SCIENTIFIC REP🞓RTS

OPEN

Predicting lncRNA-disease associations and constructing lncRNA functional similarity network based on the information of miRNA

Xing Chen[1,2]

Accumulating experimental studies have indicated that lncRNAs play important roles in various critical biological process and their alterations and dysregulations have been associated with many important complex diseases. Developing effective computational models to predict potential disease-lncRNA association could benefit not only the understanding of disease mechanism at lncRNA level, but also the detection of disease biomarkers for disease diagnosis, treatment, prognosis and prevention. However, known experimentally confirmed disease-lncRNA associations are still very limited. In this study, a novel model of HyperGeometric distribution for LncRNA-Disease Association inference (HGLDA) was developed to predict lncRNA-disease associations by integrating miRNA-disease associations and lncRNA-miRNA interactions. Although HGLDA didn't rely on any known disease-lncRNA associations, it still obtained an AUC of 0.7621 in the leave-one-out cross validation. Furthermore, 19 predicted associations for breast cancer, lung cancer, and colorectal cancer were verified by biological experimental studies. Furthermore, the model of LncRNA Functional Similarity Calculation based on the information of MiRNA (LFSCM) was developed to calculate lncRNA functional similarity on a large scale by integrating disease semantic similarity, miRNA-disease associations, and miRNA-lncRNA interactions. It is anticipated that HGLDA and LFSCM could be effective biological tools for biomedical research.

Based on the assumption of the central dogma of molecular biology, genetic information is stored in protein-coding genes and RNA is just an intermediary between a DNA sequence and its encoded protein[1,2]. However, sequence analysis shown that there were only ~20,000 protein-coding genes in the human genome and more than 98% of the human genome does not encode protein sequences[3–10], yielding tens of thousands of non-coding RNAs (ncRNAs). Based on accumulating experimental evidences, these ncRNAs have played very fundamental and critical roles in various biological processes[11]. Based on whether transcript lengths are larger than 200 nucleotides, ncRNAs can be further divided into small ncRNA (such as miRNA, siRNA, and piRNA) and long ncRNA (lncRNA). Long non-coding RNAs (lncRNAs) are a heterogeneous class of ncRNAs with non-protein-coding transcripts longer than 200 nucleotides[8,12,13]. In comparison with protein-coding genes, lncRNAs have the following differences: (1) lncRNAs have less conservation across species[14,15]; (2) lncRNAs have relatively lower expression level and much more tissue-specific pattern[16–18]. (3) lncRNAs have longer, but fewer, exons. In the early 1990 s, H19 and Xist were first identified based on traditional gene mapping approaches[19–21]. In the

[1]National Center for Mathematics and Interdisciplinary Sciences, Chinese Academy of Sciences, Beijing, 100190, China. [2]Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, 100190, China. Correspondence and requests for materials should be addressed to X.C. (email: xingchen@amss.ac.cn)

recent few years, there has been rapid development in both experimental technology and computational prediction algorithm for lncRNA discovery. Thousands of lncRNAs have been discovered in eukaryotic organisms ranging from nematodes to humans[15,16,22,23]. For example, based on tiling arrays, HOTAIR (HOX antisense intergenic RNA) and HOTTIP (HOXA transcript at the distal tip) were discovered in the homeobox gene regions (HOX clusters)[24,25]. Guttman, et al.[23] discovered 1600 novel mouse lncRNAs by integrating gene expression data, the presence of chromatin marks for promoter regions and gene bodies, and the known annotations of coding transcripts. Cabili, et al.[14] generated the human lincRNA catalog across 24 different human cell types and tissues based on chromatin marks and RNA-sequencing (RNA-seq) data.

