



Article

A Novel Probability Model for LncRNA–Disease Association Prediction Based on the Naïve Bayesian Classifier

Jingwen Yu ¹ , Pengyao Ping ¹ , Lei Wang ^{1,2,*} , Linai Kuang ^{1,2} , Xueyong Li ² and Zhelun Wu ³

¹ Key Laboratory of Intelligent Computing & Information Processing, Xiangtan University, Xiangtan 411105, China; jingwen.yu18@gmail.com (J.Y.); pengyao.ping2017@gmail.com (P.P.); kla@xtu.edu.cn (L.K.)

² College of Computer Engineering & Applied Mathematics, Changsha University, Changsha 410001, China; xueyong.li111@gmail.com

³ Department of Computer Science, Princeton University, Princeton, NJ 08544, USA; zhelunw@princeton.edu

* Correspondence: wanglei@xtu.edu.cn; Tel.: +0731-58292968

Received: 26 May 2018; Accepted: 3 July 2018; Published: 8 July 2018



Abstract: An increasing number of studies have indicated that long-non-coding RNAs (lncRNAs) play crucial roles in biological processes, complex disease diagnoses, prognoses, and treatments. However, experimentally validated associations between lncRNAs and diseases are still very limited. Recently, computational models have been developed to discover potential associations between lncRNAs and diseases by integrating multiple heterogeneous biological data; this has become a hot topic in biological research. In this article, we constructed a global tripartite network by integrating a variety of biological information including miRNA–disease, miRNA–lncRNA, and lncRNA–disease associations and interactions. Then, we constructed a global quadruple network by appending gene–lncRNA interaction, gene–disease association, and gene–miRNA interaction networks to the global tripartite network. Subsequently, based on these two global networks, a novel approach was proposed based on the naïve Bayesian classifier to predict potential lncRNA–disease associations (NBCLDA). Comparing with the state-of-the-art methods, our new method does not entirely rely on known lncRNA–disease associations, and can achieve a reliable performance with effective area under ROC curve (AUCs) in leave-one-out cross validation. Moreover, in order to further estimate the performance of NBCLDA, case studies of colorectal cancer, prostate cancer, and glioma were implemented in this paper, and the simulation results demonstrated that NBCLDA can be an excellent tool for biomedical research in the future.

Keywords: lncRNA–disease associations; tripartite network; quadruple network; prediction model; Naïve Bayesian Classifier

1. Introduction

Long non-coding RNAs (lncRNAs), those with over 200 nucleotides in length [1–3], are considered a new class of non-protein-coding transcripts. Much research evidence has shown that lncRNAs participate in almost the entire cell life cycle through various mechanisms and play significant roles in multiple biological processes including transcription, translation, epigenetic regulation, splicing, differentiation, immune response, cell cycle control, and so on [4–8]. In particular, the mutations and dysregulations of lncRNAs have been proven to be closely related to various human complex diseases [9–11], including AIDS [12], diabetes [13], Alzheimer’s Disease (AD) [14], and many types of cancers such as breast [15], prostate [16], hepatocellular [17], and bladder cancer [18]. For instance,

the expression of the lncRNA called HOTAIR was shown to be higher in primary breast tumors and metastases, and the HOTAIR expression level was proven to be a powerful predictor of eventual metastasis and death [19,20]. Additionally, the lncRNA MALAT1 was demonstrated as a prognostic indicator as well as a therapeutic target and acts as a potential therapeutic method for preventing lung cancer metastasis, which is targeted by antisense oligonucleotides (ASO) [21]. Moreover, recent studies have shown that the human *H19* gene is frequently overexpressed in the myometrium and stroma during pathological endometrial proliferative events [22].

Obviously, predicting potential associations between lncRNAs and diseases would contribute to systematically understanding the pathogenesis of complex diseases at the molecular level and facilitate the identification of biomarkers for disease diagnosis, treatment, and prediction of response to therapy. However, relatively few experiments have supported lncRNA–disease associations until now. Hence, developing effective computational methods to uncover the potential associations between lncRNAs and diseases has become a hot topic in recent years. In general, existing models for predicting potential associations between lncRNAs and diseases can be divided into three categories. Among them, the first kind of methods are based on known disease-related lncRNAs. For example, Sun et al. proposed a model named RWRlncD [23], which carried out a random walk with the restart method on an lncRNA functional similarity network. This method uncovered potential associations between lncRNAs and diseases by integrating the disease similarity network, the lncRNAs functional network, and known lncRNA–disease associations. Ping et al. developed a method based on a newly constructed bipartite network, which relies on the known associations between lncRNAs and diseases [24]. Yang et al. constructed a coding–non-coding gene–disease bipartite network based on known associations between diseases and disease-causing genes (including lncRNAs). Then, they developed an iterative algorithm to uncover the possible links in the newly constructed bipartite network [25]. Ding et al. proposed a new model named TPGLDA to predict potential lncRNA–disease associations by integrating gene–disease associations with lncRNA–disease associations [26].

Different from the first kind of methods based on known lncRNA–disease associations, the second category of prediction models does not rely on known disease-related lncRNAs. For example, Chen et al. proposed a new method called HGLDA by integrating micro-RNA (miRNA)–disease associations and lncRNA–miRNA interactions. A hypergeometric distribution test is then applied to identify potential lncRNA–disease associations [27]. Liu et al. developed a computational framework by integrating human lncRNA expression profiles, gene expression profiles, and human disease-associated gene data to predict potential human lncRNA–disease associations [28]. Li et al. put forward a prediction method on account of the information of genome location to globally discover potential human lncRNAs related to vascular disease [29]. Gu et al. proposed a random walk-based model to identify potential associations between lncRNAs and diseases, which can be applied for predicting a disease without known associated lncRNAs and for inferring an lncRNA without known associated diseases [30].

In recent years, an increasing number of studies have been developed for understanding the cellular process, molecular interactions, and the pathogenesis of complex diseases at the molecular level by integrating different types of data and molecular interaction networks [31]. Such research includes the prediction of gene–disease associations [32], and the prediction of potential disease-associated miRNAs [33,34]. An increasing number of researchers have also adopted various data frameworks to increase the reliability of association prediction between diseases and lncRNAs. Hence, a third kind of prediction models has been proposed, in which multiple data sources are integrated to identify disease-related lncRNAs. For example, Lu et al. proposed a new prediction of lncRNA–disease associations via inductive matrix completion (named SIMCLDA), by integrating known lncRNA–disease interactions, disease–gene, gene–gene ontology associations [35]. Zhang et al. developed a novel model named LncRDNetFlow, which utilized a flow propagation algorithm to integrate a variety of information including the similarity of lncRNAs, the protein–protein interactions, and the similarity of diseases to infer lncRNA–disease associations [36]. Fu et al. proposed a

model called MFLDA to predict potential lncRNA–disease associations by considering the quality and relevance of different heterogeneous data sources, which can select and integrate the data sources by assigning different weights to them [37]. Chen developed a path-based approach named KATZLDA for discovering potential lncRNA–disease associations by integrating information including known lncRNA–disease associations, lncRNA expression profiles, lncRNA functional similarity, disease semantic similarity, and Gaussian interaction profile kernel similarity [38]. All of these above data fusion-based methods can achieve effective results.

In this paper, to effectively predict potential lncRNA–disease associations, we first constructed a global tripartite network by integrating three kinds of heterogeneous networks including an lncRNA–disease association network, an miRNA–disease association network, and an miRNA–lncRNA interaction network. Then, considering that more heterogeneous networks can boost the prediction performance, we constructed a quadruple global network by appending a gene–lncRNA interaction network, a gene–disease association network, and a gene–miRNA interaction network to the tripartite network. Thereafter, based on these two newly constructed global networks, a novel probabilistic model named Naïve Bayesian Classifier used to predict potential lncRNA–Disease Associations (NBCLDA), based on the naïve Bayesian classifier, is proposed to uncover potential lncRNA–disease associations. Moreover, in order to evaluate the prediction performance of the NBCLDA, the leave-one-out cross-validation (LOOCV) framework was implemented, and the experimental results demonstrated the effective performance of the NBCLDA and illustrated that it can achieve better predictive performance than state-of-the-art methods in the terms of LOOCV.

2. Data Collection and Preprocessing

Considering that more heterogeneous data sources can boost the performance of prediction models, in this paper, to construct our novel prediction model NBCLDA—with the ultimate goal being to infer potential associations between lncRNAs and diseases—seven heterogeneous data sets were combined. These include the sets of miRNA–disease, miRNA–lncRNA, lncRNA–disease, gene–disease, and gene–lncRNA associations, as well as the sets of gene–miRNA interactions, and of diseases with disease tree numbers. The sets were collected from various databases.

2.1. Construction of miRNA–Disease and miRNA–lncRNA Association Sets

In this article, the miRNA–disease and miRNA–lncRNA association sets were downloaded from the HMDD [39] and the starBase v2.0 [40] databases in January 2015. Once these two data sets were collected, we removed any duplicate associations with conflicting evidence. Then, we further unified the names of miRNAs, and, thereafter, manually selected the common miRNAs in both sets. Finally, we retained only the associations related with those selected miRNAs in these two data sets. As a result, we obtained a data set DS_1 consisting of 4704 miRNA–disease interactions between 246 miRNAs and 373 diseases, and a data set DS_2 consisting of 9086 miRNA–lncRNA interactions between 246 miRNAs and 1089 lncRNAs (see Supplementary Materials Tables S1 and S2).

2.2. Construction of the lncRNA–Disease Association Set

In this paper, the set of lncRNA–disease associations was collected from the MNDR v2.0 database [41] in 2017. In a similar way, once the data set was collected, we removed the duplicate associations with conflicting evidence. Then, we selected the lncRNA–disease associations with diseases belonging to DS_1 and lncRNAs belonging to DS_2 simultaneously. As a result, we obtained a data set DS_3 consisting of 407 lncRNA–disease associations between 77 lncRNAs and 95 diseases (see Supplementary Materials Table S3). The data set DS_3 is utilized as the test sample in our following simulation experiments.

