SPMLMI: Predicting lncRNA–miRNA interactions in humans using a structural perturbation method

Mingmin Xu [Equal first author, 1] , **Yuanyuan Chen** [Equal first author, 2] , **Wei Lu** [3] , **Lingpeng Kong** [1] , **Jingya Fang** [1] , **Zutan Li** [1] , **Liangyun Zhang** [Corresp., 2] , **Cong Pian** [Corresp. 2]

[1] College of Agriculture, Nanjing Agricultural University, Nanjing, Jiangsu, China

[2] College of Sciences, Nanjing Agricultural University, Nanjing, Jiangsu, China

[3] College of Life Sciences, Nanjing Agricultural University, Nanjing, Jiangsu, China

Corresponding Authors: Liangyun Zhang, Cong Pian
Email address: zlyun@njau.edu.cn, piancong@njau.edu.cn

Long non-coding RNA (lncRNA)–microRNA (miRNA) interactions are quickly emerging as important mechanisms underlying the functions of non-coding RNAs. Accordingly, predicting lncRNA–miRNA interactions provides an important basis for understanding the mechanisms of action of ncRNAs. However, the accuracy of the established prediction methods is still limited. In this study, we used structural consistency to measure the predictability of interactive links based on a bilayer network by integrating information for known lncRNA–miRNA interactions, an lncRNA similarity network, and an miRNA similarity network. In particular, by using the structural perturbation method, we proposed a framework called SPMLMI to predict potential lncRNA–miRNA interactions based on the bilayer network. We found that the structural consistency of the bilayer network was higher than that of any single network, supporting the utility of bilayer network construction for the prediction of lncRNA–miRNA interactions. Applying SPMLMI to three real datasets, we obtained areas under the curves of 0.9512 ± 0.0034, 0.8767 ± 0.0033, and 0.8653 ± 0.0021 based on 5-fold cross-validation, suggesting good model performance. In addition, the generalizability of SPMLMI was better than that of the previously established methods. Case studies of two lncRNAs (i.e., SNHG14 and MALAT1) further demonstrated the feasibility and effectiveness of the method. Therefore, SPMLMI is a feasible approach to identify novel lncRNA–miRNA interactions underlying complex biological processes.

1 # SPMLMI: Predicting lncRNA–miRNA interactions in

2 # humans using a structural perturbation method

5 Mingmin Xu[1,†], Yuanyuan Chen[2,†], Wei Lu[3], Lingpeng Kong[1], Jingya Fang[1], Zutan Li[1],

6 Liangyun Zhang[2,*] and Cong Pian[2,*]

8 [1] College of Agriculture, Nanjing Agricultural University, Nanjing, Jiangsu, China

9 [2] College of Sciences, Nanjing Agricultural University, Nanjing, Jiangsu, China

10 [3] College of Life Sciences, Nanjing Agricultural University, Nanjing, Jiangsu, China

12 [†] These two authors should be regarded as joint first authors

13 [*] Corresponding Author:

14 Liangyun Zhang[2] and Cong Pian[2]

15 Weigang No.1, Nanjing, Jiangsu, China

16 Email address: zlyun@njau.edu.cn; piancong@njau.edu.cn

Abstract

19 Long non-coding RNA (lncRNA)–microRNA (miRNA) interactions are quickly emerging as

20 important mechanisms underlying the functions of non-coding RNAs. Accordingly, predicting

21 lncRNA–miRNA interactions provides an important basis for understanding the mechanisms of

22 action of ncRNAs. However, the accuracy of the established prediction methods is still limited.

23 In this study, we used structural consistency to measure the predictability of interactive links

24 based on a bilayer network by integrating information for known lncRNA–miRNA interactions,

25 an lncRNA similarity network, and an miRNA similarity network. In particular, by using the

26 structural perturbation method, we proposed a framework called SPMLMI to predict potential

27 lncRNA–miRNA interactions based on the bilayer network. We found that the structural

28 consistency of the bilayer network was higher than that of any single network, supporting the

29 utility of bilayer network construction for the prediction of lncRNA–miRNA interactions.

30 Applying SPMLMI to three real datasets, we obtained areas under the curves of 0.9512 ± 0.0034,

31 0.8767 ± 0.0033, and 0.8653 ± 0.0021 based on 5-fold cross-validation, suggesting good model

32 performance. In addition, the generalizability of SPMLMI was better than that of the previously

33 established methods. Case studies of two lncRNAs (i.e., SNHG14 and MALAT1) further

34 demonstrated the feasibility and effectiveness of the method. Therefore, SPMLMI is a feasible

35 approach to identify novel lncRNA–miRNA interactions underlying complex biological

36 processes.

37

Introduction

39   For a long time, our framework for understanding the nature of genetic programming in

40 complex organisms was biased by the assumption that protein-coding genes are the majority of

41 the bearers of genetic information. Researchers have recently abandoned their notions about the

42 regulation of complex cellular mechanisms, shifting their focus to previously unexplored RNAs.

43 We now know that RNAs are functionally versatile molecules directly involved in the regulation

44 of many cellular processes, such as epigenetic control, gene transcription, translation, RNA

45 transformation, chromosome organization, and genome defense as well as cell proliferation and

46 developmental programs (Guil & Esteller 2015; Morris & Mattick 2014). In most instances, the

47 key regulatory factors are non-coding RNAs (ncRNAs) (Huntzinger & Izaurralde 2011; Rinn &

48 Chang 2012; Sabin et al. 2013), which account for the vast majority of mammalian transcripts

49 and range in length from 22 nucleotides to hundreds of kilobases. Although there are established

50 principles that define classes of ncRNAs (e.g., tRNAs and miRNAs), it could be suggested that

51 each ncRNA has a unique function (Cech & Steitz 2014).

52   The first miRNA (microRNA) was reported in 1993 as a short RNA in *Caenorhabditis*

53 *elegans* able to control the timing of developmental transitions by base pairing to partially

54 complementary sequences in the 3′-UTR of its target mRNA (Lee et al. 1993; Wightman et al.

55 1993). Today, miRNAs are well-known gene expression repressors (Béthune et al. 2012; Braun

56 et al. 2013; Huntzinger & Izaurralde 2011), and are widely found in human and plant genomes

57 (Krol et al. 2010; Voinnet 2009). LncRNAs (long non-coding RNAs) are the most abundant

58 ncRNAs; tens of thousands of human lncRNAs have been identified. However, the roles of most

59 lncRNAs in cellular processes have yet to be determined, and we anticipate that many new

60 functions will be discovered (Quinn & Chang 2016; Chen et al. 2019a; Chen et al. 2017).