In recent several years, accumulating experimental studies have shown that lncRNAs play important roles in various critical biological process, such as cell proliferation, differentiation, chromatin remodeling, epigenetic regulation, genomic splicing, transcription, translation and so on[9,12,18,22,23,26–29]. Specifically, lncRNA can bind to proteins or miRNAs, resulting in functional inhibition of proteins and titration of miRNAs, respectively[30]. According to the molecular mechanism of lncRNAs, the emerging archetypes of molecular functions of lncRNAs could be divided into signals, decoys, guides, and scaffolds[31]. It has been demonstrated that lncRNA have a very complicated regulation network, but the underlying mechanism of lncRNA-related regulation is still remain unclear. In the light of important biological functions of lncRNAs, the alterations and dysregulations of lncRNAs have been associated with the development and progression of many different complex diseases[12,18,26], including cardiovascular diseases[32], neurological disorders[33], diabetes[34], HIV[35] and various types of cancers, such as breast cancer[36,37], hepatocellular cancer[38,39], prostate cancer[40,41], lung cancer[42,43]. In the past few years, many researchers have focused their researches on lncRNA-disease associations, and they have found some specific lncRNAs associated with various diseases. For example, lncRNA HOTAIR has 100 to approximately 2,000 times expression levels in breast cancer metastases based on quantitative PCR[37,44], and its expression level are correlated with metastasis and progression of other various cancers, such as colorectal cancer[45,46], gastric cancer[47,48], liver cancer[49], lung cancer[47] and so on. Therefore, HOTAIR was considered as potential biomarker in various types of cancers[45]. Except for HOTAIR, the dysfunction of lncRNA H19 is also involved in various diseases. For example, H19 could be used as a potential prognostic tumour marker for the early recurrence of bladder cancer[50]. Furthermore, it has been demonstrated that down-regulation of H19 significantly decreases breast and lung cancer cell clonogenicity and anchorage-independent growth based on a knockdown approach[36]. Several experimental studies have also shown that lncRNA BCAR4 is associated with breast cancer, which is expressed in 27% of primary breast tumors[51–53]. Specifically, in human ZR-75-1 and MCF7 breast cancer cells, the forced expression of BCAR4 causes cell proliferation in the absence of estrogen and in the presence of various antiestrogens, indicating BCAR4 could considered as a proper target for the treatment of antiestrogen-resistant breast cancer[51].

Considering the important roles of lncRNAs in various biological processes regulation and complex diseases development and progression, potential disease-lncRNA associations identification could not only benefit the underlying disease mechanism mining at lncRNA level, but also facilitate disease biomarkers detection and drug discovery for disease diagnosis, treatment, prognosis and prevention[29,54]. Computational models and tools can effectively decrease the time and cost of biological experiments by quantifying the association probability of each lncRNA-disease pair and verifying most promising lncRNA-disease pairs with high scores based on further biological experimental validation. Nowadays, developing effective computational models by integrating various kinds of biological datasets to prioritize disease-related lncRNAs has become one of the most important and attracting topics in the fields of both lncRNAs and complex diseases.

Some computational models have been developed to infer novel disease-lncRNA associations. In the previous study, Chen et al.[54] presented a semi-supervised learning method, LRLSLDA, to infer novel human lncRNA-disease associations. LRLSLDA was developed based on the assumption that similar diseases tend to interact with functionally similar lncRNAs and the framework of Laplacian Regularized Least Squares. LRLSLDA is a reliable tool for lncRNA-disease association prediction. More importantly, it does not need negative samples. However, the parameter selection problem and the problem of combining two different classifiers into the final classifier exist in this method. Based on the same assumption, Sun et al.[55] presented a method to constructed a lncRNA-lncRNA functional similarity network, then they proposed a global network-based computational method named RWRlncD by integrating disease similarity network, lncRNAs functional network and known lncRNA-disease associations. However, this method can't be applied to the lncRNAs without any known associated diseases. Li et al.[56] developed a simple genomic location based bioinformatics method for the prediction of novel associations between lncRNAs and vascular disease. However, not all of the lncRNAs are related with their neighbor genes and no statistical tests were used, which resulted in limitations of this method. Yang et al.[57] investigated lncRNA-disease associations by constructing the lncRNA-disease association network and coding-non-coding gene-disease bipartite network based on known associations between diseases and disease genes. Then, a propagation algorithm was applied to infer the underlying lncRNA-disease associations. This method also has some limitations, such as the lack of the information of non-coding genes and protein coding genes interactions and similarities and lncRNA functional annotation. Zhao et al.[58] developed the naive Bayesian classifier to identify cancer-related lncRNAs based on the integration of genome, regulome and transcriptome data. The important limitation of this method is that they