2.3. Construction of the Gene–Disease and Gene–lncRNA Association Sets

In this article, the set of gene–disease associations was gathered from the DisGeNET v5.0 database [42] in May 2017, and the set of gene–lncRNA associations was downloaded from the LncACTdb v1.0 database [43]. Again, we removed the duplicate associations with conflicting evidence. Then, we further unified the names of genes, and thereafter manually selected the common genes in both sets. Finally, we retained only the associations related with those selected genes in these two data sets. Additionally, we transformed some disease names included in the newly constructed set of gene–disease associations into their aliases in the DS_1 , in order to keep the uniformity of disease names. For example, the disease names “pulmonary Emphysema” and “Bladder Neoplasm” in the newly collected set of gene–disease associations was converted into “pulmonary Embolism” and “Bladder Neoplasms” in the DS_1 , respectively. Hence, we obtained a data set DS_4 consisting of 3702 gene–disease associations between 171 genes and 227 diseases, and a data set DS_5 consisting of 411 gene–lncRNA interactions between 171 genes and 66 lncRNAs (see Supplementary Materials Tables S4 and S5).

2.4. Construction of the Gene–miRNA Association Set

In this paper, the set of gene–miRNA interactions was obtained from the miRecords [44] database that was last updated in April 2013. Once the data set was collected, we removed the duplicate associations with conflicting evidence. Then, we selected the gene–miRNA interactions with genes belonging to DS_4 or DS_5 and miRNAs belonging to DS_1 or DS_2 , simultaneously. Finally, as a result, we obtained a data set DS_6 consisting of 565 gene–miRNA associations between 109 genes and 174 miRNAs (see Supplementary Materials Table S6).

2.5. Construction of the Set of Diseases with Disease Tree Numbers

In this article, the set of diseases with Disease tree numbers was gathered from the MeSH database [45]. In the MeSH database, the disease terms, described as DAGs, were classified and signified as disease tree numbers. We browsed the MeSH database and collected the disease tree numbers of diseases in DS_1 . As a result, we obtained a data set DS_7 consisting of 373 diseases with their disease tree numbers (see Supplementary Materials Table S7).

2.6. Analysis of Multi Relational Data Sources

In our model, four object types such as lncRNA, diseases, miRNA, and genes are considered. Based on these four object types, we collect six relational data sources from different databases. Figure 1 is constructed to illustrate the relationship between these different data sources more directly. In Figure 1, $R_{\#_1\Omega-\#_2\Omega}$ denotes the different associations between these four object types, where $\#_1$ represents one object, $\#_2$ represents another object and Ω denotes the dataset DS_Ω that the two objects belong to. For example, R_{m1-d1} denotes the associations between miRNAs and diseases, m represents miRNAs, d represents diseases, and ‘1’ indicates all these miRNAs and diseases belong to the dataset DS_1 . In addition, the numbers of the same objects in the different datasets and the relationships among them are shown in Figure 1. For instance, the number of diseases is 373 in R_{m1-d1} , 95 (= 29 + 66) in R_{l3-d3} and 227 (= 66 + 161) in R_{g4-d4} , and it is obvious that both the 95 diseases in R_{l3-d3} and the 227 diseases in R_{g4-d4} are part of the 373 diseases in R_{m1-d1} ; moreover, the intersect of disease in R_{l3-d3} and R_{g4-d4} includes 66 different diseases.

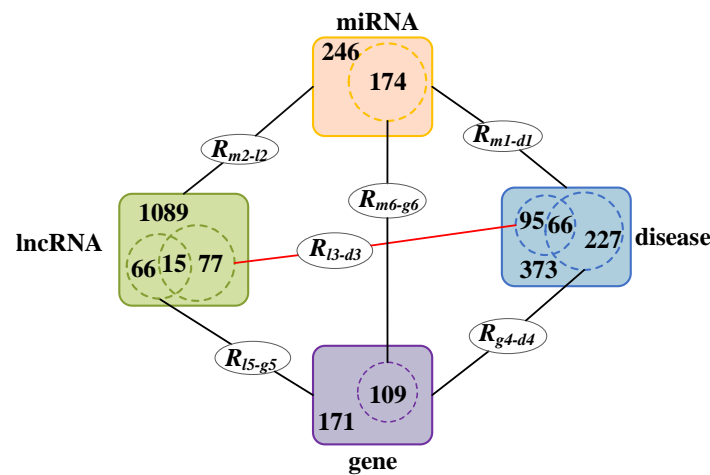


Figure 1. The relationship between the different data sources and number of data points.

3. Method

As illustrated in Figure 2, our newly proposed model NBCLDA for predicting potential associations between lncRNAs and diseases can be mainly divided into the following steps:

Step 1: As illustrated in Figure 2a, on the basis of data sets DS_1 , DS_2 , and DS_3 we can construct an miRNA–disease association network labeled MDN, an miRNA–lncRNA association network labeled MLN, and an lncRNA–disease association network labeled LDN.

Step 2: As illustrated in Figure 2b, by integrating the three association networks constructed in Step 1, we can easily obtain a global tripartite network GN_1 of lncRNA–miRNA–disease relationships.

Step 3: As illustrated in Figure 2c, in order to utilize multiple data sources to improve the prediction performance, on the basis of data sets DS_4 , DS_5 , and DS_6 obtained above, we can also construct a gene–disease association network labeled GDN, a gene–lncRNA association network labeled GLN, and a gene–miRNA association network labeled GMN.

Step 4: As illustrated in Figure 2d, by appending the three association networks constructed in Step 3 to GN_1 constructed in Step 2, we can easily obtain a global quadruple network GN_2 of lncRNA–miRNA–gene–disease relations.

Step 5: As illustrated in Figure 2e,f, after applying the naïve Bayesian classifier theory to GN_1 and GN_2 , we can obtain two kinds of prediction models: NBCLDA- GN_1 and NBCLDA- GN_2 .

Step 6: As illustrated in Figure 2g,h, in order to further improve the prediction performance of the NBCLDA, we implemented disease semantic similarity in NBCLDA- GN_1 and NBCLDA- GN_2 . Thus, we can obtain two new prediction models, NBCLDA- GN_1 -SD and NBCLDA- GN_2 -SD, to infer potential lncRNA–disease associations.

3.1. Construction of the MDN, MLN, LDN, and GN_1

Let L be the set of n lncRNAs in DS_2 , L' be the set of n' lncRNAs in DS_3 , D be the set of r diseases in DS_1 , D' be the set of r' diseases in DS_3 . Additionally, let $M = \{m_1, m_2, \dots, m_t\}$ be the set of t miRNAs in DS_1 or DS_2 . From Sections 2.1 and 2.2, it is clear that $L' \subseteq L$ and $D' \subseteq D$; hence, we can let $L' = \{l_1, l_2, \dots, l_{n'}\}$, $L = \{l_1, l_2, \dots, l_{n'}, l_{n'+1}, \dots, l_n\}$, $D' = \{d_1, d_2, \dots, d_{r'}\}$, and $D = \{d_1, d_2, \dots, d_{r'}, d_{r'+1}, \dots, d_r\}$. Thus, we can represent the miRNA–disease association network, MDN, as $MDN = (M, D, E_1)$, where $E_1 = \{e_{m_k-d_j} | m_k \in M, d_j \in D\}$ denotes the set of known interactions between the miRNAs in M and the diseases in D . That is, the edge $e_{m_k-d_j} \in E_1 \Leftrightarrow m_k$ is associated with d_j .

In the same way, we can further represent the miRNA–lncRNA interaction network, MLN, and the lncRNA–disease association network, LDN, as $MDN = (M, L, E_2)$ and $LDN = (L, D, E_3)$,

where $E_2 = \{e_{m_k-l_i} | m_k \in M, l_i \in L\}$ denotes the set of known interactions between the miRNAs in M and the lncRNAs in L ; $E_3 = \{e_{l_i-d_j} | l_i \in L', d_j \in D'\}$ represents the set of interactions between the lncRNAs in L' and the diseases in D' . Thus, the edge $e_{m_k-l_i} \in E_2 \Leftrightarrow m_k$ is associated with l_i , and the edge $e_{l_i-d_j} \in E_3 \Leftrightarrow l_i$ is associated with d_j . Finally, the global tripartite network, GN_1 , is expressed as $GN_1 = (L, D, M, E)$, where $E = E_1 \cup E_2 \cup E_3$.

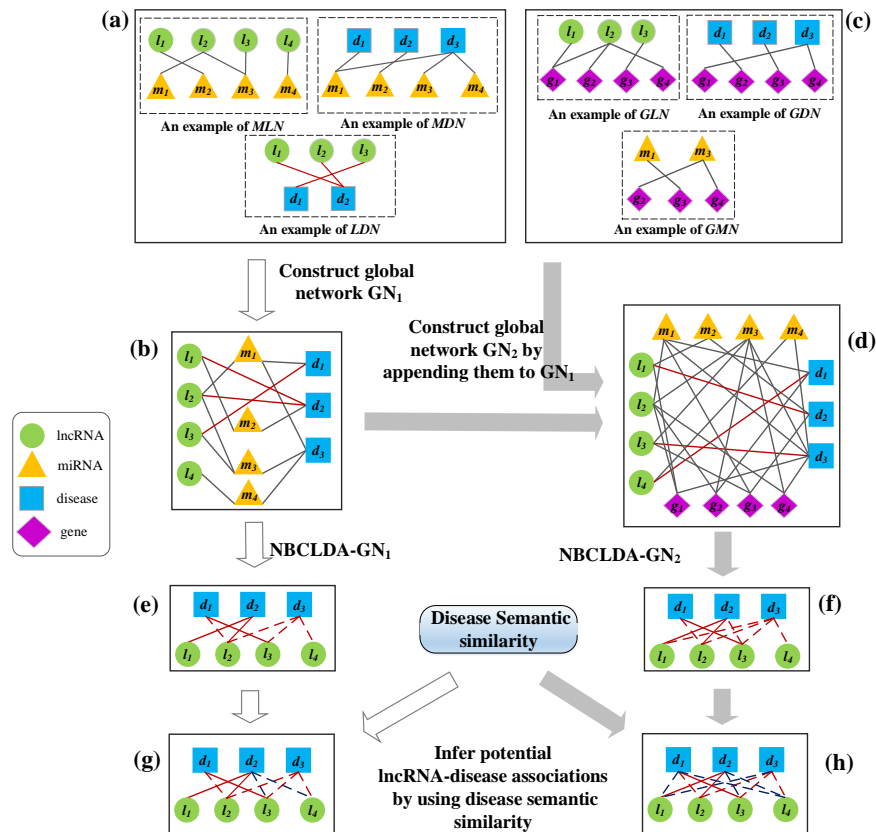


Figure 2. The flowchart of NBCLDA. In the diagram, the green circles, blue squares, orange triangles, and purple diamonds represent lncRNAs, diseases, miRNAs, and genes, respectively. (a) construction of the MDN, MLN, and LDN; (b) construction of global tripartite network GN_1 by integrating the MDN, MLN, and LDN; (c) construction of the GDN, GLN, and GMN; (d) construction of the global quadruple network GN_2 by appending the GDN, GLN, and GMN into GN_1 ; (e,f) construction of the potential lncRNA–disease association network by using the NBCLDA- GN_1 , and NBCLDA- GN_2 ; (g,h) inference of potential lncRNA–disease associations by using disease semantic similarity. Here, in (e–h), the known lncRNA–disease associations are represented as the solid edges, and the candidate lncRNA–disease associations are represented as dashed edges.