61 Previous studies have concluded that lncRNAs could regulate gene expression patterns by

62 biomolecular interactions, including lncRNA–protein, lncRNA–mRNA, and lncRNA–ncRNA

63 interactions (Li et al. 2014). Yoon et al. (2014) reviewed functional interactions among miRNAs

64 and lncRNAs with important roles in gene expression programs. We believe that detailed

65 analyses of the interactivity between lncRNAs and miRNAs will enrich our understanding of the

66 functions of cellular networks.

67   RNA–RNA interactions are rapidly emerging as one of the most important functional

68 mechanisms involving ncRNAs (Lai & Meyer 2016). Typically, the discovery of molecular

69 interactions in biological network requires extensive experimental work (Clauset et al. 2008).

70   Biological experiments are highly reliable for studies of biomolecular interactions; however, they
71   are difficult to carry out on a large scale owing to the high costs and long execution times. An
72   alternative to studying interactions between each pair of molecules in a huge and complex
73   biological network is a predictive approach, in which prior interaction information is used to
74   identify molecules that are most likely to be related, providing a basis for further experimental
75   studies. In particular, computational prediction methods are useful for prioritizing candidate
76   lncRNA–miRNA interactions on a large scale.
77   Most existing RNA–RNA interaction prediction methods are based on sequence information,
78   including sequence conservation, seed region matching, site accessibility, and minimum free
79   energy, and predict matching relationships between miRNAs, small RNAs, and small nucleolar
80   RNAs with their targets (mRNAs and rRNAs) (Lai & Meyer 2016; Quinn & Chang 2016).
81   Different from mRNAs and rRNAs, lncRNAs have substantial sequence variation and relatively
82   low conservation, and they are involved in relatively complex molecular interactions.
83   Accordingly, we cannot predict interactions between lncRNAs and miRNAs by sequence
84   information alone.
85   As a primary type of competitive endogenous RNA (ceRNA), lncRNAs can act as miRNA
86   sponges to weaken the effect of miRNAs on mRNAs; therefore, miRNAs play important roles in
87   the molecular mechanism of lncRNAs (Salmena et al. 2011; Xu et al. 2015; Thomson & Dinger
88   2016; Lv et al. 2020). ceRNA crosstalk appears to be closely related to miRNA expression
89   levels, and the specificity of interactions may depend on the expression profile of miRNAs (Ala
90   et al. 2013; Xu et al. 2015; Xu et al. 2016). In addition, some lncRNAs may be co-regulated in
91   expression networks, suggesting that multiple lncRNAs may interact with specific miRNA
92   clusters to regulate biological processes in a coordinated way (Li et al. 2015; Thomson & Dinger
93   2016). It is therefore reasonable to expected the expression patterns of lncRNAs and miRNAs to
94   be important information for predicting interactions.
95   How can known information be used to effectively predict lncRNA–miRNA interactions?
96   This problem is essentially a linkage prediction problem for the interactions network (Lü & Zhou
97   2011). Previous studies have focused on improving the accuracy of prediction methods, without
98   investigating the extent to which lncRNA–miRNA interactions can be predicted. In this study,
99   we refer to the model proposed by Lü et al. (2015) and use a structural consistency index to
100   measure the predictability of interactions. Given the adjacency matrix of a network, its first-order
101   matrix perturbation can describe the inherent properties of the network structure and can reflect
102   the link predictability without prior knowledge of the network organization. That is, the
103   consistency of structural features before and after the removal of a random component of the
104   links reflects the predictability of a network. A model based on the structural perturbation

105   method was developed to predict lncRNA–miRNA interactions. We refer to this model as

106   SPMLMI (i.e., Structural Perturbation Method for predicting LncRNA–MiRNA Interactions).

107   This study makes several key contributions. First, we constructed an lncRNA–miRNA bilayer

108   network composed of three kinds of information (lncRNA similarity, miRNA similarity, and the

109   validated lncRNA–miRNA interactions). Second, we used the structural consistency index to

110   measure the link predictability of multiple related networks and found that the structural

111   consistency of the bilayer network was greater than that of any of single network, supporting the

112   use of the bilayer network for link prediction. Furthermore, we employed SPMLMI to predict

113   potential lncRNA–miRNA interactions. Last, experimental results showed that the newly

114   developed method achieves good prediction performance using different datasets and that its

115   generalizability is better than those of many previous methods. Overall, SPMLMI is an effective

116   method for predicting unknown lncRNA–miRNA interactions.

117

Materials & Methods

119   **Data collection**

120   Our goal was to develop a method that is able to predict candidate interactions between

121   lncRNAs and miRNAs. Three kinds of data were collected from various databases and a bilayer

122   network of lncRNAs and miRNAs was constructed to represent the complex relationships. First,

123   known lncRNA–miRNA interaction data were downloaded from the lncRNASNP database

124   version 2.0 (http://bioinfo.life.hust.edu.cn/lncRNASNP) (Miao et al. 2018) and used to construct

125   an lncRNA–miRNA network. Second, expression data for lncRNAs were downloaded from the

126   NONCODE database (http://www.noncode.org/) (Fang et al. 2018) and used to calculate the

127   expression similarity between lncRNAs and to construct the lncRNA similarity network.

128   Additionally, miRNA expression data were downloaded from the miRmine database

129   (http://guanlab.ccmb.med.umich.edu/mirmine) (Panwar et al. 2017) and used to construct the

130   miRNA similarity network. Table 1 shows the data obtained from publicly available databases.

131

132   **Construction of an lncRNA–miRNA bilayer network**

133   An lncRNA–miRNA bilayer network containing two kinds of nodes (miRNAs and lncRNAs)

134   and three types of relationships (miRNA–miRNA similarity relationship, lncRNA–lncRNA

135   similarity relationship, as well as the known lncRNA–miRNA interaction relationship) was

136   constructed.

137   To construct the lncRNA–miRNA interaction network, data were downloaded from the

138   version 2.0 of the lncRNASNP database (Miao et al. 2018). This database contains known

139   lncRNA–miRNA interactions with experimental validation collected from version 2.0 of the

140   starBase (Li et al. 2014). Then, we define matrix $LMnet$ as the known lncRNA–miRNA

141   interaction network, where if lncRNA i and miRNA j are connected, the element $LMnet = 1$;

142   otherwise, $LMnet = 0$. A total of 39366 associations between 3150 lncRNAs and 262 miRNAs

143   were retained after removing duplicates. Figure 1(b) presents a simple example of lncRNA–

144   miRNA interaction network construction.

