et al., 2018b), this type of method has some weaknesses.

1. Parts of computational methods tested their performances only on one database, which may result in biased predictions because of the sparse nature of LPI data (Li et al., 2015). More importantly, the lack of known LPIs limits the further research of LPI prediction in a larger network (Ge et al., 2016).
2. It is important to unravel potential LPIs for lncRNAs/proteins without any associated information (we represent these lncRNAs/proteins as new lncRNAs/proteins); however, most network-based models fail to capture LPI candidates (Zhang et al., 2018b).
3. Current network-based methods tend to be biased to the lncRNAs/proteins with more known associated proteins. Some lncRNAs/proteins interact with multiple proteins/lncRNAs and others interact with a few or even only one protein/lncRNA in an LPI network. The unbalanced nature of degree distributions in the LPI network may affect prediction performance. Increasing resistance based on the random walk may improve predictive accuracy for LPI prediction models (Li et al., 2015).
4. Parts of methods compute lncRNA similarities based on the expression profile and may produce incomplete coverage of the lncRNA similarity network when adding LPI datasets. This problem may be solved by increasing appropriate data including LPIs (Li et al., 2015).
5. Network-based methods can be applied to an LPI network in which there exists at least one link between two nodes. Especially for a bipartite network, network-based methods require that each node in the network has at least two linkages. However, the LPI network is usually composed of a few isolated subnetworks, and most of the existing network-based models fail to identify the LPIs between the lncRNAs in one subnetwork and the proteins in another (Ge et al., 2016).
6. Most current network-based methods utilized local network information and showed better performance; however, many previous computational biology studies showed that global network information contributes to capturing the associations between two entities, such as LPIs (Karuza et al., 2016; Meng et al., 2016; Shi et al., 2017).
7. Biology finally aims at providing personalized medicine for cancer patients, and it is a key issue to predict relevant drugs/targets for a certain disease by integrating multiple heterogeneous networks and constructing multiple-partite biological networks, such as protein–lncRNA–disease association networks and drug–protein–lncRNA–disease networks. However, current network-based methods are still not applied to this type of prediction (Yao et al., 2016; Yang et al., 2017; Bester et al., 2018; Lu et al., 2018; Ping et al., 2018; Fan et al., 2019).

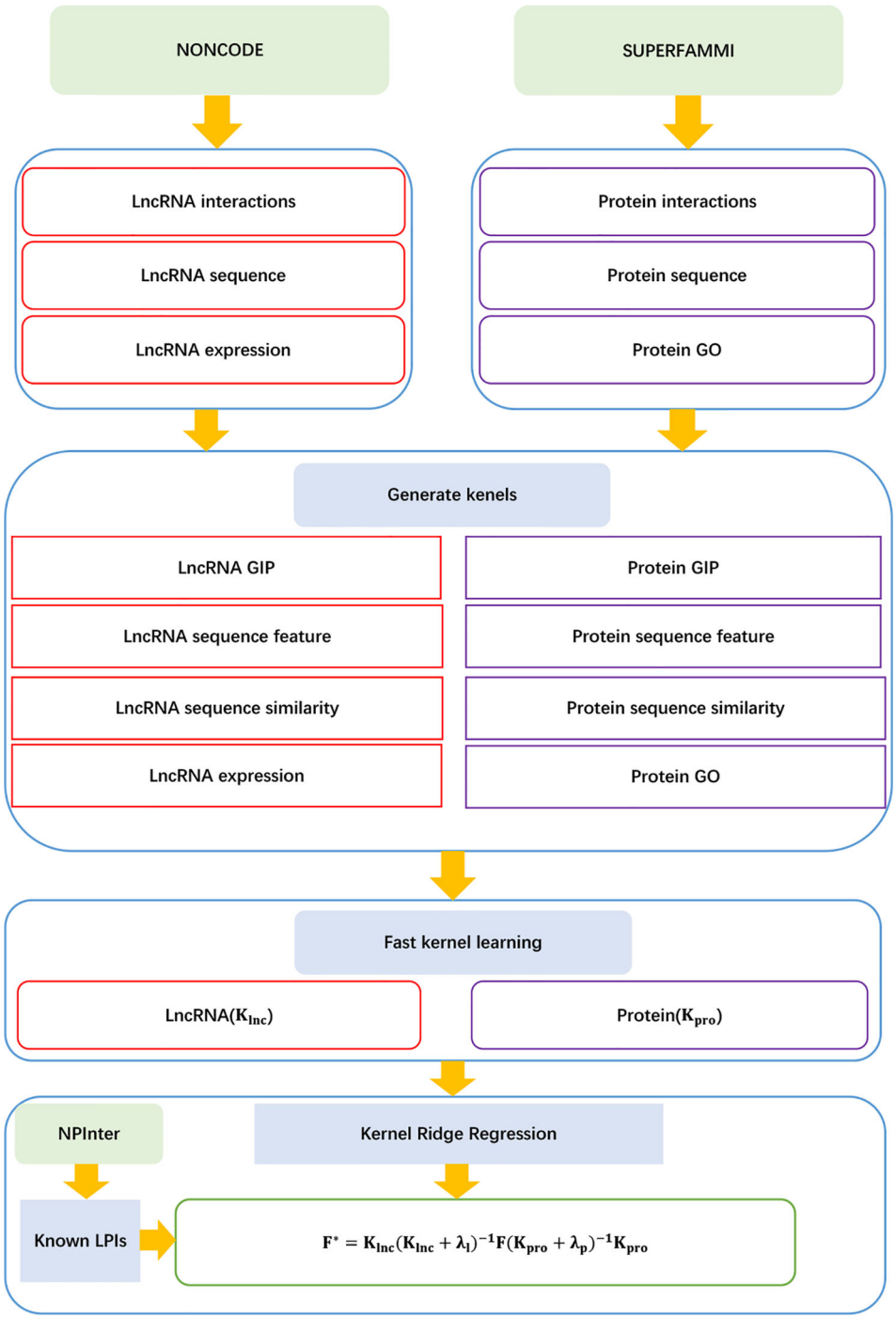

FIGURE 13 | Flowchart of LPI prediction method based on fast kernel learning with kernel ridge regression.

In summary, machine learning-based LPI prediction methods have some limitations.

1. There are no non-LPIs (negative samples) with experimental validation; therefore, most supervised learning-based LPI prediction models can only randomly select unknown lncRNA–protein pairs as negative LPIs. However, this part of randomly selected negative LPIs may contain true LPIs (positive samples) as well, which significantly influences the predictive performance (Liu et al., 2017; Zhao et al., 2018a; Zhao et al., 2018b; Zhang et al., 2018c; Shen et al., 2019). Although semi-supervised learning-based models utilized

TABLE 1 | Performance of LPI prediction methods on LOOCV.