3.2. Construction of GDN, GLN, GMN, and GN_2

Let D'' be the set of r'' diseases in DS_4 , L'' be the set of n'' lncRNAs in DS_5 , G be the set of p genes in DS_4 or DS_5 , G' be the set of p' genes in DS_6 , and M' be the set of t' miRNAs in DS_6 . Additionally, from Sections 2.3 and 2.4, it is clear that $D'' \subseteq D$, $L'' \subseteq L$, and $G' \subseteq G$; hence, we can let $D' = \{d_1, d_2, \dots, d_{r''}\}$, $L' = \{l_1, l_2, \dots, l_{n''}\}$, $G' = \{g_1, g_2, \dots, g_{p'}\}$, $G = \{g_1, g_2, \dots, g_{p'}, g_{p'+1}, \dots, g_p\}$, and $M' = \{m_1, m_2, \dots, m_{t'}\}$. We can thus represent the gene–disease association network, GDN, as $GDN = (G, D, E_4)$, where $E_4 = \{e_{g_f-d_j} | g_f \in G, d_j \in D''\}$ denotes the set of known interactions between the genes in G and the diseases in D'' . That is, the edge $e_{g_f-d_j} \in E_4 \Leftrightarrow g_f$ is associated with d_j .

In the same way, we can further represent the gene–lncRNA interaction network, GLN, and gene–miRNA interaction network, GMN, as $GLN = (G, L, E_5)$ and $GMN = (G, M, E_6)$, where $E_5 = \{e_{g_f-l_i} | g_f \in G, l_i \in L''\}$ and $E_6 = \{e_{g_f-m_k} | g_f \in G', m_k \in M'\}$ denote the set of known gene–lncRNA interactions and the set of known gene–miRNA interactions, respectively. In other words, the edge $e_{g_f-l_i} \in E_5 \Leftrightarrow g_f$ is associated with l_i and the edge $e_{g_f-m_k} \in E_6 \Leftrightarrow g_f$ is associated with m_k . Finally, it is evident that the global tripartite network GN_2 can be expressed as $GN_2 = (L, D, M, G, E')$, where $E' = E \cup E_4 \cup E_5 \cup E_6$.

3.3. Construction of NBCLDA

The naïve Bayesian classifier is a simple probabilistic classifier with a naïve independence assumption that any feature of a class is independent of the other features of the class. Abstractly, based on the Bayesian classifier probability model $p(C|F_1, F_2, \dots, F_n)$, where C is a dependent class variable and F_1, F_2, \dots, F_n are the feature variables of class C , the posterior probability can be described as follows:

$$p(C|F_1, F_2, \dots, F_n) = \frac{p(F_1, F_2, \dots, F_n|C)p(C)}{p(F_1, F_2, \dots, F_n)}. \quad (1)$$

Furthermore, according to the above assumption, since each feature F_i is conditionally independent of every other feature F_j ($i \neq j$), Equation (1) can be expressed as:

$$p(C|F_1, F_2, \dots, F_n) = \frac{p(C) \prod_{i=1}^n p(F_i|C)}{p(F_1, F_2, \dots, F_n)}. \quad (2)$$

Inspired by existing probabilistic models based on Bayesian theory to predict missing links in complex networks [46], we designed a prediction model NBCLDA to infer potential disease-related lncRNAs; we applied the naïve Bayesian theory to GN_1 and GN_2 , constructed in Sections 3.1 and 3.2, respectively. In the context of Equation (1), in NBCLDA, the associations between lncRNAs and diseases in GN_1 and GN_2 are considered as the class of variables, while the common neighboring nodes of every lncRNA–disease pair in GN_1 and GN_2 are considered as the feature variables. In particular, when applying the naïve Bayesian theory to GN_1 , for any given pair of lncRNA and disease nodes in GN_1 , we will consider that their common neighboring miRNA nodes are all conditionally independent of each other, since all of the miRNAs are different, and, therefore, we assume that each of the miRNAs will not affect the others. To illustrate this assumption more intuitively, we provide an example in Figure 3a, in which the common neighboring nodes m_1 and m_3 between l_2 and d_3 will be assumed to be conditionally independent.

However, when applying the naïve Bayesian theory to GN_2 , as there are two types of common neighboring nodes, miRNAs and genes, between a pair of lncRNA and disease nodes. In this case, it is unreasonable to consider that all of these common neighbors are conditionally independent of each other, since there may exist interactions between genes and miRNAs. Therefore, for any given pair of lncRNA and disease nodes in GN_2 , let ϕ be the set that consists of all their common neighboring nodes. Then, for any miRNA node m^* , if there is a gene node g^* that is associated with m^* , we will consider the miRNA m^* and its related gene g^* as a whole, and denote them as m^*-g^* and label this an miRNA–gene pair. By this means, it is obvious that there will be three kinds of features in ϕ —miRNAs, genes, and miRNA–gene pairs. Hence, we assume that these three kinds of elements in ϕ are conditionally independent of each other. To illustrate this assumption more intuitively, we present an example in Figure 3b, in which, m_1, m_3, g_1 , and g_4 are the common neighboring nodes between l_2 and d_3 , and we will assume that m_3-g_4, m_1 , and g_1 are conditionally independent of each other.

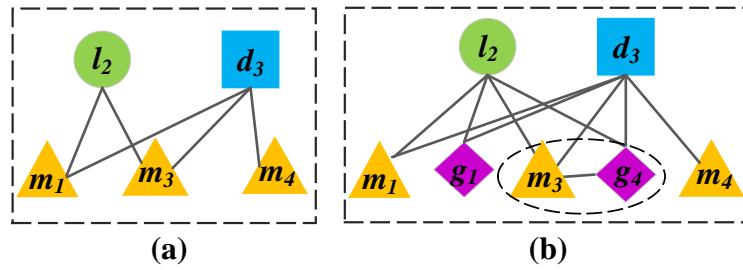


Figure 3. (a) a subnetwork of Figure 2b, in which, the common neighboring nodes m_1 and m_3 between l_2 and d_3 , are assumed to be conditionally independent; (b) a subnetwork of Figure 2d, in which, m_1 , m_3 , g_1 , and g_4 are the common neighboring nodes between l_2 and d_3 . Here, m_3 - g_4 , m_1 , and g_1 are assumed to be conditionally independent.

3.3.1. Method for Applying the Naïve Bayesian Theory into GN_1

For any given lncRNA node l_i and disease node d_j in GN_1 , let $N(l_i)$ and $N(d_j)$ be the sets of neighboring nodes that are directly connected to l_i and d_j , respectively. From this, we construct $CN(l_i, d_j) = \{m_1, m_2, \dots, m_h\}$, which denotes the set consisting of all common neighboring nodes between l_i and d_j in GN_1 . Then, the prior probabilities for the existence of an relationship edge $e_{l_i-d_j}$ are calculated via:

$$p(e_{l_i-d_j} = 1) = \frac{|M^c|}{|M|}, \tag{3}$$

$$p(e_{l_i-d_j} = 0) = 1 - p(e_{l_i-d_j} = 1), \tag{4}$$

where $|M^c|$ denotes the number of known associations between lncRNAs and diseases in LDN, and $|M| = n \times r$, where n denotes the number of lncRNAs in L and r denotes the number of diseases in D .

Based on the naïve Bayesian classifier, the posterior probabilities for an edge $e_{l_i-d_j}$, representing whether the node l_i is connected to d_j in GN_1 , are defined as follows:

$$p(e_{l_i-d_j} = 1|CN(l_i, d_j)) = \frac{p(e_{l_i-d_j} = 1)}{p(CN(l_i, d_j))} \prod_{m_\delta \in CN(l_i, d_j)} p(m_\delta|e_{l_i-d_j} = 1), \tag{5}$$

$$p(e_{l_i-d_j} = 0|CN(l_i, d_j)) = \frac{p(e_{l_i-d_j} = 0)}{p(CN(l_i, d_j))} \prod_{m_\delta \in CN(l_i, d_j)} p(m_\delta|e_{l_i-d_j} = 0). \tag{6}$$

From Equations (5) and (6), we can directly identify whether an lncRNA node is connected with a disease node or not in GN_1 . However, since it is often too complicated to calculate the value of $p(CN(l_i, d_j))$, we first define the probability of a potential association existing between l_i and d_j in GN_1 as follows:

$$S1(l_i, d_j) = \frac{p(e_{l_i-d_j} = 1)}{p(e_{l_i-d_j} = 0)} \prod_{m_\delta \in CN(l_i, d_j)} \frac{p(m_\delta|e_{l_i-d_j} = 1)}{p(m_\delta|e_{l_i-d_j} = 0)}, \tag{7}$$

where $p(m_\delta|e_{l_i-d_j}=1)$ and $p(m_\delta|e_{l_i-d_j}=0)$ are the conditional probabilities of a node m_δ belonging to $CN(l_i, d_j)$; they represent the possibilities of whether the node is a common neighboring node between l_i and d_j in GN_1 or not, respectively. Moreover, according to Bayesian theory, these two conditional probabilities can be expressed as:

$$p(m_\delta|e_{l_i-d_j} = 1) = \frac{p(e_{l_i-d_j} = 1|m_\delta)p(m_\delta)}{p(e_{l_i-d_j} = 1)}, \tag{8}$$

$$p(m_\delta | e_{l_i-d_j} = 0) = \frac{p(e_{l_i-d_j} = 0 | m_\delta) p(m_\delta)}{p(e_{l_i-d_j} = 0)}, \tag{9}$$

where $p(e_{l_i-d_j}=1|m_\delta)$ and $p(e_{l_i-d_j}=0|m_\delta)$ represent the conditional probability of whether the lncRNA node l_i is connected to the disease node d_j or not, respectively, and m_δ is one of the common neighboring nodes between l_i and d_j in GN_1 . Thus, $p(e_{l_i-d_j}=1|m_\delta)$ and $p(e_{l_i-d_j}=0|m_\delta)$ are calculated via the following formulas:

$$p(e_{l_i-d_j} = 1 | m_\delta) = \frac{N_{m_\delta}^+}{N_{m_\delta}^+ + N_{m_\delta}^-}, \tag{10}$$

$$p(e_{l_i-d_j} = 0 | m_\delta) = \frac{N_{m_\delta}^-}{N_{m_\delta}^+ + N_{m_\delta}^-}, \tag{11}$$

where $N_{m_\delta}^+$ and $N_{m_\delta}^-$ denote the number of known and unknown associations between lncRNAs and diseases whose common neighbors include m_δ , respectively.