145   The lncRNA similarity network and miRNA similarity network were constructed based on

146   expression profiles. First, expression profile data for miRNAs were downloaded from the

147   miRmine database (Panwar et al. 2017). The database contains data for 2588 miRNAs in 135

148   different human tissues and cell lines. After removing missing items and matching the miRNA

149   IDs in the lncRNA–miRNA interaction network ($LMnet$),expression data for 262 miRNAs in 124

150   different human tissues and cell lines were retained. Second, lncRNA expression profiles were

151   downloaded from the NONCODE database (Fang et al. 2018). Expression profiles of 171814

152   transcripts in 16 different human tissues and 8 cell lines were obtained. After the removal of

153   duplicates and matching lncRNA IDs in $LMnet$, 3150 lncRNA expression records for 24 different

154   human tissues and cell lines were retained.

155   To describe the expression characteristics of ncRNAs in different tissues or cell lines in an

156   unbiased fashion, the expression profiles were standardized before calculating pairwise similarity

157   values. The Pearson correlation coefficient was used to measure the similarity of ncRNAs and

158   the expression similarity matrix of lncRNAs and miRNAs was obtained. The similarity matrix

159   represented the similarity network of these two types of ncRNAs, referred to as $LSnet$ and $MSnet$,

160   respectively, as shown in Figure 1 (a) and (c). For a pair of RNAs, a higher correlation indicates

161   greater similarity in expression, in general.

162   Finally, the above three networks are integrated to build a bilayer network. Specifically, the

163   lncRNA–miRNA bilayer network can be represented by a block adjacency matrix $A \in R^{N \times N}$,

$$A = \begin{bmatrix} LSnet & LMnet \\ LMnet^T & MSnet \end{bmatrix}, \tag{1}$$

164   where $N = 3412$ (the total number of nodes in the network, 3150 lncRNAs and 262 miRNAs) and

165   $LMnet^T$ is the transpose of $LMnet$.

166   Overall, the lncRNA–miRNA bilayer network was composed of the miRNA similarity

167   network, lncRNA similarity network, and edges connecting the two networks. Figure 1

168   demonstrates the workflow for constructing the bilayer network and its matrix representation.

169

170   **Structural consistency and structural perturbation method**

171   Structural consistency was first proposed by Lü et al. (2015) and can be used to evaluate the
172   likelihood of links prediction in complex network. It is defined as the consistency of network
173   structural features before and after the elimination of partial associations at random. For the
174   rationale and the specific derivation, refer to the supplementary material
175   (Supplementary_File_S1). In this study, this method was applied to the lncRNA–miRNA bilayer
176   network A and its sub-networks ($LSnet$ and $MSnet$) and the structural consistency of each network
177   was evaluated separately.
178   Generally, the link prediction problem refers to the problem of estimating the probability of
179   the existence of unobserved links according to known topological information. The network
180   structure perturbation method involved in the structure consistency calculation process can be
181   used to predict missing links (Lü et al., 2015). After obtaining the lncRNA–miRNA bilayer
182   network (adjacency matrix, A) in Section 2.2, we derive the perturbation matrix A' by the
183   structural perturbation method. The prediction matrix $\hat{A}'$ is then obtained by averaging t
184   independent perturbation calculations. See the supplementary materials for more details
185   (Supplementary_File_S1). In this way, elements in the prediction matrix $\hat{A}'$ can be interpreted as
186   scores for each pair of nodes in the lncRNA–miRNA bilayer network. The scores in $\hat{A}'$
187   determine the extent of all unobserved lncRNA–miRNA interactions, and we assume that a
188   higher score corresponds to a greater likelihood of a potential interaction.
189
190   **Evaluation metrics**
191   To evaluate the performance of SPMLMI, k-fold cross-validation was performed based on
192   verified lncRNA–miRNA interactions downloaded from the lncRNASNP database (Miao et al.
193   2018). In the cross-validation process, the original set was randomly divided into k equally sized
194   subsets. Among these k subsets, one was used as the validation data for testing the model, and
195   the others were used as the training data. Specifically, we randomly reset 1/k known interaction
196   entries of the training dataset to unknown, and prioritize the candidate interaction entries in the
197   prediction matrix by SPMLMI scores. The known interactions are considered as positive, and the
198   rest as negative. True positive rate is the proportion of overlap between the top 1/k ranked
199   candidate interactions with the known interactions. Then, the cross-validation was repeated k
200   times, taking each of the k subsets as the validation data in turn. The cross-validation results are
201   averaged to produce a single estimation.
202   The AUC (area under the receiver operating characteristic curve) was estimated to evaluate
203   the performance of lncRNA–miRNA interaction prediction results. In particular, the TPR (true
204   positive rate) and FPR (false positive rate) were calculated by varying the threshold and the ROC
205   (receiver operating characteristic) curve and AUC value were obtained, where TPR and FPR
206   represent the percentage of test samples above or below the given thresholds, respectively.

207   In addition, we applied SPMLMI to datasets analyzed by two additional methods, EPLMI
208   (Huang et al. 2018) and SNFHGILMI (Fan et al. 2020), for an objective comparison of
209   performance.

210

Results

212   **Structural consistency of different networks**
213   We calculated the structural consistency of three datasets (i.e., the EPLMI dataset provided by
214   Huang, et al. (2018), SNFHGILMI dataset provided by Fan, et al. (2020), and SPMLMI dataset
215   obtained in this study). Each group of datasets was composed of an miRNA similarity network (

216    $MSnet$), an lncRNA similarity network ($LSnet$), and a bilayer network ($Bilayer-net$). We

217   randomly selected 10% of the total links from each network as the perturbation set (note that the
218   bilayer network contains two kinds of links). As shown in Figure 2 (a), in three different

219   datasets, the structural consistency values for the lncRNA–miRNA bilayer network ($Bilayer-net$,

220   yellow bar) were 0.2413 ± 0.0015, 0.3036 ± 0.0011, and 0.2168 ± 0.0005, respectively. The

221   structural consistency of these three bilayer networks ($Bilayer-net$) were all greater than that of

222   the similarity network (where the blue bar is $MSnet$ and the orange bar is $LSnet$ in Figure 2),
223   indicating that considering more information can improve the structural consistency of the
224   network. These results show that the construction of the bilayer network improves the inherent
225   link predictability of the network and will be helpful for the prediction of "missing links" (i.e.,
226   newly established or novel interactions).
227

228   **Performance evaluation and parameter optimization**
229   First, to evaluate the performance of our proposed SPMLMI for the prediction of lncRNA–
230   miRNA interactions, we adopted 2-fold, 5-fold and 10-fold cross-validation frameworks. To
231   avoid the bias of sample division in cross-validation, we repeated each experiment 10 times and
232   took the average value. Figure 2 (c) shows the corresponding ROC curve for each k-fold cross-
233   validation. The average areas under the curves were 0.8617 ± 0.0013, 0.9512 ± 0.0034, and
234   0.9517 ± 0.0016 for 2-fold, 5-fold, and 10-fold cross-validation, respectively. These results
235   indicate that k-fold cross-validation increase improves the performance of SPMLMI. There was
236   no significant difference (paired t-test, p-value= 0.1997) between 5-fold and 10-fold cross-
237   validation. Therefore, in subsequent experiments, we use 5-fold cross-validation to save time.