regard the unknown lncRNA-disease associations as negative samples, which would largely influence the predictive accuracy of the method. Recently, based on the findings that lncRNAs that sharing significantly enriched interacting miRNAs tend to be associated with similar diseases, Zhou *et al.*[59] proposed a novel method named RWRHLD to identify candidate lncRNA-disease associations by integrating miRNA-associated lncRNA-lncRNA crosstalk network, disease-disease similarity network, and known lncRNA-disease association network into a heterogeneous network. Then, a random walk was implemented on this heterogeneous network. This method can only predict associations for the lncRNAs that have lncRNA-miRNA interaction datasets, limiting the wide application of RWRHLD. Aforementioned methods all need the prior information of known experimentally verified lncRNA-disease association. So far, although plenty of biological datasets about lncRNA sequence and expression have been generated and stored in some publicly available databases, such as NRED[60], lncRNAdb[28], NONCODE[61], the number of lncRNAs reported to be associated with diseases is still very limited. Liu *et al.*[62] developed a method by integrating human lncRNA and gene expression profiles, and human disease-associated gene data. This method didn't rely on known lncRNA-disease associations and obtained an AUC of 0.7645 for non-tissue-specific lincRNAs. However, too many false positives would be brought based on the ROC curve in that paper.

Nowadays, plenty of experimentally confirmed miRNA-disease associations have been collected in various databases[63–66]. Therefore, the model of <u>H</u>yper<u>G</u>eometric distribution for <u>L</u>ncRNA-<u>D</u>isease <u>A</u>ssociation inference (HGLDA) was developed here to predict potential lncRNA-disease associations by integrating known miRNA-disease associations and lncRNA-miRNA interactions. Although HGLDA didn't rely on any known disease-related lncRNAs associations, it still obtained a reliable AUC of 0.7621 in the leave-one-out cross validation (LOOCV) based on known experimentally verified lncRNA-disease associations from the LncRNADisease database[29]. HGLDA was also applied to predict Breast Cancer, Lung Cancer, and Colorectal Cancer-related lncRNAs. Seven, seven, and five predicted potential associations with false discovery rate (FDR) less than 0.05 have been confirmed by recent biological experiments for these three important human complex diseases, respectively. Above results effectively demonstrated its potential ability of inferring disease-lncRNA associations and detecting biomarkers detection for human disease diagnosis, treatment, prognosis and prevention. Furthermore, the model of <u>L</u>ncRNA <u>F</u>unctional <u>S</u>imilarity <u>C</u>alculation based on the information of <u>M</u>iRNA (LFSCM) was developed to quantitatively calculate lncRNA functional similarity on a large scale by integrating disease semantic similarity, miRNA-disease associations, and miRNA-lncRNA interactions.