Hence, from Equations (8) and (9), Equation (7) can be modified as follows:

$$S1(l_i, d_j) = \frac{p(e_{l_i-d_j} = 1)}{p(e_{l_i-d_j} = 0)} \prod_{m_\delta \in CN(l_i, d_j)} \frac{p(e_{l_i-d_j} = 0) p(e_{l_i-d_j} = 1 | m_\delta)}{p(e_{l_i-d_j} = 1) p(e_{l_i-d_j} = 0 | m_\delta)}. \tag{12}$$

Moreover, given any two nodes l_i and d_j in GN_1 , the value of $\frac{p(e_{l_i-d_j}=1)}{p(e_{l_i-d_j}=0)}$ is a constant, which we denote as ϕ_m for convenience. Additionally, for each common neighboring node between l_i and d_j in GN_1 , let N_l denote the number of lncRNAs directly related to m_δ , and N_d denote the number of diseases directly related to m_δ . Then, $N_{m_\delta}^+ + N_{m_\delta}^- = N_l \times N_d$, and hence, Equation (7) can further be modified as follows:

$$S1(l_i, d_j) = \phi_m \prod_{m_\delta \in CN(l_i, d_j)} \phi_m^{-1} \frac{N_{m_\delta}^+}{N_{m_\delta}^-}. \tag{13}$$

Considering that $N_{m_\delta}^+$ may equal zero, we will introduce the Laplace calibration to guarantee that the value of $S1(l_i, d_j)$ will not be zero:

$$S1(l_i, d_j) = \phi_m \prod_{m_\delta \in CN(l_i, d_j)} \phi_m^{-1} \frac{N_{m_\delta}^+ + 1}{N_{m_\delta}^- + 1}. \tag{14}$$

Furthermore, by introducing the logarithmic function for standardization, for any given lncRNA node l_i and disease node d_j in GN_1 , we can finally define the probability of a potential association existing between them as:

$$S1'(l_i, d_j) = \frac{\log(S1(l_i, d_j))}{\lambda}, \tag{15}$$

where λ is a constant utilized for normalization.

3.3.2. Method for Applying the Naïve Bayesian Theory to GN_2

In the same manner as described in Section 3.3.1, for any given lncRNA node l_i and disease node d_j in GN_2 , we construct the set consisting of all common neighboring nodes, $CN'(l_i, d_j) = \{m_1, m_2, \dots, m_h, g_1, g_2, \dots, g_u\}$. Then, the posterior probabilities of $p'(e_{l_i-d_j}=1|CN'(l_i, d_j))$ and $p'(e_{l_i-d_j}=0|CN'(l_i, d_j))$, representing whether the node l_i is connected to d_j in GN_2 or not, respectively. Then, similarly as described in Section 3.3.1, we can define the probability of a potential

association existing between l_i and d_j in GN_2 as follows (the deep representation of scheme are described in Supplementary Material):

$$S2(l_i, d_j) = \phi_m \prod_{m_\alpha \in CN'(l_i, d_j)} \prod_{g_\beta \in CN'(l_i, d_j)} \prod_{m_{\bar{\alpha}}, g_{\bar{\beta}} \in CN'(l_i, d_j)} \phi_m^{-3} \frac{(N_{m_\alpha}^+ + 1)(N_{g_\beta}^+ + 1)(N_{m_{\bar{\alpha}}, g_{\bar{\beta}}}^+ + 1)}{(N_{m_\alpha}^- + 1)(N_{g_\beta}^- + 1)(N_{m_{\bar{\alpha}}, g_{\bar{\beta}}}^- + 1)}, \quad (16)$$

where $N_{m_{\bar{\alpha}}, g_{\bar{\beta}}}^+$ and $N_{m_{\bar{\alpha}}, g_{\bar{\beta}}}^-$ denote the number of known and unknown associations between l_i and d_j in GN_2 , respectively, conditional on $m_{\bar{\alpha}}$ and $g_{\bar{\beta}}$ being common neighboring nodes between l_i and d_j in GN_2 and $m_{\bar{\alpha}}-g_{\bar{\beta}}$ is an miRNA–gene pair. In addition, $N_{m_\alpha}^+$ and $N_{m_\alpha}^-$ denote the number of known and unknown associations between l_i and d_j in GN_2 , respectively, conditional on m_α being a common neighboring node between l_i and d_j . In addition, $N_{g_\beta}^+$ and $N_{g_\beta}^-$ represent the number of known and unknown associations between l_i and d_j in GN_2 , respectively, conditional on g_β being a common neighboring node between l_i and d_j . Finally, following the example of Equation (15), we can finally define the probability of a potential association existing between l_i and d_j in GN_2 as follows:

$$S2'(l_i, d_j) = \frac{\log(S2(l_i, d_j))}{\lambda}. \quad (17)$$

3.3.3. Method of Appending the Disease Semantic Similarity into NBCLDA

The disease semantic similarity has been widely utilized as a valuable data source for discovering potential disease-related lncRNAs in many previous studies [30,38]. In this paper, we append the disease semantic similarity into our newly constructed prediction model NBCLDA to further uncover the potential relationships between lncRNAs and diseases.

From the description given in Section 2.5, we know that each disease term in the MeSH database can be described as a directed acyclic graph (DAG), in which the nodes represent the disease MeSH descriptors and all MeSH descriptors in the DAG are linked from more general terms (parent nodes) to more specific terms (child nodes) by a direct edge. Hence, in this paper, we first obtain the disease tree numbers according to the disease terms collected from the MeSH database. Thereafter, adopting the method proposed by Wang et al. [47], while supposing that disease d_j is represented as the graph $DAG_{d_j} = (d_j, T_{d_j}, E_{d_j})$, where T_{d_j} is the set of all ancestor nodes of d_j including node d_j , E_{d_j} is the set of corresponding links, and the contribution of a disease t in DAG_{d_j} to the semantic of disease d_j can be calculated as follows:

$$D_{d_j}(t) = \begin{cases} 1, & \text{if } t = d_j, \\ \max\{\Delta \times D_{d_j}(ct) | ct \in \text{children of } t\}, & \text{if } t \neq d_j, \end{cases} \quad (18)$$

where Δ is the semantic contribution factor for edges E_{d_j} linking disease d_j with child disease t and the disease d_j is the most specific disease and its own semantic score is defined as 1. Since nodes located farther from d_j will be more general diseases that contribute less to d_j , then, based on Equation (24), we can define the semantic value of the disease d_j as follows:

$$DV(d_j) = \sum_{t \in T_{d_j}} D_{d_j}(t). \quad (19)$$

Therefore, based on the assumption that the diseases share the nodes of their DAGs, the semantic similarity between disease d_j and d_i can be defined as:

$$SD(d_j, d_i) = \frac{\sum_{t \in T_{d_j} \cap T_{d_i}} (D_{d_j}(t) + D_{d_i}(t))}{DV(d_j) + DV(d_i)}. \quad (20)$$

Finally, based on the disease semantic similarity and the similarities between lncRNAs and diseases, we can reconstruct a new recommended measurement for inferring potential associations between lncRNAs and diseases as follows:

$$S = S' \times SD, \quad (21)$$

where S' denotes either $S1'(l_i, d_j)$ or $S2'(l_i, d_j)$ and SD , which is computed via Equation (20) denotes the disease semantic similarity.

4. Results

4.1. Performance Evaluation

The performance of the NBCLDA, for inferring potential associations between lncRNAs and diseases, is evaluated by implementing LOOCV and is based on experimentally verified lncRNA–disease associations. At each round, a known lncRNA–disease association is used as a test sample, whereas all the remaining associations are taken as training cases for model learning. This step continues until each sample is treated as a verification sample. Moreover, the value of area under the receiver operating characteristic (ROC) curve (AUC) can be applied for measuring the overall performance of the method. The closer the AUC value is to 1, the better the performance is, and an AUC value of 0.5 refers to a random guess. We calculate a series of true positive rates (TPR or sensitivity) and false positive rates (FPR or 1–specificity) by setting different classification thresholds, and the ROC curve is plotted with the functional relationship between them. Specifically, TPR corresponds to the ratio of the successfully predicted lncRNA–disease associations to the total experimentally verified lncRNA–disease associations, and FPR refers to the percentage of candidate lncRNAs ranked below the threshold.

First, in order to estimate the influence of the addition of new types of nodes and the introduction of the disease semantic similarity on the predictions of potential associations between lncRNAs and diseases, we implemented the NBCLDA on the two constructed global networks GN_1 and GN_2 in the framework of LOOCV. The simulation results are shown in Figures 4 and 5. From Figure 4, the NBCLDA achieved an AUC of 0.8240 on GN_1 and an AUC of 0.8604 on GN_2 when the disease semantic similarity was not utilized. On the other hand, from Figure 5, an AUC of 0.8519 on GN_1 and an AUC of 0.8819 on GN_2 were achieved when the disease semantic similarity was included. This demonstrates that the prediction performance of our method not only benefits from the addition of the new types of nodes for predicting potential associations between lncRNAs and diseases, but also is significantly improved by the introduction of disease semantic similarity.

In order to further assess the performance of the NBCLDA, we compared it with other state-of-the-art models including HGLDA [27], SIMCLDA [35], MFLDA [37], Yang et al. method [26], KATZLDA [38] and TPGLDA [26] in the framework of LOOCV. For comparing with the HGLDA, a data set consisting of 183 experimentally validated lncRNA–disease associations was previously constructed and taken as the test set to evaluate its performance. Hence, for convenience, we compared our model, the NBCLDA, with the HGLDA on that data set using the framework of LOOCV. The simulation results are illustrated in Table 1 and Figure 6, from which it is evident that our approach outperformed the HGLDA. For comparing with SIMCLDA, a data set consisting of 101 known lncRNA–disease associations between 30 lncRNAs and 79 diseases was collected from the data set containing of 293 experimentally validated lncRNA–disease associations which was used in method SIMCLDA. These selected lncRNAs and diseases all belong to DS_3 in our paper. The simulation results are illustrated in Table 1, from which it is evident that our approach outperformed the SIMCLDA. While comparing with MFLDA, six relational data sources including lncRNA–miRNA associations, lncRNA–gene function associations, lncRNA–disease associations, miRNA–gene interactions, miRNA–disease associations and gene–disease associations, which were used in the method MFLDA, were collected to implement NBCLDA. The data set of experimentally

validated lncRNA–disease associations was taken as the test set to evaluate its performance. The simulation results are illustrated in Table 1, from which it is evident that our approach outperformed the MFLDA.