238   Second, to optimize t (i.e., the number of perturbations), we tuned the parameter from 1 to 16.
239   Figure 2 (b) shows boxplots of the variation of AUC values with increasing t. The AUC values
240   increased as t increased; however, for values of t greater than or equal to 8, AUC was relatively
241   stable. For simplicity, we set the parameter t to 8 for subsequent experiments.
242
243   **Comparison with previous methods**
244   Furthermore, we compared SPMLMI with the previously established EPLMI (Huang et al.,
245   2018) based on the two-way diffusion algorithm and SNFHGILMI (Fan et al., 2020) based on
246   the heterogeneous graph inference method. Initially, we evaluated the same dataset (i.e., the
247   SPMLMI dataset obtained in this study, see Section 2 for details) by these three methods. Figure
248   3 (a) shows the ROC curves for the prediction performance of these three models. The average
249   AUC for SPMLMI was 0.9512 ± 0.0034, which was higher than the AUC values of $0.7999 \pm$
250   0.0014 for EPLMI and 0.5163 ± 0.0013 for SNFHGILMI, respectively. We also compared the
251   methods using the original datasets. Using the EPLMI dataset provided by Huang, et al. (2018),
252   the average AUC values for the three methods (SPMLMI, EPLMI, and SNFHGILMI) were
253   0.8767 ± 0.0033, 0.8417 ± 0.0017, and 0.5184 ± 0.0023, respectively (Figure 3 (b)). For the
254   SNFHGILMI dataset provided by Fan et al. (2020), the average AUC values were $0.8653 \pm$
255   0.0021, 0.8570 ± 0.0012, and 0.9426 ± 0.0035, respectively (Figure 3 (c)).
256   These results prove that the newly established SPMLMI has good performance and
257   generalizability across datasets.
258
259   **Case studies**
260   Two lncRNAs, SNHG14 (UBE3A-ATS/NONHSAT041137) and MALAT1
261   (NONHSAT022132), with important biological functions were selected to validate the prediction
262   performance of SPMLMI. SNHG14, also known as UBE3A-ATS, overlaps with the entire
263   *UBE3A* gene and its paternal expression is thought to negatively regulate paternal *UBE3A*
264   expression in neurons, thereby determining tissue-specific imprinting of *UBE3A* in the brain
265   (Galiveti et al. 2014; Stanurova et al. 2016). MALA1 is located less than 70 kb from
266   NCRNA00084 on chromosome 11q13.1 (Hutchinson et al. 2007) and is dysregulated in many
267   diseases, such as alcohol-related diseases, endometrioid endometrial carcinoma, and non-small
268   cell lung cancer (Ji et al. 2003; Kryger et al. 2012; Li et al. 2016). For each of this two lncRNAs
269   in the interaction network, known interaction entries were randomly reset to unknown. We
270   implemented SPMLMI using the reset data set and prioritized all candidate miRNAs according
271   to their scores in the prediction matrix.
272   As shown in Table 2, among the top 10% (26) predicted SNHG14-related miRNAs, 18
273   lncRNA–miRNA interactions were confirmed by experimental validation (containing 17 of the

274 21 known interactions we collected from the lncRNASNP2 database (Miao et al. 2018)). It
275 should be noted that the miRNA hsa-miR-126-3p (marked with an asterisk in Table 2), ranked
276 16th, was not collected from the lncRNASNP2 database (Miao et al. 2018) but was reported by
277 an experimental study of transcriptome-wide Ago2:RNA interactions in human brain tissues
278 using HITS-CLIP (i.e., CLIP-seq) (Boudreau et al. 2014). This information was retrieved from
279 DIANA-LncBase v3 (Karagkouni et al. 2020). Several studies have shown that miR-126 directly
280 targets the 3'-UTR of insulin receptor substrate-1 (IRS-1) (Ryu et al. 2011; Zhang et al. 2008;
281 Zhou et al. 2013).Over-expression of miR-126 leads to downregulation of IRS-1 expression, and
282 a consequent impairment in insulin signaling (Ryu et al. 2011). Moreover, the research results of
283 Zhou et al. (2013) showed over-expression of miR-126 down-regulated IRS-1, suppressed AKT
284 and ERK1/2 activation, CRC cells proliferation, migration, invasion, and caused cell cycle arrest.
285 Therefore, lncRNA SNHG14, as an interacting molecule of miR-126, may play a potential role
286 in regulating cell metabolism and cell cycle.
287   Similarly, for MALAT1, we examined the candidate miRNAs ranked by prediction scores.
288 The results showed that the top 10% of predicted MALAT1-related miRNAs were all verified in
289 published studies (Supplementary_Table_S1). The reason is most likely that lncRNA MALAT1
290 has more known interaction information (containing 113 known interactions collected from the
291 lncRNASNP2 database (Miao et al. 2018)). Therefore, we examined more predicted candidate
292 miRNAs and found that a total of 106 were verifiable. Additionally, in a genome-wide analysis
293 of miRNA–mRNA interactions in marrow stromal cells, the interaction between lncRNA
294 MALAT1 and miRNA hsa-miR-487b-3p was confirmed (Balakrishnan et al. 2014); this
295 candidate miRNA was ranked 91st (Supplementary_Table_S1) based on prediction scores. The
296 experimental study by Xi et al. (2013) demonstrated that miR-487b directly targets gene Wnt5a.
297 The protein Wnt5a is an important activator of the Wnt/Ca^{2+} signaling pathway, which plays an
298 important role in embryonic development, cell differentiation, inflammation and other
299 physiological processes (De 2011). Because of sharing miRNA miR-487b with Wnt5a gene,
300 lncRNA MALAT1 may also be involved in this complex signaling network.
301   The above results suggest that SPMLMI has good prediction ability and can be used to screen
302 candidate target miRNAs for a given lncRNA.
303

Discussion

305   The identification of associations among biomolecules has important implications for our
306 understanding of biological processes. In view of the complexity of intermolecular interaction
307 networks, it is necessary to establish a prediction model that makes full use of prior information.
308 Lü et al. (2015) studied the link predictability of complex networks and proposed the structural