Methods	AUC	precision	accuracy	F1
IRWNRLPI	0.9150	0.7178	0.9009	0.6516
LPBNI	0.8586	0.9681	0.9581	0.3868
LPGNMF	0.8520	**1**	0.7854	**0.6871**
LPI-BNPRA	0.8754	0.6540	0.8799	0.5564
LPI-ETSLP	0.8876	0.5932	0.8834	0.5978
LPIHN	0.8030	0.3713	0.9581	0.3868
LPI-NRLMF	0.9025	0.6129	0.8804	0.6197
LPLNP	0.9594	0.1153	0.9592	0.1621
SFPEL-LPI	**0.9735**	0.0016	**0.9731**	0.0033

These bolded texts represent that the corresponding method is the best among comparison methods.

unlabeled information to decrease the limitations of negative LPI selection, it still has the same disadvantage as classifier combination (Liu et al., 2017; Zhang et al., 2018a; Shen et al., 2019).

2. Some machine learning-based methods constructed two different classifiers, based on lncRNAs and proteins, respectively. The final results are an average of the performances of two predictive models. This type of model will produce biased results (Zhao et al., 2018b).
3. Many lncRNAs/proteins do not have known association information with any proteins/lncRNAs, and we represent them as new lncRNAs/proteins. Most current predictive models are unable to capture possible proteins/lncRNAs for new lncRNAs/proteins (Zhang et al., 2018c).
4. The proposed methods rely heavily on known LPI data; however, the current number of known LPIs is still very low. Therefore, most machine learning-based models are trained using RNA–protein interaction information instead of LPI data. This results in limited predictive performances (Liu et al., 2017; Zhao et al., 2018a). With the increase in experimentally validated LPIs, the prediction performances of models will improve (Zhao et al., 2018b).
5. The better performances of existing machine learning methods rely severely on data called features (Goodfellow et al., 2016). Current computational methods utilize various lncRNA features and protein features. However, identifying more appropriate features for a given task is still a challenge (Liu et al., 2017; Min et al., 2017). More importantly, these features are not available for all proteins or lncRNAs (Liu et al., 2017; Zhang et al., 2018c).
6. Most experimental data are provided by the NPInter database. NPInter is a relatively abundant database for lncRNA and protein data, but it only provides gene–protein interaction data corresponding to relevant lncRNAs instead of direct LPIs. Gene–protein interactions were directly applied to machine learning-based methods to find possible ncRNA–protein associations and did not discover true LPIs (Liu et al., 2017; Zhao et al., 2018a; Zhao et al., 2018b).
7. Most current computational models for LPI interaction prediction are measured based on cross-validation. Park and Marcotte (2012) used a proteochemometrics model (Wikberg and Mutulis, 2008) for drug–protein interaction prediction and observed that the paired nature of input samples has significant implications on the cross-validation of these pair-input methods. That is to say, there are significant cross-validation differences between input sample and out-of-sample interactions (Park and Marcotte, 2012). For drug–target interaction identification problems, the paired feature of input samples may produce a natural partition of test pairs, and thus the pair-input methods may obtain significantly distinct prediction accuracies for different test classes (Chen et al., 2015). The same situation applies to LPI prediction, which is still a pair-input computational identification problem.

TABLE 2 | Performance of LPI prediction methods on fivefold cross-validation.

Methods	AUC	Precision	Accuracy	F1
LPBNI	0.84177	0.2898	0.9431	0.3336
LPI-ETSLP	0.8876	**0.5932**	0.8834	**0.5978**
LPIHN	0.8531	0.4139	0.9581	0.3868
LPLNP	0.9104	0.4102	**0.9646**	0.4520
SFPEL-LPI	**0.9200**	0.4490	0.9600	0.4702

These bolded texts represent that the corresponding method is the best among comparison methods.

CONCLUSION AND FURTHER RESEARCH

There are a few LPIs and numerous unknown lncRNA–protein pairs not validated by experimental methods in the existing databases. In addition, similar lncRNAs tend to interact with similar proteins, and vice versa (Xiao et al., 2017; Zhang et al., 2018a). Therefore, LPI data have a sparse, low-rank, and unbalanced nature (Li et al., 2015; Zhang et al., 2018a; Shen et al., 2019). With the development of experimental technology, more LPIs will be confirmed, and thus the prediction accuracy of computational models will increase. In this section, we present some suggestions for further research based on the nature of LPI data.

Fusing Comprehensive LPI Datasets

Parts of computational methods tested their performances only on one database, which may result in biased predictions because of the sparse nature of LPI data (Li et al., 2015). More importantly, existing computational models utilize various biological information from proteins and lncRNAs, for example, physicochemical properties including hydrogen bonding, secondary structure, and van der Waals propensities (Belluci et al., 2011; Xiao et al., 2017). It is important to utilize diverse biological features to improve the performances of LPI prediction models. However, these features are not available for all proteins or lncRNAs, and thus computational methods cannot capture LPI candidates when information is unavailable (Zhang et al., 2018c). Therefore, exploring advanced data fusion methods to integrate more available data sources may further boost the performance of LPI identification.

Focusing on the drawbacks of current network-based LPI identification methods, future research can begin with

integrating more heterogeneous networks, such as protein–protein interaction network (Zhang et al., 2019a), lncRNA–miRNA interaction network (Zeng et al., 2016; Huang et al., 2018c; Zhao et al., 2019), lncRNA–mRNA interaction network (Alaei et al., 2019), lncRNA–disease association network (Fu et al., 2017; Wang et al., 2019), and lncRNA–miRNA–mRNA regulatory network (Chen et al., 2018; Zhang et al., 2019b). However, how to address the data conflict problems while integrating diverse LPI data from different repositories is a challenge.

Although there are not currently data conflict solutions for LPI prediction, we can find some clues by other problems in the area of bioinformatics. For example, Liu et al. (2015) set a confidence level for each DTI and gave a higher score to a DTI from a more reliable data repository. For example, the STITCH database assigns a score with a range [0, 1,000] to each DTI based on four types of different sources: model prediction, text mining, manually curated databases, and experimental validation. Particularly, Liu et al. (2015) gave DTIs from Matador and DrugBank the highest values (1,000) because DTIs from these two databases are reported by biochemical experiments and relevant studies. Lou et al. (2017) exploited another type of data fusion from a multiple-views perspective. This involved five steps: screening relevant information from different data sources; removing isolated nodes without edges in the networks; fusing various types of nodes and edges and building a heterogeneous network; constructing multiple similarity networks to boost the network heterogeneity; and excluding homologous nodes from the constructed heterogeneous networks to further reduce the possible redundancy of associated information. Inspired by these two methods, we can fuse diverse heterogeneous data to improve performance in future research. More importantly, new exploited network-based methods should be implemented on a constructed heterogeneous network rather than a single network.