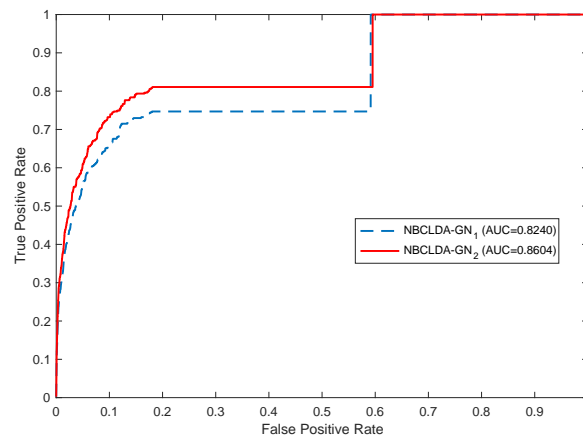


Figure 4. Performance evaluation for the NBCLDA in terms of ROC curves and AUCs based on the experimentally known associations (data set DS_3), in the framework of LOOCV. Here, NBCLDA- GN_1 and NBCLDA- GN_2 represent the simulation results while implementing our algorithm on the global networks GN_1 and GN_2 , respectively.

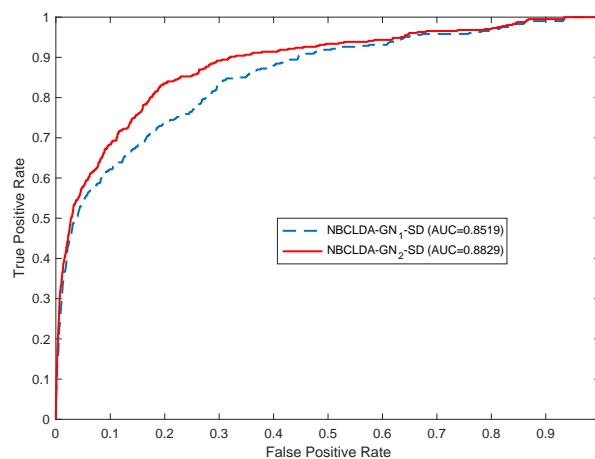


Figure 5. Same as Figure 4, but additionally including disease semantic similarity. Here, NBCLDA- GN_1 -SD and NBCLDA- GN_2 -SD represent the simulation results when appending the disease semantic similarity to the NBCLDA on networks GN_1 and GN_2 , respectively.

Table 1. Performance comparisons between the NBCLDA and other state-of-the-art models in terms of AUCs based on the different data sets of known lncRNA–disease associations in the framework of the LOOCV.

Methods	AUCs	Methods	AUCs
NBCLDA- GN_2 -SD	0.8982	NBCLDA- GN_2 -SD	0.9169
HGLDA	0.7621	Yang et al. method	0.8568
NBCLDA- GN_2 -SD	0.8897	NBCLDA- GN_2 -SD	0.8829
SIMCLDA	0.8526	KATZLDA	0.8283
NBCLDA- GN_2 -SD	0.8704	NBCLDA- GN_2 -SD	0.8897
MFLDA	0.7945	TPGLDA	0.92

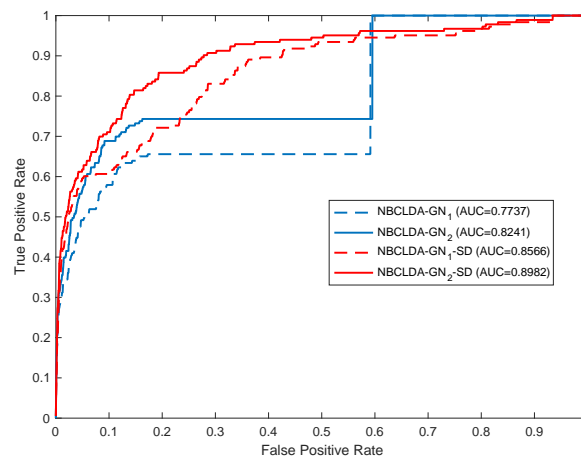


Figure 6. The performance of the NBCLDA in terms of ROC curves and AUCs based on 183 known lncRNA–disease associations, in the framework of the LOOCV.

Furthermore, we compared the NBCLDA with Yang et al.’s method based on the data set DS₃ consisting of 407 lncRNA–disease associations between 77 lncRNAs and 95 diseases. In order to make a comparison with Yang et al.’s method, according to their description, we first deleted the nodes with a degree equal to 1. As a result, we obtained a data set consisting of 319 lncRNA–disease associations between 37 lncRNAs and 52 diseases. Then, we took this data set as the test set to compare the two methods in the framework of the LOOCV. The simulation results are shown in Figure 7, from which it is seen that the NBCLDA achieved an AUC of 0.9169 while being implemented on GN₂, which is much better than the AUC of 0.8568 achieved by Yang et al.’s method. We also compared the NBCLDA with the KATZLDA, which is a path-based method designed to predict potential lncRNA–disease associations by integrating multiple pieces of information including known lncRNA–disease associations, lncRNA expression profiles, lncRNA functional similarity, disease semantic similarity, and the Gaussian interaction profile kernel similarity. Executing the simulation, we could not obtain information on the expression profiles of corresponding lncRNAs; thus, we compared the two methods without this information. The simulation results are shown in Figure 8, which indicate that the NBCLDA achieves higher AUCs (of 0.8519 and 0.8829) than the KATZLDA with a corresponding AUC of 0.8323. This also demonstrates the superiority of our newly constructed prediction model, the NBCLDA. Finally, comparing with TPGLDA, a data set consisting of 312 experimentally validated lncRNA–disease associations including 68 lncRNAs and 67 diseases and a data set consisting of 1941 gene–disease associations between 165 genes and 67 diseases were constructed, respectively. The data set of known lncRNA–disease associations was taken as the test set to evaluate its performance. The simulation results are illustrated in Table 1, from which it is obvious that TPGLDA can achieve a better performance with an AUC of 0.92, which is higher than that of ours with the AUC value of 0.8982. The main reason that TPGLDA can achieve a better performance is probably that the contribution of resource moved in both directions are taken into consideration by a consistence-based resource allocation algorithm. However, NBCLDA does not entirely rely on known lncRNA–disease associations and can integrate multiple data sources to predict potential associations.

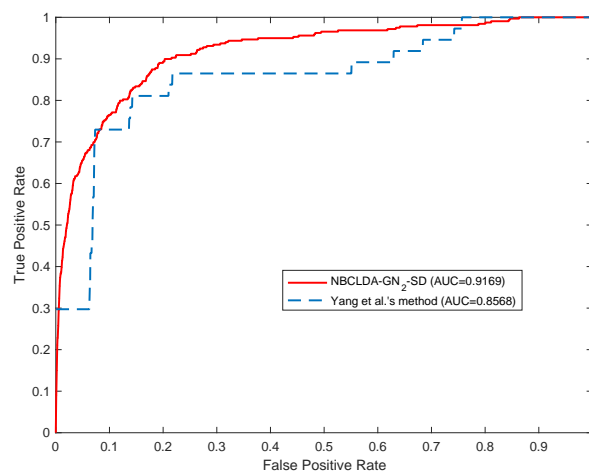


Figure 7. Comparison of the performance of the NBCLDA and Yang et al.'s method in terms of ROC curves and AUCs based on a data set of 319 lncRNA–disease associations between 37 lncRNAs and 52 diseases in the framework of the LOOCV.

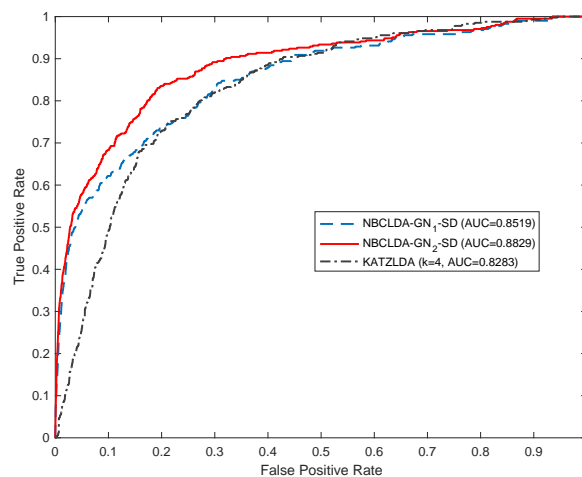


Figure 8. Comparison of the performance of the NBCLDA and KATZLDA approaches in terms of ROC curves and AUCs based on data set DS_3 , in the framework of the LOOCV.

In order to further evaluate the performance of NBCLDA, 20 percent of the known lncRNA–disease associations are randomly chosen as training set, while the remaining known and all the unknown associations are taken as testing set. We then compare with the six methods on the predicted top- k associations by using F1-score measure, which is a measure of a test's accuracy [48]. Since the sparse known lncRNA–disease associations, we set different threshold k based on the different set of known associations when comparing with other methods and the comparison results are illustrated in Table 2. From Table 2, we could see that NBCLDA outperforms several other methods in terms of F1-score. However, TPGLDA could achieve higher values than that of our approach, this is likely due to that resource moved in both directions are taken into consideration by consistence-based resource allocation algorithm. However, comparing with TPGLDA, our new method does not entirely rely on known lncRNA–disease associations and can integrate multiple data sources to predict potential associations. These advantages may be an excellent addition for biomedical research in the future.