309 perturbation method for predicting "missing links" in networks. Zeng et al. (2018) applied this
310 algorithm to biological networks. They constructed a bilayer network and predicted unknown
311 miRNA–disease associations by the perturbation of specific diseases in the network, achieving
312 good results. We further extended the structural perturbation method; the key difference in our
313 study was the application of structural perturbation to the entire known lncRNA–miRNA
314 interaction network, without focusing on a specific ncRNA (lncRNA or miRNA).
315 The ceRNA hypothesis (Salmena et al. 2011) predicts that similarity in the expression profile
316 of a pair of lncRNAs is beneficial if they interact with the same miRNA. In our study, the
317 successful prediction of lncRNA–miRNA interactions based on ncRNA expression profile
318 similarity seems to support this hypothesis; however, the reality is more complex for a few
319 reasons. First, the ncRNA expression profile data used in this study were collected from different
320 human tissues and cell lines (see section 2.2). The similarity network reflects the generalized
321 expression pattern of a pair of ncRNAs, rather than their expression pattern in a specific cellular
322 environment. Second, an interaction between an lncRNA and miRNA may involve other
323 molecules. Because lncRNAs are abundant in vivo and have various interactions with other
324 molecules (Liu et al. 2017; Liu et al. 2015). Therefore, the indirect interactions of lncRNAs and
325 miRNAs via other molecules, such as protein complexes, cannot be ignored.
326 Based on our results, we suggest that the network structure perturbation method has two main
327 advantages over other approaches: (1) it can make full use of prior information in the network (2)
328 and it captures network topology features well. However, the method has limitations. First, it
329 requires prior knowledge of network nodes and links to achieve good performance. Second, our
330 predicted candidate interaction pairs may contain indirect relationships involving other
331 molecules (e.g., proteins, nucleic acids, etc.). The above summary of advantages and
332 disadvantages are aspects of the method that should be noted, and also point out a direction we
333 need to further research in the future. In addition, previous researchers have many excellent
334 research work on miRNA-disease associations (Chen et al. 2019b) as well as lnRNA-disease
335 associations (Chen & Yan 2013). The studies of ncRNA-disease association by Chen et al.
336 (2013; 2019b) inspired us to expand ncRNA-disease association studies by combining a priori
337 information and reliable lncRNA-miRNA prediction information in our future research work.
338 Overall, the approach is valuable for large biomolecular networks and the generalized
339 prediction of unknown links in the network or for analyses of specific molecules in a network.
340 The method can be adapted according to the specific research aims.
341

342 ## Conclusions

343  We propose a framework named SPMLMI based on a bilayer network to predict interactions
344 between lncRNAs and miRNAs. The bilayer network integrates similarity information for two
345 types of ncRNA molecules and known interaction information. By using SPMLMI, we obtained
346 a perturbation matrix and corresponded each matrix element to the correlation score between
347 pairs of nodes in the network. In this way, missing links (i.e., unknown lncRNA–miRNA
348 interactions) in the network are assigned correlation scores, where high scores indicate probable
349 interactions. SPMLMI was feasible and effective for predicting lncRNA–miRNA interactions
350 and represents a substantial improvement over SPMLMI.
351  Extensive regulatory functions of ncRNAs have been confirmed. Based on expression data,
352 our method infers potential interactions between lncRNAs and miRNAs, providing a basis for
353 further functional analyses. However, other regulatory interactions, such as lncRNA–mRNA
354 interactions, miRNA–mRNA interactions, and ncRNA–protein interactions, have not been
355 considered and will be a focus of future research.
356

Data availability

358  The codes and datasets are available online at https://github.com/xumingmin/SPMLMI.git
359

Acknowledgements

361

References

363 Ala U, Karreth FA, Bosia C, Pagnani A, Taulli R, Leopold V, Tay Y, Provero P, Zecchina R, and
364   Pandolfi PP. 2013. Integrated transcriptional and competitive endogenous RNA networks are
365   cross-regulated in permissive molecular environments. *Proc Natl Acad Sci U S A* 110:7154-7159
366   DOI: 10.1073/pnas.1222509110.
367 Balakrishnan I, Yang X, Brown J, Ramakrishnan A, Torok-Storb B, Kabos P, Hesselberth JR, and Pillai
368   MM. 2014. Genome-wide analysis of miRNA-mRNA interactions in marrow stromal cells. *Stem*
369   *Cells* 32:662-673 DOI: 10.1002/stem.1531.
370 Béthune J, Artus-Revel CG, and Filipowicz W. 2012. Kinetic analysis reveals successive steps leading to
371   miRNA-mediated silencing in mammalian cells. *EMBO Rep* 13:716-723 DOI:
372   10.1038/embor.2012.82.
373 Boudreau RL, Jiang P, Gilmore BL, Spengler RM, Tirabassi R, Nelson JA, Ross CA, Xing Y, and
374   Davidson BL. 2014. Transcriptome-wide discovery of microRNA binding sites in human brain.
375   *Neuron* 81:294-305 DOI: 10.1016/j.neuron.2013.10.062.

376   Braun JE, Huntzinger E, and Izaurralde E. 2013. The Role of GW182 Proteins in miRNA-Mediated Gene
377   Silencing. In: Chan EKL, and Fritzler MJ, eds. *Ten Years of Progress in GW/P Body Research*.
378   New York, NY: Springer New York, 147-163.
379   Cech TR, and Steitz JA. 2014. The noncoding RNA revolution-trashing old rules to forge new ones. *Cell*
380   157:77-94 DOI: 10.1016/j.cell.2014.03.008.
381   Chen X, and Yan GY. 2013. Novel human lncRNA-disease association inference based on lncRNA
382   expression profiles. *Bioinformatics* 29:2617-2624 DOI: 10.1093/bioinformatics/btt426.
383   Chen X, Sun YZ, Guan NN, Qu J, Huang ZA, Zhu ZX, and Li JQ. 2019a. Computational models for
384   lncRNA function prediction and functional similarity calculation. *Brief Funct Genomics* 18:58-82
385   DOI: 10.1093/bfgp/ely031.
386   Chen X, Xie D, Zhao Q, and You ZH. 2019b. MicroRNAs and complex diseases: from experimental
387   results to computational models. *Brief Bioinform* 20:515-539 DOI: 10.1093/bib/bbx130.
388   Chen X, Yan CC, Zhang X, and You ZH. 2017. Long non-coding RNAs and complex diseases: from
389   experimental results to computational models. *Brief Bioinform* 18:558-576 DOI:
390   10.1093/bib/bbw060.
391   Clauset A, Moore C, and Newman ME. 2008. Hierarchical structure and the prediction of missing links in
392   networks. *Nature* 453:98-101 DOI: 10.1038/nature06830.
393   De A. 2011. Wnt/Ca2+ signaling pathway: a brief overview. *Acta Biochim Biophys Sin (Shanghai)*
394   43:745-756 DOI: 10.1093/abbs/gmr079.
395   Fan YX, Cui J, and Zhu QQ. 2020. Heterogeneous graph inference based on similarity network fusion for
396   predicting lncRNA-miRNA interaction. *Rsc Advances* 10:11634-11642 DOI:
397   10.1039/c9ra11043g.
398   Fang S, Zhang L, Guo J, Niu Y, Wu Y, Li H, Zhao L, Li X, Teng X, Sun X, Sun L, Zhang MQ, Chen R,
399   and Zhao Y. 2018. NONCODEV5: a comprehensive annotation database for long non-coding
400   RNAs. *Nucleic Acids Res* 46:D308-D314 DOI: 10.1093/nar/gkx1107.
401   Galiveti CR, Raabe CA, Konthur Z, and Rozhdestvensky TS. 2014. Differential regulation of non-protein
402   coding RNAs from Prader-Willi Syndrome locus. *Sci Rep* 4:6445 DOI: 10.1038/srep06445.
403   Guil S, and Esteller M. 2015. RNA-RNA interactions in gene regulation: the coding and noncoding
404   players. *Trends Biochem Sci* 40:248-256 DOI: 10.1016/j.tibs.2015.03.001.
405   Huang YA, Chan KCC, and You ZH. 2018. Constructing prediction models from expression profiles for
406   large scale lncRNA-miRNA interaction profiling. *Bioinformatics* 34:812-819 DOI:
407   10.1093/bioinformatics/btx672.
408   Huntzinger E, and Izaurralde E. 2011. Gene silencing by microRNAs: contributions of translational
409   repression and mRNA decay. *Nat Rev Genet* 12:99-110 DOI: 10.1038/nrg2936.
410   Hutchinson JN, Ensminger AW, Clemson CM, Lynch CR, Lawrence JB, and Chess A. 2007. A screen for
411   nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains.
412   *BMC Genomics* 8:39 DOI: 10.1186/1471-2164-8-39.
413   Ji P, Diederichs S, Wang W, Boing S, Metzger R, Schneider PM, Tidow N, Brandt B, Buerger H, Bulk E,
414   Thomas M, Berdel WE, Serve H, and Muller-Tidow C. 2003. MALAT-1, a novel noncoding
415   RNA, and thymosin beta4 predict metastasis and survival in early-stage non-small cell lung
416   cancer. *Oncogene* 22:8031-8041 DOI: 10.1038/sj.onc.1206928