Results

Performance evaluation of potential lncRNA-disease association prediction. HGLDA was applied to the known experimentally verified lncRNA-disease associations in the lncRNADisease database in the framework of LOOCV. Each known disease-lncRNA association was left out in turn as test sample. How well this test sample was ranked relative to the candidate samples (all the disease-lncRNA pairs without the evidence to confirm their association) was evaluated. When the rank of this test sample exceeds the given threshold, this model was considered to provide a successful prediction. When the thresholds were varied, true positive rate (TPR, sensitivity) and false positive rate (FPR, 1-specificity) could be obtained. Here, sensitivity refers to the percentage of the test samples whose ranking is higher than the given threshold. Specificity refers to the percentage of samples that are below the threshold. Receiver-operating characteristics (ROC) curve was drawn by plotting TPR versus FPR at different thresholds. Area under ROC curve (AUC) was further calculated to evaluate the performance of HGLDA. AUC = 1 indicates perfect performance and AUC = 0.5 indicates random performance. As a result, HGLDA achieved an AUC of 0.7621 (see Fig. 1). One important fact must be pointed out is that HGLDA predict potential lncRNA-disease association without relying on the information of known disease-lncRNA associations. Although previous study of predicting potential lncRNA-disease associations by integrating disease-gene associations and gene-lncRNA co-expression relationship obtained a comparable AUC of 0.7645, the ROC curve in that study is much below the ROC curve in this study when FPR is small, which is particularly important for practical biological research. More importantly, available experimentally verified disease-miRNA associations are still comparatively rare relative to the known disease-gene associations. The performance of HGLDA would be further improved when more known miRNA-disease associations could be obtained in the future.

Case studies of potential lncRNA-disease association prediction. HGLDA was applied to predict potential disease-lncRNA associations for all the diseases investigated in this article. Potential predictive associations with significant FDR values were publicly released to benefit the biological experimental validation (see Supplementary Table 1). It is anticipated that these potential lncRNA-disease associations which significantly share common miRNAs could be validated by biological experiments and provide important complementary for experimental studies. Especially, plenty of evidences have demonstrated that lncRNAs plays important roles in various kinds of human cancers[36–41]. Therefore, case studies about three kinds of important cancers were implemented to show the predictive performance of HGLDA. Predictive results were confirmed based on recent experimental literatures.

As the second leading cause of female cancer death, breast cancer comprises 22% of all cancers in women[67,68]. Breast cancer is caused because of multiple molecular alterations and traditionally diagnosed
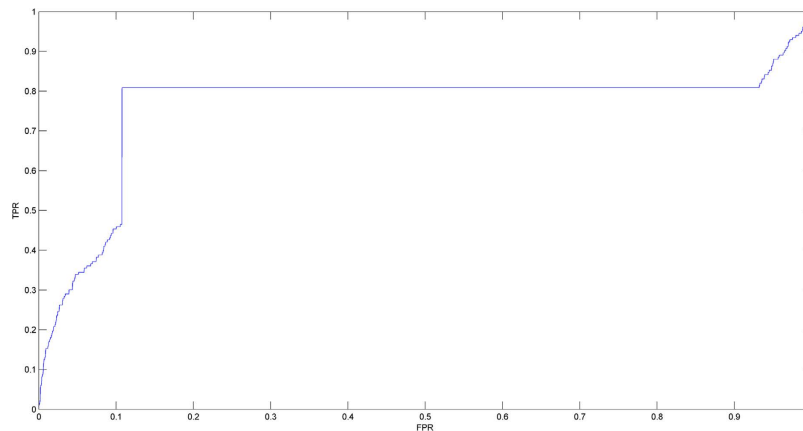
Figure 1. Performance evaluation for the HGLDA in terms of ROC curve and AUC based on LOOCV.
As a result, HGLDA achieved an AUC of 0.7621, demonstrating its reliable predictive ability even if potential lncRNA-disease associations were predicted without relying on the information of known disease-lncRNA associations in the model of HGLDA.