Table 2. F1-scores of NBCLDA, SIMCLDA, MFLDA, Yang et al.'s method, KATZLDA, TPGLDA at different top-*k* cutoffs

Methods	F1-Score		
NBCLDA	0.1536 (k = 20)	0.1582 (k = 40)	null (k = 60)
SIMCLDA	0.0635 (k = 20)	0.0482 (k = 40)	null (k = 60)
NBCLDA	0.1773 (k = 20)	0.2415 (k = 40)	null (k = 60)
MFLDA	0.2012 (k = 20)	0.1139 (k = 40)	null (k = 60)
NBCLDA	0.2575 (k = 20)	0.2855 (k = 34)	null (k = 60)
Yang et al.'s method	0.2707 (k = 20)	0.2769 (k = 34)	null (k = 60)
NBCLDA	0.1183 (k = 20)	0.1088 (k = 40)	0.1139 (k = 60)
KATZLDA	0.1274 (k = 20)	0.0869 (k = 40)	0.0779 (k = 60)
NBCLDA	0.1295 (k = 20)	0.1510 (k = 40)	0.1320 (k = 60)
TPGLDA	0.2070 (k = 20)	0.1644 (k = 40)	0.1301 (k = 60)

4.2. Case Studies

To further estimate the performance of the NBCLDA, case studies of three types of lncRNA-related diseases—colorectal cancer, prostate cancer, and glioma—are analyzed in this section. During the simulation experiment, the known lncRNA–disease associations in the data set DS_3 were considered as the training samples, while the experimentally validated lncRNA–disease associations beyond DS_3 were used for testing. As for the simulation results, the top 20 disease-related lncRNAs, predicted by the NBCLDA, were verified via relevant literature, and the corresponding evidence is listed in Table 3. In addition, the predicted results of the top 20 disease-related lncRNAs were presented in the Supplementary Table S8.

Colorectal cancer (CRC) is one of the most common cancer types in western countries and its morbidity increases with age [49]. Accumulating studies have shown that lncRNAs play important roles in several steps of carcinogenesis and cancer metastasis and additionally interact with various cancers including CRC [50,51]. Therefore, we implemented the NBCLDA to discover possible CRC-associated lncRNAs. As illustrated in Table 3, seven of the top 20 lncRNAs have been validated to be related to colorectal cancer by recent biological literature, and five of them are ranked in the top 10 of the prioritized prediction results. The other two are lncRNAs SNHG16 (ranked 12th) and TUG1 (ranked 18th). For example, Chen et al. indicated that the lncRNA XIST can regulate the process of CRC development by competing for miR-200b-3p and thus it may be considered as a biomarker for prognosis [52]. Additionally, it has been demonstrated that the lncRNA MALAT1 may be considered as a potential prognostic and therapeutic target of colorectal cancer patients as it can fulfill a chemoresistant function in colorectal cancer [53]. Nakano et al. found that the epigenetic destruction and loss of imprinting of the lncRNA KCNQ1OT1 play a significant role in the occurrence of colorectal cancer [54]. Han et al. suggested that H19 can be considered as a candidate therapeutic biomarker and a new target for human CRC therapy when it is used as a growth regulator [55].

Prostate cancer is the second most common cause of cancer-related mortality in males worldwide [56]. Increasing studies show that lncRNA have become a promising target for the treatment of cancers including prostate cancer [57,58]. Hence, we carried out the NBCLDA to uncover possible prostate cancer-associated lncRNAs, and five of the top 20 predicted lncRNAs were verified and are listed in Table 3 according to the relevant literature. For example, Ren et al. evaluated the expression of MALAT1 in prostate cancer and showed that it may be considered as a perspective therapeutic target for refractory prostate cancer [59]. Zhu et al. found that the lncRNA H19 and its derived miRNA H19-miR-675 were significantly downregulated in advanced prostate cancer and they may be used for diagnostic and therapeutic treatment in advanced prostate cancer because H19-miR-675 could act as a suppressor of prostate cancer metastasis [60]. Additionally, Tian et al. showed that targeting the

lncRNA NEAT1 axis could be used as a potential application in improving chemotherapy of prostate cancer [61].

Table 3. The lncRNAs in the top 20 for the three case studies.

Disease	lncRNA	Evidence (PMID)	Rank
Colorectal cancer	XIST	17143621	1
Colorectal cancer	MALAT1	25446987,25031737,21503572,25025966,24244343,26887056	3
Colorectal cancer	KCNQ1OT1	16965397	6
Colorectal cancer	H19	11120891,19926638,22427002,26068968,26989025	8
Colorectal cancer	NEAT1	26314847	9
Colorectal cancer	SNHG16	24519959	12
Colorectal cancer	TUG1	26856330	18
Prostate cancer	MALAT1	23845456,23726266,26516927,22349460	3
Prostate cancer	KCNQ1OT1	23728290	6
Prostate cancer	H19	24063685,24988946	8
Prostate cancer	NEAT1	23728290,25415230	10
Prostate cancer	TUG1	26975529	19
Glioma	MALAT1	26649278,25613066,26619802,27134488,26938295	4
Glioma	H19	24466011,26983719	6
Glioma	TUG1	25645334,27363339	10
Glioma	NEAT1	26582084	12

Glioma is one of the most common malignant forms of brain tumors, and 6 out of 100,000 people may have gliomas [62]. Accumulating research has shown that lncRNAs play a significant role in the process of glioma development [63]. Therefore, we applied the NBCLDA to predict potential lncRNAs associated with glioma. Four of the top 20 glioma-related lncRNAs were validated by recent literature on biological experiments, and the results are illustrated in Table 3. For example, the lncRNA MALAT1 plays an important role in the progression and therapy of glioma and it may be considered an effective prognostic biomarker for the treatment of glioma [64]. Zhang et al. demonstrated that the lncRNA H19 was overexpressed in glioma tissue and cell lines, and also promotes cell proliferation of glioma [65]. Furthermore, Li et al. suggested that the lncRNA TUG1 can promote cell apoptosis of glioma cells and may act as a tumor suppressor in human glioma [66].

5. Discussion

Accumulating studies have indicated that lncRNAs play crucial roles in biological processes, complex disease diagnoses, prognoses, and treatments. Furthermore, computational models for predicting novel lncRNA–disease associations by integrating varieties of biological data are among the most noticeable topics. This is helpful to explore the understanding of disease mechanisms at the lncRNA level. In this paper, we construct a global tripartite network and a quadruple network by integrating various biological information and propose a novel approach, the NBCLDA, to predict potential lncRNA–disease associations by applying the naïve Bayesian classifier into the two constructed networks. Compared with current models, the NBCLDA does not entirely rely on known lncRNA–disease associations, and can achieve a reliable performance with effective AUCs in the LOOCV framework. This means that our method can not only predict the possible associations between lncRNAs and diseases included in the known associations set, but can also predict the potential associations whose elements are not in the known data set.

To evaluate the predictive performance of our method, the LOOCV is implemented based on the experimentally verified lncRNA–disease associations obtained from the MNDR database. Simulation experiment results of the NBCLDA show a strong performance and its predictive accuracy has been significantly improved by the addition of new types of nodes and the disease semantic similarity for predicting potential associations between lncRNAs and diseases. It also shows that

the NBCLDA can achieve better performance than the other three state-of-the-art models with more effective AUCs in the framework of the LOOCV. Moreover, in order to further estimate the performance of the NBCLDA, case studies of colorectal cancer, prostate cancer, and glioma were implemented in this paper. These simulation results demonstrated that the NBCLDAs can be an excellent tool for future biomedical research.

Despite the reliable experimental results of the NBCLDA, there are also some biases in our method. For example, the known experimentally validated lncRNA–disease associations are still limited. Therefore, the prediction performance of the NBCLDA would be improved by a more comprehensive data set. Furthermore, the data sources in this paper need to be strictly preprocessed according to the proposed method, which restricts the richness of the data sources to a certain extent.

6. Conclusions

In this paper, we mainly summed up the following contributions: (1) we constructed a global tripartite network by integrating a variety of biological information including miRNA–disease, miRNA–lncRNA and lncRNA–diseases associations and interactions; (2) we constructed a global quadruple network by appending gene–lncRNA interaction, gene–disease association, and gene–miRNA interaction networks to the global tripartite network; (3) we developed a novel approach NBCLDA based on the naïve Bayesian classifier and applied it into the two global networks to predict potential lncRNA–disease associations; (4) we appended the disease semantic similarity into our newly constructed prediction model NBCLDA to further uncover the potential relationships between lncRNAs and diseases; (5) NBCLDA can not only predict the possible associations between lncRNAs and diseases included in the known associations set, but can also predict the potential associations whose elements are not in the known data set; (6) NBCLDA can integrate multiple heterogeneous biological data for discovering potential relationships between lncRNAs and diseases; (7) in the future work, more biological data can be collected and pre-processed to be utilized in the newly proposed method for predicting potential lncRNA–disease associations.

Supplementary Materials: The following are available at www.mdpi.com/xxx/s1, Supplementary Table S1: The known miRNA–disease associations of the data set DS₁ consisting of 4704 miRNA–disease interactions which were collected from the HMDD database; Supplementary Table S2: The known miRNA–lncRNA associations of the data set DS₂ consisting of 9086 miRNA–lncRNA interactions which were collected from the starBase v2.0 database; Supplementary Table S3: The known lncRNA–disease associations of the data set DS₃ consisting of 407 lncRNA–disease associations which were downloaded from the MNDR v2.0 database; Supplementary Table S4: The known gene–disease associations of the data set DS₄ consisting of 3702 gene–disease associations which were gathered from the DisGeNET v5.0 database; Supplementary Table S5: The known gene–lncRNA associations of the data set DS₅ consisting of 411 gene–lncRNA interactions which were downloaded from the LncACTdb database; Supplementary Table S6: The known gene–miRNA associations of the data set DS₆ consisting of 565 gene–miRNA association was obtained from the miRecords database; Supplementary Table S7: The Disease tree numbers of the data set DS₇ consisting of 373 diseases with their disease tree numbers which were gathered from the MeSH database; Supplementary Table S8: The results of top 20 lncRNAs related to these three diseases. Supplementary Materials: Deep representation of the probabilistic scheme in our method.

Author Contributions: Conceptualization, J.Y. and L.W.; Methodology, J.Y., P.P. and L.W.; Validation, L.K., X.L. and Z.W.; Formal Analysis, J.Y. and L.W.; Investigation, L.K. and Z.W.; Resources, P.P. and Z.W.; Data Curation, J.Y. and P.P.; Writing—Original Draft Preparation, J.Y. and P.P.; Writing—Review and Editing, L.W. and X.L.; Supervision, L.W.; Project Administration, L.K. and X.L.; Funding Acquisition, L.W.