Screening Credible Negative Samples

There are some known LPIs (positive samples) and abundant unknown lncRNA–protein pairs in existing LPI data resources. More importantly, there are no experimentally validated non-LPIs, and thus most supervised learning-based models have no other choice but to randomly screen negative LPIs from unlabeled lncRNA–protein pairs or even regarded all unlabeled lncRNA–protein pairs as negative samples (Liu et al., 2017; Zhao et al., 2018b). However, the randomly screened negative LPIs may contain positive LPIs as well, and thus there are severe biases in supervised learning-based techniques. Therefore, exploiting an efficient model to select high-quality negative samples is a challenging task for boosting LPI prediction accuracy.

Cheng et al. (2017) designed a FInding Reliable nEgative samples method (FIRE) to select negative RNA–protein interactions. FIRE was based on the following assumption: given a known RNA–protein interaction between an RNA *i* and a protein *j*, for an RNA *k*, the more differences between *i* and *k*, the less possibility that *k* interacts with *j*, and vice versa. FIRE screened negative RNA–protein interactions through the following steps: computing the protein similarity matrix, building a positive sample set based on known interaction information, scoring an unknown RNA–protein pair not included in positive sample set based on protein similarities, generating *m* negative samples by sorting these RNA–protein pairs *via* their scores in increasing order, and selecting the top-*m* RNA–protein pairs. Similarly, we may generate negative LPIs based on lncRNA–lncRNA similarities, protein–protein similarities, and the above assumption.

Positive-unlabeled (PU) learning (de Campos et al., 2018; Sansone et al., 2018; Yang et al., 2018) is applied to various situations. In PU learning, a supervised learning-based method is designed to learn a classification model from a positive sample set and an unlabeled dataset from an unknown class. Yang et al. (2018) designed an adaptive sampling framework with class label noise based on PU learning and introduced two new bioinformatic applications: identifying kinase-substrates and identifying transcription factor target genes. Therefore, PU learning may be one strong way to solve the problem of lacking negative LPIs.

Deep Learning

Existing computational methods have utilized different lncRNA features and protein features. For example, Bellucci et al. (2011) integrated three types of physicochemical properties, including hydrogen bonding, secondary structure, and van der Waals propensities; meanwhile, Lu et al. (2013) used six types of RNA secondary structures (besides physicochemical properties), which were provided by Bellucci et al. (2011). Therefore, designing more powerful models to integrate relevant biological features is a key issue. However, features are typically exploited by human biomedical engineers, and determining which features are more suitable for LPI prediction remains difficult. More importantly, encoding vectors that are too short may restrict the prediction accuracy of classification model. More importantly, most computational models only used sequence information but did not consider structure information (Peng et al., 2019).

Deep learning-based computational models composed of multiple processing layers require very little engineering knowledge and can efficiently extract features from raw data and construct high-level representations (Wei et al., 2018; Peng et al., 2019). These types of models have been applied to diverse analysis problems, and have obtained better performance due to the excellent power of feature learning (Jurtz et al., 2017; Min et al., 2017; Peng et al., 2019). Therefore, it is valuable and feasible to exploit deep learning-based methods to highly and effectively represent biological features for relevant entities in bioinformatics (Min et al., 2017; Zhang et al., 2018d; Peng et al., 2019; Zeng et al., 2019), such as information relevant to LPI prediction (Xiao et al., 2017; Shen et al., 2019; Zhu et al., 2019). More importantly, although deep learning demonstrated promising performance, it is not a silver bullet in LPI prediction. There still exist many challenges in LPI identification, such as the imbalanced nature of LPI data, limited LPI data, appropriate architecture selection, hyper parameter selection, and interpretation of learning results (Min et al., 2017). Therefore, solving these problems is the key to promoting deep learning-based LPI prediction models in future research.

Particularly, deep learning can be combined with PU learning and improve the performance of computational models (Bepler et al., 2018; Pati et al., 2018). For example, Bepler et al., 2018 designed the first particle-picking framework, Topaz. Topaz combined a convolutional neural network with a generalized-expectation-binomial-based objective function. The convolutional neural network was used to train classification models using only positive and unlabeled samples. Meanwhile, the generalized-expectation-binomial-based objective function was used to learn model parameters based on positive and unlabeled samples. Topaz utilized convolutional neural network classifiers to fit labeled particles (samples) and the remaining unlabeled samples based on the minibatched stochastic gradient decent method. Deep learning methods based on PU learning provide valuable insight and may be a starting point for deep learning applied to LPI prediction in future research.

Capturing LPI Candidates for New LncRNAs/Proteins

Network-based methods can be applied to an LPI network that has least one link between two nodes. For a bipartite network especially, network-based methods require that each node in the network has at least two linkages. That is to say, network-based methods cannot discover possible proteins for any lncRNA–protein pair without any known reachable paths in the LPI network (Ge et al., 2016; Zhang et al., 2018c). These lncRNAs/proteins without any interaction information are regarded as new lncRNAs/proteins (Zhang et al,. 2018c).

Given a known LPI dataset, we aim to predict (S1) LPIs between known lncRNAs and known proteins; (S2) LPIs between new lncRNAs and known proteins; (S3) LPIs between known lncRNAs and new proteins; and (S4) LPIs between new lncRNAs and new proteins. S1 has the most abundant association information, S2 and S3 have less data, and S4 has the least data. Computational models appropriate for S2 can still be applied to S3, and vice versa.

To the best of our knowledge, SFPEL-LPI provided by Zhang et al. (2018c) may be one of the rare computational methods for predicting possible LPIs for new lncRNAs/proteins. Although few computational models can be applied to the last three situations, some methods have been designed to solve similar problems in other areas in bioinformatics, and thus provide some clues for LPI prediction. For example, Shi et al. (2015) enhanced the similarity measures and introduced the concept of a "super-target" to capture the missing interactions for new drugs/targets. Furthermore, Chen et al. (2016b) exploited a miRNA–disease association prediction model based on within and between scores (WBSMDA) to uncover possible miRNA–disease associations for new miRNAs/diseases. These solutions provide clues for capturing LPI candidates for new lncRNAs/proteins.