Disease	lncRNA	Evidence (PMID)
Breast cancer	MALAT1	24525122;19379481
Breast cancer	H19	16707459;14729626;12419837
Breast cancer	CDKN2B-AS1	17440112;20956613
Breast cancer	NEAT1	25417700;23825647
Breast cancer	XIST	17545591
Breast cancer	KCNQ1OT1	21304052
Breast cancer	HOTAIRM1	25296969
Lung cancer	EPB41L4A-AS1 BCYRN1	16973895;9490301
Lung cancer	MALAT1	20937273;24757675;24667321
Lung cancer	TUG1	24853421
Lung cancer	GAS5	24357161;23676682
Lung cancer	HOTAIR	25491133;24591352;24155936
Lung cancer	H19	16707459;8838103;7700644
Lung cancer	NEAT1	25010625
Colorectal cancer	XIST	17143621;22879877
Colorectal cancer	HOTAIR	24531795;21862635;24667321
Colorectal cancer	MALAT1	21503572;25446987;25031737
Colorectal cancer	KCNQ1OT1	16965397;11340379;23660942
Colorectal cancer	H19	18719115;19926638;22121898

Table 1. HGLDA was applied to three kinds of important cancer (breast cancer, lung cancer, and colorectal cancer). As a result, 19 predicted lncRNA-disease pairs with significant FDR less than 0.05 have been confirmed based on recent experimental literatures.

based on histopathological features such as tumor size, grade and lymph node status[69]. Researches showed that lncRNA plays an important role in many biological processes and is strongly associated with the formation of various cancers including breast cancer[69,70]. To better diagnose and treat breast cancer, it is necessary to predict breast cancer-related lncRNAs and identify lncRNA biomarkers[70]. HGLDA was implemented to prioritize candidate lncRNAs for breast cancer. As a result, seven lncRNAs with significant FDR less than 0.05 have been confirmed based on recent experimental literatures (see Table 1). For example, XIST, KCNQ1OT1 and NEAT1 are there experimentally confirmed breast cancer related lncRNAs, which have been ranked 1st, 8th, and 12th in the predicted list based on the model of HGLDA, respectively. The XIST RNA signal variability in the BRCA1 breast tumor is correlated with chromosomal genetic abnormalities, and BRCA1 breast tumors often contain cells showing multiple XIST RNA domains per nucleus[71]. KCNQ1OT1 is induced by estrogen in estrogen receptor-alpha (ERα) expressing

breast cancer cells and further mediate CDKN1C repression through epigenetic repression[72]. The alternative splicing of NEAT1 may play important role in nicotine induced breast cancer development[73] and breast cancer patients with high level of NEAT1 expression shows low survival rate[74].

Lung cancer, which can be roughly divided into two groups: non-small cell lung cancer (80.4%) and small cell lung cancer (16.8%) considering disease patterns and treatment strategies, is the leading cause of cancer-related death worldwide in both men and women[75,76]. There are estimated 1.4 million deaths resulting from lung cancer each year[77,78]. Data show that the risk of lung cancer mortality is even greater than the combination of the next three most common cancers (colon, breast and prostate)[75]. Specially, five-year survival rate of lung cancer patients is only approximately 15%, which is much lower than other cancers types[79,80]. To diagnose and treat lung cancer in a better and more efficiently way, more attentions are focused on the deregulation of protein-coding genes to identify oncogenes and tumor suppressors in the last decades[75,81,82]. Recent researches have shown that lncRNAs play a critical role the development and progression of lung cancers[75,82]. Potential lung cancer-related lncRNAs were obtained by selecting candidate lncRNAs with FDR less than 0.05. Seven predicted lncRNAs have been confirmed by independent experimental literatures (see Table 1). According to biological experiments in several studies, it has been confirmed that MALAT1 is a non-coding RNA which plays important roles in many different cancers[47]. Specially it has been shown to be highly associated with metastasis of lung cancer[83–86] and promote lung cancer cell motility by regulating motility related gene expression[87]. Therefore, it could be an important biomarker for metastasis development in lung cancer[49]. TUG is another lung cancer related lncRNA, which can be regulated by P53 to affect non-small cell lung cancer (NSCLC) cell proliferation in part by epigenetically controlling the expression of HOXB7[88]. GAS5, which can also be mediated by P53 pathway, is shown to be a tumor suppressor and down-regulated in NSCLC[89]. These three lncRNAs were all ranked in the top of prediction list for lung cancer (10th, 14th, and 41st, respectively).