Funding: This research is partly sponsored by the Natural Science Foundation of Hunan Province (No. 2018JJ4058, No. 2017JJ5036), the National Natural Science Foundation of China (No. 61640210, No. 61672447), the CERNET Next Generation Internet Technology Innovation Project (No. NGII20160305), the project of “12th Five-Year” planning of Education Science in Hunan Province (No. XJK015BZY031) and the CERNET Next Generation Internet Technology Innovation Project (No. NGII20160305, No. NGII20170109).

Acknowledgments: The authors thank the anonymous referees for suggestions that helped improve the paper substantially.

Conflicts of Interest: The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

1. Li, Y.; Zhang, J.; Pan, J.; Feng, X.; Duan, P.; Yin, X.; Xu, Y.; Wang, X.; Zou, S. Insights into the roles of lncRNAs in skeletal and dental diseases. *Cell Biosci.* **2018**, *8*, 8. [[CrossRef](#)] [[PubMed](#)]
2. Garitano-Trojaola, A.; Agirre, X.; Prósper, F.; Fortes, P. Long non-coding RNAs in haematological malignancies. *Int. J. Mol. Sci.* **2013**, *14*, 15386. [[CrossRef](#)] [[PubMed](#)]
3. Guttman, M.; Russell, P.; Ingolia, N.T.; Weissman, J.S.; Lander, E.S.R. Ribosome Profiling Provides Evidence that Large Noncoding RNAs Do Not Encode Proteins. *Cell* **2013**, *154*, 240–251. [[CrossRef](#)] [[PubMed](#)]
4. Guttman, M.; Rinn, J.L. Modular regulatory principles of large non-coding RNAs. *Nature* **2012**, *482*, 339–346. [[CrossRef](#)] [[PubMed](#)]
5. Wang, K.C.; Chang, H.Y. Molecular mechanisms of long noncoding RNAs. *Mol. Cell* **2011**, *43*, 904–914. [[CrossRef](#)] [[PubMed](#)]
6. Wapinski, O.; Chang, H.Y. Long noncoding RNAs and human disease. *Trends Cell Biol.* **2011**, *21*, 354–361. [[CrossRef](#)] [[PubMed](#)]
7. Derrien, T.; Johnson, R.; Bussotti, G.; Tanzer, A.; Djebali, A.; Tilgner, H.; Guernec, G.; Martin, D.; Merkel, A.; Knowles, D. The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. *Genome Res.* **2012**, *22*, 1775–1789. [[CrossRef](#)] [[PubMed](#)]
8. Zhao, W.; Luo, J.; Jiao, S. Comprehensive characterization of cancer subtype associated long non-coding RNAs and their clinical implications. *Sci. Rep.* **2014**, *4*, 6591. [[CrossRef](#)] [[PubMed](#)]
9. Cheetham, S.W.; Gruhl, F.; Mattick, J.S.; Dinger, M.E. Long noncoding RNAs and the genetics of cancer. *Br. J. Cancer* **2013**, *108*, 2419–2425. [[CrossRef](#)] [[PubMed](#)]
10. Mercer, T.R.; Dinger, M.E.; Mattick, J.S. Long non-coding RNAs: Insights into functions. *Nat. Rev. Genet.* **2009**, *10*, 155–159. [[CrossRef](#)] [[PubMed](#)]
11. Taft, R.J.; Pang, K.C.; Mercer, T.R.; Dinger, M.; Mattick, J.S. Non-coding RNAs: Regulators of disease. *J. Pathol.* **2010**, *220*, 126–139. [[CrossRef](#)] [[PubMed](#)]
12. Zhang, Q.; Chen, C.Y.; Yedavalli, V.S.; Jeang, K.T. NEAT1 long noncoding RNA and paraspeckle bodies modulate HIV-1 posttranscriptional expression. *MBio* **2013**, *4*, e00596-12. [[CrossRef](#)] [[PubMed](#)]
13. Pasmant, E.; Sabbagh, A.; Vidaud, M.; Bièche, I. ANRIL, a long, noncoding RNA, is an unexpected major hotspot in GWAS. *FASEB J.* **2011**, *25*, 444–448. [[CrossRef](#)] [[PubMed](#)]
14. Faghihi, M.A.; Modarresi, F.; Khalil, A.M.; Wood, D.E.; Sahagan, B.G.; Morgan, T.E.; Finch, C.E.; Laurent, G.S.; Kenny, P.J.; Wahlestedt, C. Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of β -secretase. *Nat. Med.* **2008**, *14*, 723–730. [[CrossRef](#)] [[PubMed](#)]
15. Malih, S.; Saidijam, M.; Malih, N. A brief review on long noncoding RNAs: a new paradigm in breast cancer pathogenesis, diagnosis and therapy. *Tumor Biology* **2016**, *37*, 1479–1485. [[CrossRef](#)] [[PubMed](#)]
16. Cui, Z.; Ren, S.; Lu, J.; Wang, F.; Xu, W.; Sun, Y.; Wei, M.; Chen, J.; Gao, X.; Xu, C.; Mao, J.H.; Sun, Y. The prostate cancer-up-regulated long noncoding RNA PlncRNA-1 modulates apoptosis and proliferation through reciprocal regulation of androgen receptor. *Urol. Oncol.* **2013**, *31*, 1117–1123. [[CrossRef](#)] [[PubMed](#)]
17. Wang, J.; Liu, X.; Wu, H.; Ni, P.; Gu, Z.; Qiao, Y.; Chen, N.; Sun, F.; Fan, Q. CREB up-regulates long noncoding RNA, HULC expression through interaction with microRNA-372 in liver cancer. *Nucleic Acids Res.* **2010**, *38*, 5366–5383. [[CrossRef](#)] [[PubMed](#)]
18. Ma, Z.; Xue, S.; Zeng, B.; Qiu, D. lncRNA SNHG5 is associated with poor prognosis of bladder cancer and promotes bladder cancer cell proliferation through targeting p27. *Oncol. Lett.* **2018**, *15*, 1924–1930. [[CrossRef](#)] [[PubMed](#)]
19. Spizzo, R.; Almeida, M.I.; Colombatti, A.; Calin, G.A. Long non-coding RNAs and cancer: A new frontier of translational research? *Oncogene* **2012**, *31*, 4577–4587. [[CrossRef](#)] [[PubMed](#)]
20. Gupta, R.A.; Shah, N.; Wang, K.C.; Kim, J.; Horlings, H.M.; Wong, D.J.; Tsai, M.C.; Hung, T.; Argani, P.; Rinn, J.L. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* **2010**, *464*, 1071–1076. [[CrossRef](#)] [[PubMed](#)]
21. Gutschner, T.; Hammerle, M.; Eissmann, M.; Hsu, J.; Kim, Y.; Hung, G.; Revenko, A.; Arun, G.; Stentrup, M.; Gross, M. The noncoding RNA MALAT1 is a critical regulator of the metastasis phenotype of lung cancer cells. *Cancer Res.* **2013**, *73*, 1180–1189. [[CrossRef](#)] [[PubMed](#)]

22. Lottin, S.; Adriaenssens, E.; Berteaux, N.; Leprêtre, A.; Vilain, M.O.; Denhez, E.; Coll, J.; Dugimont, T.; Cury, J.J. The human H19 gene is frequently overexpressed in myometrium and stroma during pathological endometrial proliferative events. *Eur. J. Cancer* **2005**, *41*, 168–177. [[CrossRef](#)] [[PubMed](#)]
23. Sun, J.; Shi, H.; Wang, Z.; Zhang, C.; Liu, L.; Wang, L.; He, W.; Hao, D.; Liu, S.; Zhou, M. Inferring novel lncRNA–disease associations based on a random walk model of a lncRNA functional similarity network. *Mol. Biosyst.* **2014**, *10*, 2074–2081. [[CrossRef](#)] [[PubMed](#)]
24. Ping, P.; Wang, L.; Kuang, L.; Ye, S.; Lqbal, M.F.B. A Novel Method based on lncRNA–disease association Network for lncRNA–Disease Association Prediction. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2018**. [[CrossRef](#)]
25. Yang, X.; Gao, L.; Guo, X.; Shi, X.; Wu, H.; Song, F.; Wang, B. A network based method for analysis of lncRNA–disease associations and prediction of lncRNAs implicated in diseases. *PLoS ONE* **2014**, *9*, e87797. [[CrossRef](#)] [[PubMed](#)]
26. Ding, L.; Wang, M.; Sun, D.; Li, A. TPGLDA: Novel prediction of associations between lncRNAs and diseases via lncRNA–disease–gene tripartite graph. *Sci. Rep.* **2018**, *8*, 1065. [[CrossRef](#)] [[PubMed](#)]
27. Chen, X. Predicting lncRNA–disease associations and constructing lncRNA functional similarity network based on the information of miRNA. *Sci. Rep.* **2015**, *5*, 13186. [[CrossRef](#)] [[PubMed](#)]
28. Liu, M.X.; Chen, X.; Chen, G.; Cui, Q.H.; Yan, G.Y. A computational framework to infer human disease-associated long noncoding RNAs. *PLoS ONE* **2014**, *9*, e84408. [[CrossRef](#)] [[PubMed](#)]
29. Li, J.W.; Cheng, G.; Wang, Y.C.; Gao, C.; Wang, Y.; Ma, W.; Tu, J.; Wang, J.; Chen, Z.; Kong, W.; Cui, Q. A bioinformatics method for predicting long noncoding RNAs associated with vascular disease. *Sci. China Life Sci.* **2014**, *57*, 852–857. [[CrossRef](#)] [[PubMed](#)]
30. Gu, C.; Liao, B.; Li, X.; Cai, L.; Li, Z.; Li, K.; Yang, J. Global network random walk for predicting potential human lncRNA–disease associations. *Sci. Rep.* **2017**, *7*, 12442. [[CrossRef](#)] [[PubMed](#)]
31. Gligorijević, V.; Pržulj, N. Methods for biological data integration: perspectives and challenges. *J. R. Soc. Interface* **2015**, *12*, 20150571. [[CrossRef](#)] [[PubMed](#)]
32. Zeng, X.; Ding, N.; Rodríguez-Patón, A.; Zou, Q. Probability-based collaborative filtering model for predicting gene–disease associations. *BMC Med. Genom.* **2017**, *10* (Suppl. 5), 76. [[CrossRef](#)] [[PubMed](#)]
33. Zhao, H.; Kuang, L.; Wang, L.; Ping, P.; Xuan, Z.; Pei, T.; Wu, Z. Prediction of microRNA–disease associations based on distance correlation set. *BMC Bioinform.* **2018**, *19*, 141. [[CrossRef](#)] [[PubMed](#)]
34. Zeng, X.; Zhang, X.; Zou, Q. Integrative approaches for predicting microRNA function and prioritizing disease-related microRNA using biological interaction networks. *Brief. Bioinform.* **2016**, *17*, 193. [[CrossRef](#)] [[PubMed](#)]
35. Lu, C.; Yang, M.; Luo, F.; Wu, F.X.; Li, M.; Pan, Y.; Li, Y.; Wang, J. Prediction of lncRNA–disease associations based on inductive matrix completion. *Bioinformatics* **2018**. [[CrossRef](#)] [[PubMed](#)]
36. Zhang, J.; Zhang, Z.; Chen, Z.; Deng, L. Integrating multiple heterogeneous networks for novel lncRNA–disease association inference. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2017**. [[CrossRef](#)] [[PubMed](#)]
37. Fu, G.; Wang, J.; Domeniconi, C.; Yu, G. Matrix factorization based data fusion for the prediction of lncRNA–disease associations. *Bioinformatics* **2017**. [[CrossRef](#)] [[PubMed](#)]
38. Chen, X. KATZLDA: KATZ measure for the lncRNA–disease association prediction. *Sci. Rep.* **2015**, *5*, 16840. [[CrossRef](#)] [[PubMed](#)]
39. Li, Y.; Qiu, C.; Tu, J.; Geng, B.; Yang, J.; Jiang, T.; Cui, Q. HMDD v2.0: A database for experimentally supported human microRNA and disease associations. *Nucleic Acids Res.* **2014**, *42*, D1070. [[CrossRef](#)] [[PubMed](#)]
40. Li, J.H.; Liu, S.; Zhou, H.; Qu, L.H.; Yang, J.H. starBase v2.0: Decoding miRNA–ceRNA, miRNA–ncRNA and protein–RNA interaction networks from large-scale CLIP–Seq data. *Nucleic Acids Res.* **2014**, *42*, D92. [[CrossRef](#)] [[PubMed](#)]
41. Cui, T.; Lin, Z.; Yan, H.; Yi, Y.; Tan, P.; Zhao, Y.; Hu, Y.; Xu, L.; Li, E.; Wang, D. MNDR v2.0: An updated resource of ncRNA–disease associations in mammals. *Nucleic Acids Res.* **2018**, *46*, D371–D374. [[CrossRef](#)] [[PubMed](#)]
42. Piñero, J.; Àlex, B.; Queraltrosinach, N.; Gutierrez-Sacristan, A.; Deu-Pons, J.; Centeno, E.; Garcia-Garcia, J.; Sanz, F.; Furlong, L.I. DisGeNET: A comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res.* **2017**, *45*, D833–D839. [[CrossRef](#)] [[PubMed](#)]