Cross-Validation

Inspired by the evaluation methods proposed by Park and Marcotte (2012) and Chen et al. (2015), the test samples of LPIs could be categorized into four different groups: C1 is composed of the test samples sharing both lncRNAs and proteins with the training samples; C2 is composed of the test samples sharing only lncRNA with the training samples; C3 is composed of the test samples sharing only proteins with the training samples; and C4 is composed of the test samples sharing neither lncRNAs nor proteins with the training samples (Chen et al. (2015)). Therefore, it is vital to give cross-validation results under the above four independent test classes for LPI prediction.

AUTHOR CONTRIBUTIONS

LP and FL contributed equally to this work. LP, FL, XD, CP, and LZ introduced LPI data repositories and computational models. LP and FL wrote the paper. XL and YM revised original draft. LP, JY, GT, and LZ discussed the computational models and gave conclusion and further research. All authors read and approved the final manuscript.

FUNDING

This research was funded by the Natural Science Foundation of China (Grant 61803151), the Natural Science Foundation of Hunan province (Grant 2018JJ2461, 2018JJ3570), and the Project of Scientific Research Fund of Hunan Provincial Education Department (Grant 17A052).

ACKNOWLEDGMENTS

We would like to thank all authors of the cited references.

REFERENCES

Agirre, X., Meydan, C., Jiang, Y., Garate, L., Doane, A. S., Li, Z., et al. (2019). Long non-coding rnas discriminate the stages and gene regulatory states of human humoral immune response. *Nat. Commun.* 10, 821. doi: 10.1038/s41467-019-08679-z

Alaei, S., Sadeghi, B., Najafi, A., and Masoudi-Nejad, A. (2019). Lncrna and mrna integration network reconstruction reveals novel key regulators in esophageal squamous-cell carcinoma. *Genomics* 111, 76–89. doi: 10.1016/j.ygeno.2018.01.003

Bao, Z., Yang, Z., Huang, Z., Zhou, Y., Cui, Q., and Dong, D. (2018). Lncrnadisease 2.0: an updated database of long non-coding rna-associated diseases. *Nucleic Acids Res.* 47, D1034–D103D, 1037. doi: 10.1093/nar/gky905

Bellucci, M., Agostini, F., Masin, M., and Tartaglia, G. G. (2011). Predicting protein associations with long noncoding RNAs. *Nat. Methods (Nature Publishing Group)* 8 (6), 444. doi: 10.1038/nmeth.1611

Bepler, T., Morin, A., Noble, A. J., Brasch, J., Shapiro, L., and Berger, B. (2018). "Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs," in *Research in computational molecular biology: Annual International Conference, RECOM: proceedings. RECOMB (Conference: 2005-) (NIH Public Access)* (Nature Publishing Group), vol. 10812, p 245–247.

Bester, A. C., Lee, J. D., Chavez, A., Lee, Y.-R., Nachmani, D., Vora, S., et al. (2018). An integrated genome-wide crispra approach to functionalize lncrnas in drug resistance. *Cell* 173, 649–664. doi: 10.1016/j.cell.2018.03.052

Cantini, L., Kairov, U., De Reyniès, A., Barillot, E., Radvanyi, F., and Zinovyev, A. (2019). Assessing reproducibility of matrix factorization methods in

independent transcriptomes. *Bioinformatics*. doi: 10.1093/bioinformatics/btz225/5426054

Chen, X., and Yan, G.-Y. (2013). Novel human lncrna-disease association inference based on lncrna expression profiles. *Bioinformatics* 29, 2617–2624. doi: 10.1093/bioinformatics/btt426

Chen, X., Yan, C. C., Zhang, X., Zhang, X., Dai, F., Yin, J., et al. (2015). Drug–target interaction prediction: databases, web servers and computational models. *Briefings In Bioinf.* 17, 696–712. doi: 10.1093/bib/bbv066

Chen, X., Yan, C. C., Zhang, X., and You, Z.-H. (2016a). Long non-coding rnas and complex diseases: from experimental results to computational models. *Briefings In Bioinf.* 18, 558–576. doi: 10.1093/bib/bbw060

Chen, X., Yan, C. C., Zhang, X., You, Z.-H., Deng, L., Liu, Y., et al. (2016b). Wbsmda: within and between score for mirna-disease association prediction. *Sci. Rep.* 6, 21106. doi: 10.1038/srep21106

Chen, X., Sun, Y.-Z., Guan, N.-N., Qu, J., Huang, Z.-A., Zhu, Z.-X., et al. (2018). Computational models for lncrna function prediction and functional similarity calculation. *Briefings In Funct. Genomics* 18, 58–82. doi: 10.1093/bfgp/ely031

Cheng, Z., Huang, K., Wang, Y., Liu, H., Guan, J., and Zhou, S. (2017). Selecting high-quality negative samples for effectively predicting protein-rna interactions. *BMC Syst. Biol.* 11, 9. doi: 10.1186/s12918-017-0390-8

Cheng, L., Wang, P., Tian, R., Wang, S., Guo, Q., Luo, M., et al. (2018). Lncrna2target v2. 0: a comprehensive database for target genes of lncrnas in human and mouse. *Nucleic Acids Res.* 47, D140–D144. doi: 10.1093/nar/gky1051

Consortium, U., Apweiler, R., Bairoch, A., Wu, C. H., Barker, W. C., Boeckmann, B., et al. (2018). UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* 46, 2699. doi: 10.1093/nar/gky092

Cui, T., Zhang, L., Huang, Y., Yi, Y., Tan, P., Zhao, Y., et al. (2017). Mndr v2. 0: an updated resource of ncrna–disease associations in mammals. *Nucleic Acids Res.* 46, D371–D374. doi: 10.1093/nar/gkx1025

Cunningham, F., Achuthan, P., Akanni, W., Allen, J., Amode, M. R., Armean, I. M., et al. (2018). Ensembl 2019. *Nucleic Acids Res.* 47, D745–D751. doi: 10.1093/nar/gky1113

Dallner, O. S., Marinis, J. M., Lu, Y.-H., Birsoy, K., Werner, E., Fayzikhodjaeva, G., et al. (2019). Dysregulation of a long noncoding rna reduces leptin leading to a leptin-responsive form of obesity. *Nat. Med.* 1, 507–516. doi: 10.1038/s41591-019-0370-1

de Campos, L. M., Fernández-Luna, J. M., Huete, J. F., and Redondo-Expósito, L. (2018). Positive unlabeled learning for building recommender systems in a parliamentary setting. *Inf. Sci.* 433, 221–232. doi: 10.1016/j.ins.2017.12.046

Esteller, M. (2011). Non-coding rnas in human disease. *Nat. Rev. Genet.* 12, 861. doi: 10.1038/nrg3074