417   Karagkouni D, Paraskevopoulou MD, Tastsoglou S, Skoufos G, Karavangeli A, Pierros V,
418   Zacharopoulou E, and Hatzigeorgiou AG. 2020. DIANA-LncBase v3: indexing experimentally
419   supported miRNA targets on non-coding transcripts. *Nucleic Acids Res* 48:D101-D110 DOI:
420   10.1093/nar/gkz1036.
421   Krol J, Loedige I, and Filipowicz W. 2010. The widespread regulation of microRNA biogenesis, function
422   and decay. *Nat Rev Genet* 11:597-610 DOI: 10.1038/nrg2843.
423   Kryger R, Fan L, Wilce PA, and Jaquet V. 2012. MALAT-1, a non protein-coding RNA is upregulated in
424   the cerebellum, hippocampus and brain stem of human alcoholics. *Alcohol* 46:629-634 DOI:
425   10.1016/j.alcohol.2012.04.002.
426   Lai D, and Meyer IM. 2016. A comprehensive comparison of general RNA-RNA interaction prediction
427   methods. *Nucleic Acids Res* 44:e61 DOI: 10.1093/nar/gkv1477.
428   Lee RC, Feinbaum RL, and Ambros V. 1993. The C. elegans heterochronic gene lin-4 encodes small
429   RNAs with antisense complementarity to lin-14. *Cell* 75:843-854 DOI: 10.1016/0092-
430   8674(93)90529-y.
431   Li J, Ma W, Zeng P, Wang J, Geng B, Yang J, and Cui Q. 2015. LncTar: a tool for predicting the RNA
432   targets of long noncoding RNAs. *Brief Bioinform* 16:806-812 DOI: 10.1093/bib/bbu048.
433   Li JH, Liu S, Zhou H, Qu LH, and Yang JH. 2014. starBase v2.0: decoding miRNA-ceRNA, miRNA-
434   ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids
435   Res* 42:D92-97 DOI: 10.1093/nar/gkt1248.
436   Li Q, Zhang C, Chen R, Xiong H, Qiu F, Liu S, Zhang M, Wang F, Wang Y, Zhou X, Xiao G, Wang X,
437   and Jiang Q. 2016. Disrupting MALAT1/miR-200c sponge decreases invasion and migration in
438   endometrioid endometrial carcinoma. *Cancer Lett* 383:28-40 DOI: 10.1016/j.canlet.2016.09.019.
439   Liu CJ, Gao C, Ma Z, Cong R, Zhang Q, and Guo AY. 2017. lncRInter: A database of experimentally
440   validated long non-coding RNA interaction. *J Genet Genomics* 44:265-268 DOI:
441   10.1016/j.jgg.2017.01.004.
442   Liu N, Dai Q, Zheng G, He C, Parisien M, and Pan T. 2015. N(6)-methyladenosine-dependent RNA
443   structural switches regulate RNA-protein interactions. *Nature* 518:560-564 DOI:
444   10.1038/nature14234.
445   Lü L, Pan L, Zhou T, Zhang YC, and Stanley HE. 2015. Toward link predictability of complex networks.
446   *Proc Natl Acad Sci U S A* 112:2325-2330 DOI: 10.1073/pnas.1424644112.
447   Lü LY, and Zhou T. 2011. Link prediction in complex networks: A survey. *Physica a-Statistical
448   Mechanics and Its Applications* 390:1150-1170 DOI: 10.1016/j.physa.2010.11.027.
449   Lv D, Xu K, Jin X, Li J, Shi Y, Zhang M, Jin X, Li Y, Xu J, and Li X. 2020. LncSpA: LncRNA Spatial
450   Atlas of Expression across Normal and Cancer Tissues. *Cancer Res* 80:2067-2071 DOI:
451   10.1158/0008-5472.CAN-19-2687.
452   Miao YR, Liu W, Zhang Q, and Guo AY. 2018. lncRNASNP2: an updated database of functional SNPs
453   and mutations in human and mouse lncRNAs. *Nucleic Acids Res* 46:D276-D280 DOI:
454   10.1093/nar/gkx1004.
455   Morris KV, and Mattick JS. 2014. The rise of regulatory RNA. *Nat Rev Genet* 15:423-437 DOI:
456   10.1038/nrg3722.

457   Panwar B, Omenn GS, and Guan Y. 2017. miRmine: a database of human miRNA expression profiles.
458   *Bioinformatics* 33:1554-1560 DOI: 10.1093/bioinformatics/btx019.

459   Quinn JJ, and Chang HY. 2016. Unique features of long non-coding RNA biogenesis and function. *Nat*
460   *Rev Genet* 17:47-62 DOI: 10.1038/nrg.2015.10.