43. Wang, P.; Ning, S.; Zhang, Y.; Li, R.; Ye, J.; Zhao, Z.; Zhi, H.; Wang, T.; Guo, Z.; Li, X. Identification of lncRNA-associated competing triplets reveals global patterns and prognostic markers for cancer. *Nucleic Acids Res.* **2015**, *43*, 3478. [[CrossRef](#)] [[PubMed](#)]
44. Xiao, F.; Zuo, Z.; Cai, G.; Xiao, F.; Zuo, Z.; Cai, G.; Kang, S.; Gao, X.; Li, T. miRecords: An integrated resource for microRNA-target interactions. *Nucleic Acids Res.* **2008**, *37*, D105–D110. [[CrossRef](#)] [[PubMed](#)]
45. U.S. National Library of Medicine. Medical Subject Headings 2018 [Internet]. Available online: <https://meshb.nlm.nih.gov/search> (accessed on 6 July 2018).
46. Liu, Z.; Zhang, Q.; Lu, L.; Zhou, T. Link prediction in complex networks: A local naïve Bayes model. *Europhys. Lett.* **2011**, *96*, 48007. [[CrossRef](#)] [[CrossRef](#)]
47. Wang, D.; Wang, J.; Lu, M.; Song, F.; Cui, Q. Inferring the human microRNA functional similarity and functional network based on microRNA-associated diseases. *Bioinformatics* **2010**, *26*, 1644–1650. [[CrossRef](#)] [[PubMed](#)]
48. Luo, J.; Ding, P.; Liang, C.; Cao, B.; Chen, X. Collective prediction of disease-associated miRNAs based on transduction learning. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2016**, *14*, 1468–1475. [[CrossRef](#)] [[PubMed](#)]
49. Berger, F.G. Interview: Screening and treatment for colorectal cancer. *Colorectal Cancer* **2013**, *2*, 117–120. [[CrossRef](#)]
50. Prensner, J.R.; Chinnaiyan, A.M. The emergence of lncRNAs in cancer biology. *Cancer Dis.* **2011**, *1*, 391–407. [[CrossRef](#)] [[PubMed](#)]
51. Gutschner, T.; Diederichs, S. The hallmarks of cancer: A long non-coding RNA point of view. *RNA Biol.* **2012**, *9*, 703–719. [[CrossRef](#)] [[PubMed](#)]
52. Chen, D.L.; Chen, L.Z.; Lu, Y.X.; Zhang, D.S.; Zeng, Z.L.; Pan, Z.Z.; Huang, P.; Wang, F.H.; Li, Y.H.; Ju, H.Q. Long noncoding RNA XIST expedites metastasis and modulates epithelial-mesenchymal transition in colorectal cancer. *Econ. Theory Bus. Manag.* **2017**, *8*, e3011. [[CrossRef](#)] [[PubMed](#)]
53. Li, P.; Zhang, X.; Wang, H.; Wang, L.; Liu, T.; Du, L.; Yang, Y.; Wang, C. MALAT1 is associated with poor response to oxaliplatin-based chemotherapy in colorectal cancer patients and promotes chemoresistance through EZH2. *Mol. Cancer Ther.* **2017**, *16*, 739–751. [[CrossRef](#)] [[PubMed](#)]
54. Nakano, S.; Murakami, K.; Meguro, M.; Soejima, H.; Higashimoto, K.; Urano, T.; Kugoh, H.; Mukai, T.; Ikeguchi, M.; Oshimura, M. Expression profile of LIT1/KCNQ1OT1, and epigenetic status at the KvDMR1 in colorectal cancers. *Cancer Sci.* **2006**, *97*, 1147–1154. [[CrossRef](#)] [[PubMed](#)]
55. Han, D.; Gao, X.; Wang, M.; Qiao, Y.; Xu, Y.; Yang, J.; Dong, N.; He, J.; Sun, Q.; Lv, G. Long noncoding RNA H19 indicates a poor prognosis of colorectal cancer and promotes tumor growth by recruiting and binding to eIF4A3. *Oncotarget* **2016**, *7*, 22159–22173. [[CrossRef](#)] [[PubMed](#)]
56. Weiss, M.; Plass, C.; Gerhauser, C. Role of lncRNAs in prostate cancer development and progression. *Biol. Chem.* **2014**, *395*, 1275–1290. [[CrossRef](#)] [[PubMed](#)]
57. Yang, G.; Lu, X.; Yuan, L. LncRNA: A link between RNA and cancer. *Biochim. Biophys. Acta* **2014**, *1839*, 1097–1109. [[CrossRef](#)] [[PubMed](#)]
58. Chakravarty, D.; Sboner, A.; Nair, S.S.; Giannopoulou, E.; Li, R.; Hennig, S.; Mosquera, J.M.; Pauwels, J.; Park, K.; Kossai, M. The oestrogen receptor alpha-regulated lncRNA NEAT1 is a critical modulator of prostate cancer. *Nat. Commun.* **2014**, *5*, 5383. [[CrossRef](#)] [[PubMed](#)]
59. Ren, S.; Liu, Y.; Xu, W.; Sun, Y.; Lu, J.; Wang, F.; Wei, M.; Shen, J.; Hou, J.; Gao, X.; Xu, C. Long noncoding RNA MALAT-1 is a new potential therapeutic target for castration resistant prostate cancer. *J. Urol.* **2013**, *190*, 2278–2287. [[CrossRef](#)] [[PubMed](#)]
60. Zhu, M.; Chen, Q.; Liu, X.; Sun, Q.; Zhao, X.; Deng, R.; Wang, Y.; Huang, J.; Xu, M.; Yan, J.; Yu, J. lncRNA H19/miR-675 axis represses prostate cancer metastasis by targeting TGFBI. *FEBS J.* **2014**, *281*, 3766–3775. [[CrossRef](#)] [[PubMed](#)]
61. Tian, X.; Zhang, G.; Zhao, H.; Li, Y.; Zhu, C. Long non-coding RNA NEAT1 contributes to docetaxel resistance of prostate cancer through inducing RET expression by sponging miR-34a. *RSC Adv.* **2017**, *7*, 42986–42996. [[CrossRef](#)]
62. Boele, F.W.; Rooney, A.G.; Grant, R.; Klein, M. Psychiatric symptoms in glioma patients: from diagnosis to management. *Neuropsychiatr. Dis. Treat.* **2015**, *11*, 1413–1420. [[CrossRef](#)] [[PubMed](#)]
63. Zhou, Q.; Liu, J.; Quan, J.; Liu, W.; Tan, H.; Li, W. lncRNAs as potential molecular biomarkers for the clinicopathology and prognosis of glioma: A systematic review and meta-analysis. *Gene* **2018**. [[CrossRef](#)] [[PubMed](#)]

64. Ma, K.X.; Wang, H.J.; Li, X.R.; Li, T.; Su, G.; Yang, P.; Wu, J.W. Long noncoding RNA MALAT1 associates with the malignant status and poor prognosis in glioma. *Tumour Biol. J. Int. Soc. Oncodev. Biol. Med.* **2015**, *36*, 3355–3359. [[CrossRef](#)] [[PubMed](#)]
65. Zhang, T.; Wang, Y.R.; Zeng, F.; Cao, H.Y.; Zhou, H.D.; Wang, Y.J. LncRNA H19 is overexpressed in glioma tissue, is negatively associated with patient survival, and promotes tumor growth through its derivative miR-675. *Eur. Rev. Med. Pharmacol. Sci.* **2016**, *20*, 4891–4897. [[PubMed](#)]
66. Li, J.; Zhang, M.; An, G.; Ma, Q. LncRNA TUG1 acts as a tumor suppressor in human glioma by promoting cell apoptosis. *Exp. Biol. Med.* **2016**, *241*, 644. [[CrossRef](#)] [[PubMed](#)]



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).