Fan, X.-N., Zhang, S.-W., Zhang, S.-Y., Zhu, K., and Lu, S. (2019). Prediction of lncrna-disease associations by integrating diverse heterogeneous information sources with rwr algorithm and positive pointwise mutual information. *BMC Bioinf.* 20, 87. doi: 10.1186/s12859-019-2675-y

Fu, G., Wang, J., Domeniconi, C., and Yu, G. (2017). Matrix factorization-based data fusion for the prediction of lncrna–disease associations. *Bioinformatics* 34, 1529–1537. doi: 10.1093/bioinformatics/btx794

Gao, Y., Wang, P., Wang, Y., Ma, X., Zhi, H., Zhou, D., et al. (2018). Lnc2cancer v2. 0: updated database of experimentally supported long non-coding rnas in human cancers. *Nucleic Acids Res.* 47, D1028–D102D, 1033. doi: 10.1093/nar/gky1096

Ge, M., Li, A., and Wang, M. (2016). A bipartite network-based method for prediction of long non-coding rna–protein interactions. *Genomics Proteomics Bioinf.* 14, 62–71. doi: 10.1016/j.gpb.2016.01.004

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep learning* (Cambridge, Massachusetts, USA: MIT press).

Hao, Y., Wu, W., Li, H., Yuan, J., Luo, J., Zhao, Y., et al. (2016). Npinter v3. 0: an upgraded database of noncoding rna-associated interactions. *Database* 2016, 1–9. doi: 10.1093/database/baw057

He, W., Ju, Y., Zeng, X., Liu, X., and Zou, Q. (2018a). Sc-ncdnapred: a sequence-based predictor for identifying non-coding dna in saccharomyces cerevisiae. *Front. In Microbiol.* 9, 2174. doi: 10.3389/fmicb.2018.02174

He, Y., Zuo, Q., Edwards, J., Zhao, K., Lei, J., Cai, W., et al. (2018b). Dna methylation and regulatory elements during chicken germline stem cell differentiation. *Stem Cell Rep.* 10, 1793–1806. doi: 10.1016/j.stemcr.2018.03.018

Hentze, M. W., Castello, A., Schwarzl, T., and Preiss, T. (2018). A brave new world of rna-binding proteins. *Nat. Rev. Mol. Cell Biol.* 19, 327. doi: 10.1038/nrm.2017.130

Hon, C.-C., Ramilowski, J. A., Harshbarger, J., Bertin, N., Rackham, O. J., Gough, J., et al. (2017). An atlas of human long non-coding rnas with accurate 5' ends. *Nature* 543, 199. doi: 10.1038/nature21374

Hu, H., Zhu, C., Ai, H., Zhang, L., Zhao, J., Zhao, Q., et al. (2017). Lpi-etslp: lncrna–protein interaction prediction using eigenvalue transformation-based semi-supervised link prediction. *Mol. Biosyst.* 13, 1781–1787. doi: 10.1039/c7mb00290d

Hu, H., Zhang, L., Ai, H., Zhang, H., Fan, Y., Zhao, Q., et al. (2018). Hlpi-ensemble: Prediction of human lncrna-protein interactions based on ensemble strategy. *RNA Biol.* 15, 797–806. doi: 10.1080/15476286.2018.1457935

Huang, X., Zhou, X., Hu, Q., Sun, B., Deng, M., Qi, X., et al. (2018a). Advances in esophageal cancer: a new perspective on pathogenesis associated with long non-coding rnas. *Cancer Lett.* 413, 94–101. doi: 10.1016/j.canlet.2017.10.046

Huang, Z., Shi, J., Gao, Y., Cui, C., Zhang, S., Li, J., et al. (2018b). Hmdd v3. 0: a database for experimentally supported human microrna–disease associations. *Nucleic Acids Res.* 47, D1013–D101D, 1017. doi: 10.1093/nar/gky1010

Huang, Z.-A., Huang, Y.-A., You, Z.-H., Zhu, Z., and Sun, Y. (2018c). Novel link prediction for large-scale mirna-lncrna interaction network in a bipartite graph. *BMC Med. Genomics* 11, 113. doi: 10.1186/s12920-018-0429-8

Jiang, Q., Wang, Y., Hao, Y., Juan, L., Teng, M., Zhang, X., et al. (2008). mir2disease: a manually curated database for microrna deregulation in human disease. *Nucleic Acids Res.* 37, D98–D104. doi: 10.1093/nar/gkn714

Jiang, C., Ding, N., Li, J., Jin, X., Li, L., Pan, T., et al. (2018). Landscape of the long non-coding rna transcriptome in human heart. *Brief. Bioinform.* 20 (5), 1812–1825. doi: 10.1093/bib/bby052

Jurtz, V. I., Johansen, A. R., Nielsen, M., Almagro Armenteros, J. J., Nielsen, H., Sønderby, C. K., et al. (2017). An introduction to deep learning on biological sequence data: examples and solutions. *Bioinformatics* 33, 3685–3690. doi: 10.1093/bioinformatics/btx531

Karuza, E. A., Thompson-Schill, S. L., and Bassett, D. S. (2016). Local patterns to global architectures: influences of network topology on human learning. *Trends In Cogn. Sci.* 20, 629–640. doi: 10.1016/j.tics.2016.06.003

Lan, W., Li, M., Zhao, K., Liu, J., Wu, F.-X., Pan, Y., et al. (2016). Ldap: a web server for lncrna-disease association prediction. *Bioinformatics* 33, 458–460. doi: 10.1093/bioinformatics/btw639

Li, J.-H., Liu, S., Zhou, H., Qu, L.-H., and Yang, J.-H. (2013). starbase v2. 0: decoding mirna-cerna, mirna-ncrna and protein–rna interaction networks from large-scale clip-seq data. *Nucleic Acids Res.* 42, D92–D97. doi: 10.1093/nar/gkt1248

Li, Y., Wang, C., Miao, Z., Bi, X., Wu, D., Jin, N., et al. (2014). Virbase: a resource for virus–host ncRNA-associated interactions. *Nucleic Acids Res.* (Oxford University Press) 43, D578–D582. doi: 10.1093/nar/gku903

Li, A., Ge, M., Zhang, Y., Peng, C., and Wang, M. (2015). Predicting long noncoding rna and protein interactions using heterogeneous network model. *BioMed. Res. Int.* 2015, 1–11. doi: 10.1155/2015/671950

Li, Y., Egranov, S. D., Yang, L., and Lin, C. (2019a). Molecular mechanisms of long noncoding rnas-mediated cancer metastasis. *Genes Chromosomes Cancer* 58, 200–207. doi: 10.1002/gcc.22691