461   Rinn JL, and Chang HY. 2012. Genome regulation by long noncoding RNAs. *Annu Rev Biochem* 81:145-
462   166 DOI: 10.1146/annurev-biochem-051410-092902.

463   Ryu HS, Park SY, Ma D, Zhang J, and Lee W. 2011. The induction of microRNA targeting IRS-1 is
464   involved in the development of insulin resistance under conditions of mitochondrial dysfunction
465   in hepatocytes. *PLOS ONE* 6:e17343 DOI: 10.1371/journal.pone.0017343.

466   Sabin LR, Delas MJ, and Hannon GJ. 2013. Dogma derailed: the many influences of RNA on the
467   genome. *Mol Cell* 49:783-794 DOI: 10.1016/j.molcel.2013.02.010.

468   Salmena L, Poliseno L, Tay Y, Kats L, and Pandolfi PP. 2011. A ceRNA hypothesis: the Rosetta Stone of
469   a hidden RNA language? *Cell* 146:353-358 DOI: 10.1016/j.cell.2011.07.014.

470   Stanurova J, Neureiter A, Hiber M, de Oliveira Kessler H, Stolp K, Goetzke R, Klein D, Bankfalvi A,
471   Klump H, and Steenpass L. 2016. Angelman syndrome-derived neurons display late onset of
472   paternal UBE3A silencing. *Sci Rep* 6:30792 DOI: 10.1038/srep30792.

473   Thomson DW, and Dinger ME. 2016. Endogenous microRNA sponges: evidence and controversy. *Nat*
474   *Rev Genet* 17:272-283 DOI: 10.1038/nrg.2016.20.

475   Voinnet O. 2009. Origin, biogenesis, and activity of plant microRNAs. *Cell* 136:669-687 DOI:
476   10.1016/j.cell.2009.01.046.

477   Wightman B, Ha I, and Ruvkun G. 1993. Posttranscriptional regulation of the heterochronic gene lin-14
478   by lin-4 mediates temporal pattern formation in C. elegans. *Cell* 75:855-862 DOI: 10.1016/0092-
479   8674(93)90530-4.

480   Xi S, Xu H, Shan J, Tao Y, Hong JA, Inchauste S, Zhang M, Kunst TF, Mercedes L, and Schrump DS.
481   2013. Cigarette smoke mediates epigenetic repression of miR-487b during pulmonary
482   carcinogenesis. *J Clin Invest* 123:1241-1261 DOI: 10.1172/JCI61271.

483   Xu J, Feng L, Han Z, Li Y, Wu A, Shao T, Ding N, Li L, Deng W, Di X, Wang J, Zhang L, Li X, Zhang
484   K, and Cheng S. 2016. Extensive ceRNA-ceRNA interaction networks mediated by miRNAs
485   regulate development in multiple rhesus tissues. *Nucleic Acids Res* 44:9438-9451 DOI:
486   10.1093/nar/gkw587.

487   Xu J, Li Y, Lu J, Pan T, Ding N, Wang Z, Shao T, Zhang J, Wang L, and Li X. 2015. The mRNA related
488   ceRNA-ceRNA landscape and significance across 20 major cancer types. *Nucleic Acids Res*
489   43:8169-8182 DOI: 10.1093/nar/gkv853.

490   Yoon JH, Abdelmohsen K, and Gorospe M. 2014. Functional interactions among microRNAs and long
491   noncoding RNAs. *Semin Cell Dev Biol* 34:9-14 DOI: 10.1016/j.semcdb.2014.05.015.

492   Zeng X, Liu L, Lu L, and Zou Q. 2018. Prediction of potential disease-associated microRNAs using
493   structural perturbation method. *Bioinformatics* 34:2425-2432 DOI:
494   10.1093/bioinformatics/bty112.

495   Zhang J, Du YY, Lin YF, Chen YT, Yang L, Wang HJ, and Ma D. 2008. The cell growth suppressor,
496   mir-126, targets IRS-1. *Biochem Biophys Res Commun* 377:136-140 DOI:
497   10.1016/j.bbrc.2008.09.089.

498 Zhou Y, Feng X, Liu YL, Ye SC, Wang H, Tan WK, Tian T, Qiu YM, and Luo HS. 2013. Down-
499   regulation of miR-126 is associated with colorectal cancer cells proliferation, migration and
500   invasion by targeting IRS-1 via the AKT and ERK1/2 signaling pathways. *PLOS ONE* 8:e81203
501   DOI: 10.1371/journal.pone.0081203.

Summary of data used in this study

Database	Data types	Number of links/nodes
lncRNASNP	lncRNA–miRNA interactions (validated)	39366 (interactions/edges/links)
NONCODE	lncRNA expression profiles	3150 (lncRNA nodes)
miRmine	miRNA expression profiles	262 (miRNA nodes)

Prediction results for the top 10% SNHG14-related miRNAs ranked by prediction scores.

Rank	MiRNAs	Reference (PMID)	Rank	MiRNAs	Reference (PMID)
1	hsa-miR-425-5p	23313552	14	hsa-miR-32-5p	24668909; 28030800; 22291592
2	hsa-miR-181b-5p	23313552; 28030800; 22012620	15	hsa-miR-92a-3p	24668909; 28030800; 22291592
3	hsa-miR-340-5p	24668909; 23313552	16	hsa-miR-126-3p *	24389009
4	hsa-miR-25-3p	24668909; 28030800	17	hsa-miR-181c-5p	23313552; 28030800; 22012620
5	hsa-miR-363-3p	24668909; 28030800; 22291592	18	hsa-miR-382-5p	23313552
6	hsa-miR-543	23313552; 28030800	19	hsa-miR-551a	Unconfirmed
7	hsa-miR-433-3p	22012620	20	hsa-miR-200c-3p	Unconfirmed
8	hsa-miR-136-5p	22012620	21	hsa-miR-187-3p	Unconfirmed
9	hsa-miR-181a-5p	23313552; 28030800; 22012620	22	hsa-miR-378b	Unconfirmed
10	hsa-miR-181d-5p	23313552; 28030800; 22012620	23	hsa-miR-100-5p	Unconfirmed
11	hsa-miR-376c-3p	24668909; 22100165	24	hsa-miR-758-3p	Unconfirmed
12	hsa-miR-206	24906430; 22012620	25	hsa-miR-10b-5p	Unconfirmed
13	hsa-miR-92b-3p	24668909; 28030800; 22291592	26	hsa-miR-300	Unconfirmed

Figure 1

Flowchart of SPMLMI.