As the third most common cancer in men and the second in women, colorectal cancer is one of the most common malignancies in the world and an important threat to human health[90,91]. Data shows that the 5.2% of men and 4.8% of women have the risk of colorectal cancer in the United States and the mortality rate caused by colorectal cancer is nearly 33% in the developed world[90–92]. Some critical mutations underlying the pathogenic mechanism of colorectal cancer have been confirmed[93]. Especially, mutations and dysregulations of some lncRNAs have been linked with the development and progression of colorectal cancer. Five predicted colorectal cancer-related lncRNAs have been confirmed by experimental literature (see Table 1). XIST, MALAT1, H19, and KCNQ1OT1 were ranked in the top four prediction list of colorectal cancer. As a result, recent biological experiments indicated these four lncRNAs all showed high correlation with colorectal cancer. For example, evidences show that expression level change of or DNA amplification of XIST is associated with colorectal carcinoma[94,95]. Also, MALAT1 plays important role in colorectal cancer development by promoting its invasion and metastasis[96–99], and down-regulation of MALAT1 will inhibit colorectal invasion by attenuating Wnt/β-catenin signaling[100]. Moreover, the methylation state of H19 locus is highly related with colorectal cancer[101–105], and the H19-derived microRNA also regulates colorectal cancer development[106]. Loss of imprinting of KCNQ1OT1 is considered as a useful marker for diagnosis of colorectal cancer because of its frequent occurrences in colorectal cancer samples[107].

lncRNA functional similarity. LFSCM was applied to all the lncRNAs investigated in this study. Therefore, pairwise functional similarity among 1114 lncRNAs has been obtained (See Supplementary Table 2).

Discussions

Predicting potential disease-related lncRNAs by integrating various kinds of biological datasets is one of the most important and attracting topics for computational biology research, which is critical for understanding disease mechanism at the lncRNA level and disease biomarkers detection for disease diagnosis, prognosis and prevention. In this study, considering many miRNA-disease associations have been confirmed by recent biological experiments, the model of HGLDA was developed to predict potential disease-lncRNA associations on a large scale by selecting disease-lncRNA pairs which significantly share common miRNA partners. The important difference from previous computational researches about lncRNA-disease inference is that HGLDA doesn't rely on any known lncRNA-disease associations. To validate the performance of HGLDA, LOOCV was implemented on lncRNA-disease association dataset obtained from lncRNADisease database and case studies were further implemented to three important cancers (Breast cancer, Lung Cancer, and Colorectal Cancer). Reliable performance has been obtained in the above validations. Therefore, to facilitate further biological experiment confirmation, significant lncRNA-disease pairs for all the diseases investigated in this study were publicly released. It is anticipated that HGLDA could further demonstrate its potential value for disease-lncRNA association inference and disease biomarker detection in the future.

Calculating lncRNA functional similarity could benefit lncRNA function inference and disease-related lncRNA prioritization. Therefore, based on the assumption that functionally similar lncRNAs tend to interact with functionally similar miRNAs, the model of LFSCM was further developed to quantitatively calculate lncRNA functional similarity. In this model, disease semantic similarity, miRNA-disease associations, and miRNA-lncRNA interactions were integrated on a large scale.

HGLDA obtained the reliable performance in both LOOCV and case studies about three kinds of important cancers, which could be largely attributed to the following several factors. Firstly, known experimentally verified disease-miRNA associations and lncRNA-miRNA interactions were integrated to infer the potential associations between lncRNAs and diseases. Secondly, both miRNA and lncRNA are ncRNAs, which don't encode protein sequences. Therefore, predicting lncRNA-disease associations from miRNA-related datasets is more reasonable than previous study of integrating disease genes and gene-lncRNA co-expression relationship. More importantly, HGLDA doesn't need the prior information of known lncRNA-disease associations, which ensure that this method could be applied to the diseases without any known related lncRNAs. Therefore, HGLDA represents a novel, effective, and important bioinformatics tool for the research of both complex diseases and lncRNAs.