Li, Y.-P., Duan, F.-F., Zhao, Y.-T., Gu, K.-L., Liao, L.-Q., Su, H.-B., et al. (2019b). A trim71 binding long noncoding rna trincr1 represses fgf/erk signaling in embryonic stem cells. *Nat. Commun.* 10, 1368. doi: 10.1002/gcc.22691

Liu, H., Sun, J., Guan, J., Zheng, J., and Zhou, S. (2015). Improving compound–protein interaction prediction by building up highly credible negative samples. *Bioinformatics* 31, i221–i229. doi: 10.1093/bioinformatics/btv256

Liu, H., Ren, G., Hu, H., Zhang, L., Ai, H., Zhang, W., et al. (2017). Lpi-nrlmf: lncrna-protein interaction prediction by neighborhood regularized logistic matrix factorization. *Oncotarget* 8, 103975. doi: 10.18632/oncotarget.21934

Liu, M., Zhang, H., Li, Y., Wang, R., Li, Y., Zhang, H., et al. (2018). Hotair, a long noncoding rna, is a marker of abnormal cell cycle regulation in lung cancer. *Cancer Sci.* 109, 2717. doi: 10.1111/cas.13745

Lu, Q., Ren, S., Lu, M., Zhang, Y., Zhu, D., Zhang, X., et al. (2013). Computational prediction of associations between long non-coding rnas and proteins. *BMC Genomics* 14, 651. doi: 10.1186/1471-2164-14-651

Lu, C., Yang, M., Luo, F., Wu, F.-X., Li, M., Pan, Y., et al. (2018). Prediction of lncrna–disease associations based on inductive matrix completion. *Bioinformatics* 34, 3357–3364. doi: 10.1093/bioinformatics/bty327

Luo, Y., Zhao, X., Zhou, J., Yang, J., Zhang, Y., Kuang, W., et al. (2017). A network integration approach for drug-target interaction prediction and computational drug repositioning from heterogeneous information. *Nat. Commun.* 8, 573. doi: 10.1038/s41467-017-00680-8

Ma, L., Li, A., Zou, D., Xu, X., Xia, L., Yu, J., et al. (2014). Lncrnawiki: harnessing community knowledge in collaborative curation of human long non-coding rnas. *Nucleic Acids Res.* 43, D187–D192. doi: 10.1093/nar/gku1167

Meng, L., Striegel, A., and Milenković, T. (2016). Local versus global biological network alignment. *Bioinformatics* 32, 3155–3164. doi: 10.1093/bioinformatics/btw348

Miao, Y.-R., Liu, W., Zhang, Q., and Guo, A.-Y. (2017). lncrnasnp2: an updated database of functional snps and mutations in human and mouse lncrnas. *Nucleic Acids Res.* 46, D276–D280. doi: 10.1093/nar/gkx1004

Min, S., Lee, B., and Yoon, S. (2017). Deep learning in bioinformatics. *Briefings In Bioinf.* 18, 851–869. doi: 10.1093/bib/bbw068

Mørk, S., Pletscher-Frankild, S., Palleja Caro, A., Gorodkin, J., and Jensen, L. J. (2013). Protein-driven inference of mirna–disease associations. *Bioinformatics* 30, 392–397. doi: 10.1093/bioinformatics/btt677

Munschauer, M., Nguyen, C. T., Sirokman, K., Hartigan, C. R., Hogstrom, L., Engreitz, J. M., et al. (2018). The norad lncrna assembles a topoisomerase complex critical for genome stability. *Nature* 561, 132. doi: 10.1038/s41586-018-0453-z

Ning, S., Yue, M., Wang, P., Liu, Y., Zhi, H., Zhang, Y., et al. (2016). Lincsnp 2.0: an updated database for linking disease-associated SNPs to human long non-coding rnas and their TFBS. *Nucleic Acids Res*. gkw945 45 (D1), D74–D78. doi: 10.1093/nar/gkw945

Nozawa, R.-S., and Gilbert, N. (2019). Rna: Nuclear glue for folding the genome. *Trends In Cell Biol*. 29 (3), 201–211. doi: 10.1016/j.tcb.2018.12.003

Pan, X., Fan, Y.-X., Yan, J., and Shen, H.-B. (2016). Ipminer: hidden ncrna-protein interaction sequential pattern mining with stacked autoencoder for accurate computational prediction. *BMC Genomics* 17, 582. doi: 10.1186/s12864-016-2931-8

Pandurangan, A. P., Stahlhacke, J., Oates, M. E., Smithers, B., and Gough, J. (2018). The superfamily 2.0 database: a significant proteome update and a new webserver. *Nucleic Acids Res.* 47, D490–D494. doi: 10.1093/nar/gky1130

Park, Y., and Marcotte, E. M. (2012). Flaws in evaluation schemes for pair-input computational predictions. *Nat. Methods* 9, 1134. doi: 10.1038/nmeth.2259

Pati, P., Andani, S., Pediaditis, M., Viana, M. P., Ruschoff, J. H., Wild, P., et al. (2018). "Deep positive-unlabeled learning for region of interest localization in breast tissue images," in *Medical Imaging 2018: Digital Pathology (International Society for Optics and Photonics)* (SPIE), vol. 10581, p 1058107.

Peng, W., Li, M., Chen, L., and Wang, L. (2017). Predicting protein functions by using unbalanced random walk algorithm on three biological networks. *IEEE/ACM Trans. Comput. Biol. Bioinf.* 14, 360–369. doi: 10.1109/TCBB.2015.2394314

Peng, C., Han, S., Zhang, H., and Li, Y. (2019). Rpiter: A hierarchical deep learning framework for ncrna–protein interaction prediction. *Int. J. Mol. Sci.* 20, 1070. doi: 10.3390/ijms20051070

Ping, P., Wang, L., Kuang, L., Ye, S., Iqbal, M. F. B., and Pei, T. (2018). A novel method for lncrna-disease association prediction based on an lncrna-disease association network. *IEEE/ACM Trans. Comput. Biol. Bioinf.* 16, 688–693. doi: 10.1039/c9mo00092e

Pyfrom, S. C., Luo, H., and Payton, J. E. (2019). Plaidoh: a novel method for functional prediction of long non-coding rnas identifies cancer-specific lncrna activities. *BMC Genomics* 20, 137. doi: 10.1186/s12864-019-5497-4