(a) The lncRNA expression profiles were downloaded from the NONCODE database and used to calculate the lncRNA similarity network (*LSnet*). (b) Known lncRNA–miRNA interactions were downloaded from the lncRNASNP database and used to construct the lncRNA–miRNA interaction network (*LMnet*). (c) The miRNA expression profiles were downloaded from the miRmine database and used to construct the miRNA similarity network (*MSnet*). (d) Three networks were integrated to construct the lncRNA–miRNA bilayer network. (e) The perturbed network was obtained by using SPMLMI.
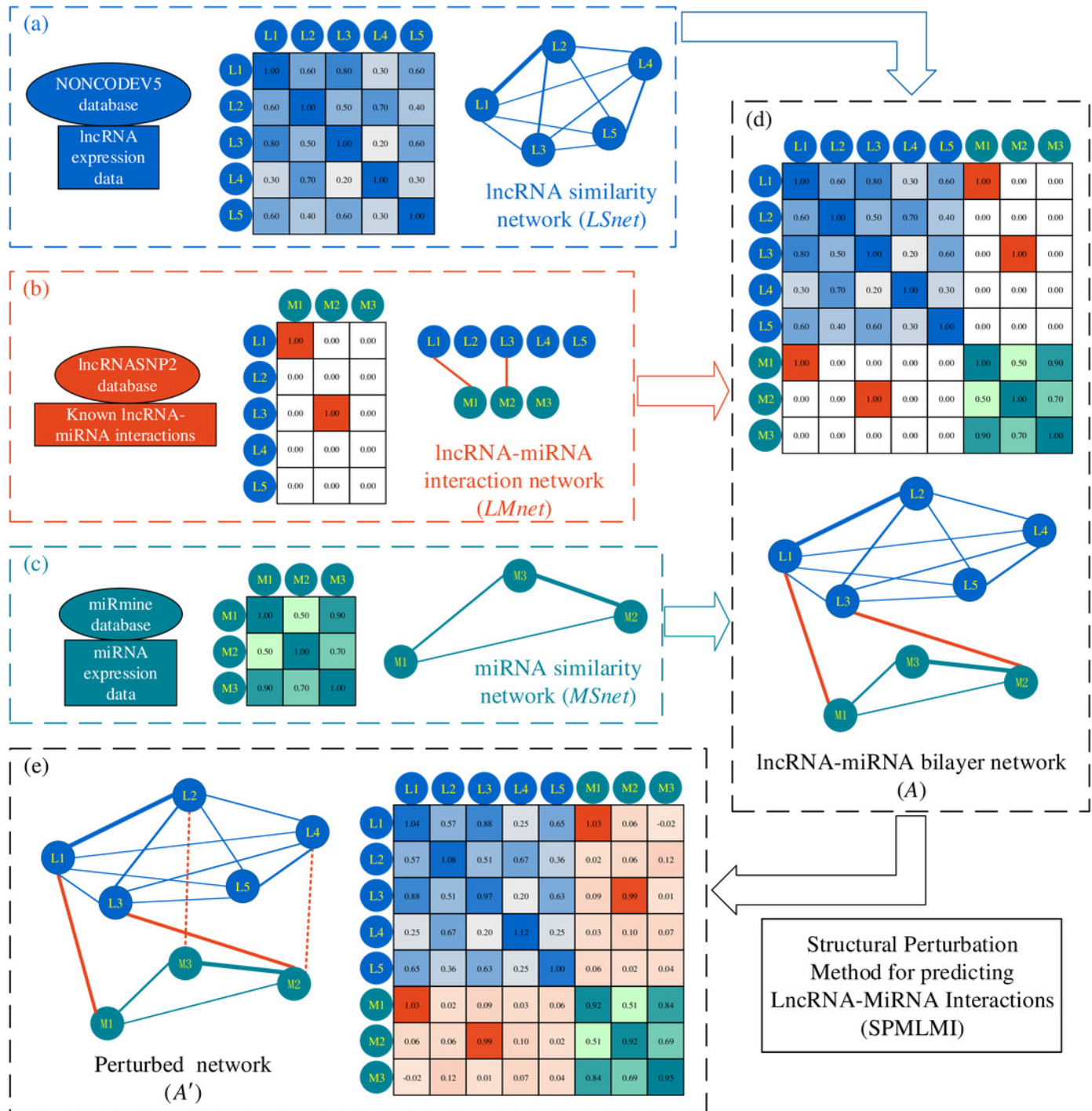
Figure 2

Experimental results on structural consistency, parameter t optimization and performance evaluation.

(a) Structural consistency of related networks. (b) Effect of the parameter t on model performance. (c) Performance of SPMLMI under different k-fold cross-validation frameworks.
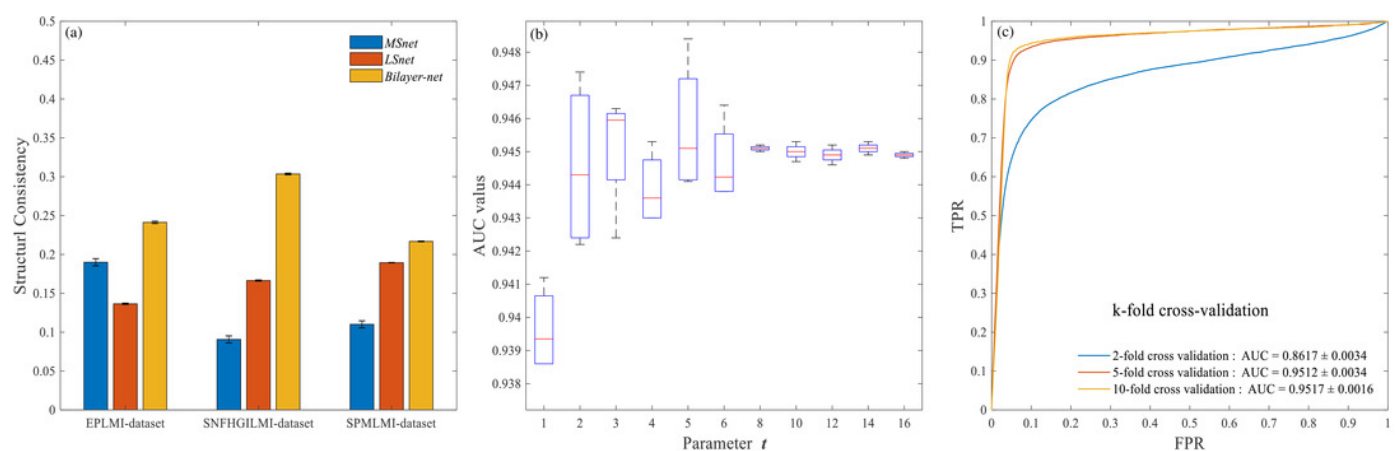
Figure 3

Comparison with previous methods.

(a) Experimental results for the three methods on the SPMLMI dataset. (b) Experimental results for the three methods on the EPLMI dataset. (c) Experimental results for three methods on the SNFHGILMI dataset.