Qian, X., Zhao, J., Yeung, P. Y., Zhang, Q. C., and Kwok, C. K. (2018). Revealing lncrna structures and interactions by sequencing-based approaches. *Trends In Biochem. Sci.* 44 (1), 33–52. doi: 10.1016/j.tibs.2018.09.012

Quek, X. C., Thomson, D. W., Maag, J. L., Bartonicek, N., Signal, B., Clark, M. B., et al. (2014). lncrnadb v2. 0: expanding the reference database for functional long noncoding rnas. *Nucleic Acids Res.* 43, D168–D173. doi: 10.1093/nar/gku988

Rajput, B., Pruitt, K. D., and Murphy, T. D. (2018). Refseq curation and annotation of stop codon recoding in vertebrates. *Nucleic Acids Res.* 47, 594–606. doi: 10.1093/nar/gky1234

Rinn, J. L., and Chang, H. Y. (2012). Genome regulation by long noncoding rnas. *Annu. Rev. Biochem.* 81, 145–166. doi: 10.1146/annurev-biochem-051410-092902

Ruepp, A., Kowarsch, A., and Theis, F. (2012). "Phenomir: micrornas in human diseases and biological processes," in *Next-Generation MicroRNA Expression Profiling Technology* (Totowa, NJ, USA: Humana Press), p 249–260.

Sanchez Calle, A., Kawamura, Y., Yamamoto, Y., Takeshita, F., and Ochiya, T. (2018). Emerging roles of long non-coding rna in cancer. *Cancer Sci.* 109, 2093–2100. doi: 10.1111/cas.13642

Sansone, E., De Natale, F. G., and Zhou, Z.-H. (2018). Efficient training for positive unlabeled learning. *IEEE Trans. Pattern Anal. Mach. Intell*. 41 (1), 2584–2598. doi: 10.1109/TPAMI.2018.2860995

Shen, C., Ding, Y., Tang, J., and Guo, F. (2018). Multivariate information fusion with fast kernel learning to kernel ridge regression in predicting lncrna-protein interactions. *Front. In Genet.* 9, 716. doi: 10.3389/fgene.2018.00716

Shen, C., Ding, Y., Tang, J., Jiang, L., and Guo, F. (2019). Lpi-ktaslp: Prediction of lncrna-protein interaction by semi-supervised link learning with multivariate information. *IEEE Access* 7, 13486–13496. doi: 10.1109/ACCESS.2019.2894225

Shi, J.-Y., Yiu, S.-M., Li, Y., Leung, H. C., and Chin, F. Y. (2015). Predicting drug–target interaction for new drugs using enhanced similarity measures and super-target clustering. *Methods* 83, 98–104. doi: 10.1016/j.ymeth.2015.04.036

Shi, C., Li, Y., Zhang, J., Sun, Y., and Philip, S. Y. (2017). A survey of heterogeneous information network analysis. *IEEE Trans. Knowl. Data Eng.* 29, 17–37. doi: 10.1109/TKDE.2016.2598561

Shi, J.-Y., Zhang, A.-Q., Zhang, S.-W., Mao, K.-T., and Yiu, S.-M. (2018). A unified solution for different scenarios of predicting drug-target interactions *via* triple matrix factorization. *BMC Syst. Biol.* 12, 136. doi: 10.1186/s12918-018-0663-x

Szklarczyk, D., Morris, J. H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., et al. (2016). The string database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res*. gkw937 45 (D1), D362–D368. doi: 10.1093/nar/gkw937

Volders, P.-J., Anckaert, J., Verheggen, K., Nuytens, J., Martens, L., Mestdagh, P., et al. (2018). Lncipedia 5: towards a reference set of human long non-coding rnas. *Nucleic Acids Res.* 47, D135–D139. doi: 10.1093/nar/gky1031

Wang, Y., Chen, L., Chen, B., Li, X., Kang, J., Fan, K., et al. (2013). Mammalian ncrna-disease repository: a global view of ncrna-mediated disease network. *Cell Death Dis.* 4, e765. doi: 10.1038/cddis.2013.292

Wang, J., Cao, Y., Zhang, H., Wang, T., Tian, Q., Lu, X., et al. (2016a). Nsdna: a manually curated database of experimentally supported ncrnas associated with nervous system diseases. *Nucleic Acids Res.* 45, D902–D907. doi: 10.1093/nar/gkw1038

Wang, J., Ma, R., Ma, W., Chen, J., Yang, J., Xi, Y., et al. (2016b). Lncdisease: a sequence based bioinformatics tool for predicting lncrna-disease associations. *Nucleic Acids Res.* 44, e90–e90. doi: 10.1093/nar/gkw093

Wang, Y., Yu, G., Wang, J., Fu, G., Guo, M., and Domeniconi, C. (2019). Weighted matrix factorization on multi-relational data for lncrna-disease association prediction. *Methods*. doi: 10.1016/j.ymeth.2019.06.015

Weber, A., Schwarz, S. C., Tost, J., Trümbach, D., Winter, P., Busato, F., et al. (2018). Epigenome-wide dna methylation profiling in progressive supranuclear palsy reveals major changes at dlx1. *Nat. Commun.* 9, 2929. doi: 10.1038/s41467-018-05325-y

Wei, L., Ding, Y., Su, R., Tang, J., and Zou, Q. (2018). Prediction of human protein subcellular localization using deep learning. *J. Parallel Distrib. Comput.* 117, 212–217. doi: 10.1016/j.jpdc.2017.08.009

Wikberg, J. E., and Mutulis, F. (2008). Targeting melanocortin receptors: an approach to treat weight disorders and sexual dysfunction. *Nat. Rev. Drug Discovery* 7, 307. doi: 10.1038/nrd2331

Xiao, Y., Zhang, J., and Deng, L. (2017). Prediction of lncrna-protein interactions using hetesim scores based on heterogeneous networks. *Sci. Rep.* 7, 3664. doi: 10.1038/s41598-017-03986-1

Xie, B., Ding, Q., Han, H., and Wu, D. (2013). mircancer: a microrna–cancer association database constructed by text mining on literature. *Bioinformatics* 29, 638–644. doi: 10.1093/bioinformatics/btt014

Xie, G., Wu, C., Sun, Y., Fan, Z., and Liu, J. (2019). Lpi-ibnra: Long non-coding rna-protein interaction prediction based on improved bipartite network recommender algorithm. *Front. In Genet.* 10, 343. doi: 10.3389/fgene.2019.00343

Yang, Z., Wu, L., Wang, A., Tang, W., Zhao, Y., Zhao, H., et al. (2016). dbdemc 2.0: updated database of differentially expressed mirnas in human cancers. *Nucleic Acids Res.* 45, D812–D818. doi: 10.1093/nar/gkw1079

Yang, H., Shang, D., Xu, Y., Zhang, C., Feng, L., Sun, Z., et al. (2017). The lncrna connectivity map: using lncrna signatures to connect small molecules, lncrnas, and diseases. *Sci. Rep.* 7, 6655. doi: 10.1038/s41598-017-06897-3

Yang, P., Ormerod, J. T., Liu, W., Ma, C., Zomaya, A. Y., and Yang, J. Y. (2018). Adasampling for positive-unlabeled and label noise learning with bioinformatics applications. *IEEE Trans. cybernetics*, 49, 1–12. doi: 10.1109/TCYB.2018.2816984

Yao, B., Ma, F., Su, J., Wang, X., Zhao, X., and Yao, M. (2016). "Scale-free multiple-partite models towards information networks," in *2016 IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC)* (Hobeken, NJ, USA: IEEE press), p 549–554.

Yi, Y., Zhao, Y., Li, C., Zhang, L., Huang, H., Li, Y., et al. (2016). Raid v2. 0: an updated resource of rna-associated interactions across organisms. *Nucleic Acids Res.* 45, D115–D118. doi: 10.1093/nar/gkw1052

You, Z.-H., Huang, Z.-A., Zhu, Z., Yan, G.-Y., Li, Z.-W., Wen, Z., et al. (2017). Pbmda: a novel and effective path-based computational model for mirna-disease association prediction. *PloS Comput. Biol.* 13, e1005455. doi: 10.1371/journal.pcbi.1005455

Zeng, X., Zhang, X., and Zou, Q. (2016). Integrative approaches for predicting microrna function and prioritizing disease-related microrna using biological interaction networks. *Briefings In Bioinf.* 17, 193–203. doi: 10.1093/bib/bbv033

Zeng, X., Zhong, Y., Lin, W., and Zou, Q. (2019). Predicting disease-associated circular rnas using deep forests combined with positive-unlabeled learning methods. *Briefings In Bioinf*. doi: 10.1093/bib/bbz080

Zhang, T., Wang, M., Xi, J., and Li, A. (2018a). Lpgnmf: Predicting long non-coding rna and protein interaction using graph regularized nonnegative matrix factorization. *IEEE/ACM Trans. Comput. Biol. Bioinf*. doi: 10.1109/TCBB.2018.2861009

Zhang, W., Qu, Q., Zhang, Y., and Wang, W. (2018b). The linear neighborhood propagation method for predicting long non-coding rna–protein interactions. *Neurocomputing* 273, 526–534. doi: 10.1016/j.jpdc.2017.08.009

Zhang, W., Yue, X., Tang, G., Wu, W., Huang, F., and Zhang, X. (2018c). Sfpel-lpi: Sequence-based feature projection ensemble learning for predicting lncrna-protein interactions. *PloS Comput. Biol.* 14, e1006616. doi: 10.1371/journal.pcbi.1006616

Zhang, Z., Zhao, Y., Liao, X., Shi, W., Li, K., Zou, Q., et al. (2018d). Deep learning in omics: a survey and guideline. *Briefings In Funct. Genomics* 18, 41–57. doi: 10.1093/bfgp/ely030

Zhang, L., Yu, G., Xia, D., and Wang, J. (2019a). Protein-protein interactions prediction based on ensemble deep neural networks. *Neurocomputing* 324, 10–19. doi: 10.1016/j.compbiolchem.2019.107147

Zhang, W., Li, Z., Guo, W., Yang, W., and Huang, F. (2019b). A fast linear neighborhood similarity-based network link inference method to predict microrna-disease associations. *IEEE/ACM Trans. Comput. Biol. Bioinf*. doi: 10.1109/TCBB.2019.2931546

Zhao, Y., Li, H., Fang, S., Kang, Y., Wu, W., Hao, Y., et al. (2015). Noncode 2016: an informative and valuable data source of long non-coding rnas. *Nucleic Acids Res.* 44, D203–D208. doi: 10.1093/nar/gkv1252

Zhao, Q., Yu, H., Ming, Z., Hu, H., Ren, G., and Liu, H. (2018a). The bipartite network projection-recommended algorithm for predicting long non-coding rna-protein interactions. *Mol. Ther.-Nucleic Acids* 13, 464–471. doi: 10.1016/j.omtn.2018.09.020

Zhao, Q., Zhang, Y., Hu, H., Ren, G., Zhang, W., and Liu, H. (2018b). Irwnrlpi: integrating random walk and neighborhood regularized logistic matrix factorization for lncrna-protein interaction prediction. *Front. In Genet.* 9, 239. doi: 10.3389/fgene.2018.00239

Zhao, X.-Y., Xiong, X., Liu, T., Mi, L., Peng, X., Rui, C., et al. (2018c). Long noncoding rna licensing of obesity-linked hepatic lipogenesis and nafld pathogenesis. *Nat. Commun.* 9, 2986. doi: 10.1038/s41467-018-05383-2

Zhao, X., Tang, D.-Y., Zuo, X., Zhang, T.-D., and Wang, C. (2019). Identification of lncrna–mirna–mrna regulatory network associated with epithelial ovarian cancer cisplatin-resistant. *J. Cell. Physiol.* 234 (11), 19886–19894. doi: 10.1002/jcp.28587

Zheng, X., Wang, Y., Tian, K., Zhou, J., Guan, J., Luo, L., et al. (2017). Fusing multiple protein-protein similarity networks to effectively predict lncrna-protein interactions. *BMC Bioinf.* 18, 420. doi: 10.1186/s12859-017-1819-1

Zhou, K.-R., Liu, S., Sun, W.-J., Zheng, L.-L., Zhou, H., Yang, J.-H., et al. (2016). Chipbase v2. 0: decoding transcriptional regulatory networks of non-coding rnas and protein-coding genes from chip-seq data. *Nucleic Acids Res.* gkw965 45 (D1), D43–D50. doi: 10.1093/nar/gkw965

Zhu, Y., Xu, G., Yang, Y. T., Xu, Z., Chen, X., Shi, B., et al. (2018). Postar2: deciphering the post-transcriptional regulatory logics. *Nucleic Acids Res.* 47, D203–D211. doi: 10.1093/nar/gky830

Zhu, R., Li, G., Liu, J.-X., Dai, L.-Y., and Guo, Y. (2019). Accbn: ant-colony-clustering-based bipartite network method for predicting long non-coding rna–protein interactions. *BMC Bioinf.* 20, 16. doi: 10.1186/s12859-018-2586-3

Conflict of Interest: Authors GT and JY were employed by the company Geneis (Beijing) Co. Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.