Despite of the reliable performance of HGLDA, there are also some limitations in the model of HGLDA. Although HGLDA doesn't rely on any known experimentally verified lncRNA-disease associations, its performance was not very satisfactory based on the evaluation of LOOCV and could be further improved by integrating more reliable biological datasets, such as disease semantic similarity, disease phenotypic similarity, lncRNA functional similarity, and lncRNA-related various interactions. Although the model of LFSCM can be applied to the lncRNAs without any known related diseases, it can't be applied to those lncRNAs without any known miRNA interaction partners. Furthermore, lncRNA functional similarity was calculated based on known miRNA-disease associations and lncRNA-miRNA interactions, hence LFSCM tends to cause bias to lncRNAs with more miRNA interaction partners or/and lncRNAs with miRNA interaction partners which has been associated with more diseases. LFSCM would be further improved when more known datasets could be available and more reliable types of biological datasets could be integrated. More importantly, as what has been pointed out in the literature[108], it is unwise to use a single disease-related lncRNA to judge cancer risks for all the persons. Therefore, I planned to construct various cancer hallmark networks to effectively evaluate cancer risks based on the lncRNA profiles of each person[108]. Finally, obtaining the tumor recurrence and metastases probability, predicting potential consequences after applying a specific drug to the patients, and identifying molecular signatures to evaluate and predict therapeutic results after cancer treatment in the framework of lncRNAs are three important problems in the personalized medicine[108,109], which could be considered in the future.

Methods

Human miRNA-disease associations. The human miRNA-disease association dataset was downloaded from HMDD in January, 2015, which included 10368 high-quality experimentally verified human miRNA-diseases associations from 3511 papers about 572 miRNA and 378 diseases[110]. Then, duplicate associations with the different evidences were discarded and different miRNA copies were merged which produce the same mature miRNA. Finally, 5430 miRNA–disease associations were obtained, including 383 diseases and 495 miRNAs (see Supplementary Table 3).

lncRNA–miRNA interactions. lncRNA–miRNA interaction dataset was downloaded from star-Base v2.0 database in January, 2015, which provided the most comprehensive experimentally confirmed lncRNA–miRNA interactions based on large scale CLIP-Seq data[111]. After getting rid of duplicate interactions, 10112 lncRNA-miRNA interactions about 132 miRNAs and 1114 lncRNAs were obtained (see Supplementary Table 4).

Disease-lncRNA associations. To validate the performance of HGLDA, the recent version of lncRNA-disease association dataset in the LncRNADisease database was downloaded[29] and LOOCV was implemented based on this golden-standard dataset. For this dataset, I got rid of duplicate associations with different evidences and the lncRNA-disease associations involved with either diseases or lncRNAs which were not contained in the dataset used in this paper. As a result, 183 lncRNA-disease associations were obtained and LOOCV was implemented based on these experimentally verified high-quality associations (see Supplementary Table 5).

HGLDA. The model of HGLDA was developed to predict potential disease-related lncRNAs (See Fig. 2). The hypergeometric distribution test was implemented for each lncRNA-disease pair by examining whether this lncRNA and disease significantly shared common miRNAs which can interact with both of them. The significance was measured by the P-value defined as follows:

$$P = 1 - \sum_{i=0}^{x-1} \frac{\binom{L}{i}\binom{N-L}{M-i}}{\binom{N}{M}}$$

(1)

where N is the total number of miRNAs which are associated with lncRNAs or diseases, M is the number of miRNAs interacting with this given lncRNA, L is the number of miRNAs interacting with this given disease, and x is the number of miRNAs that interact with both of them, respectively. Furthermore, FDR
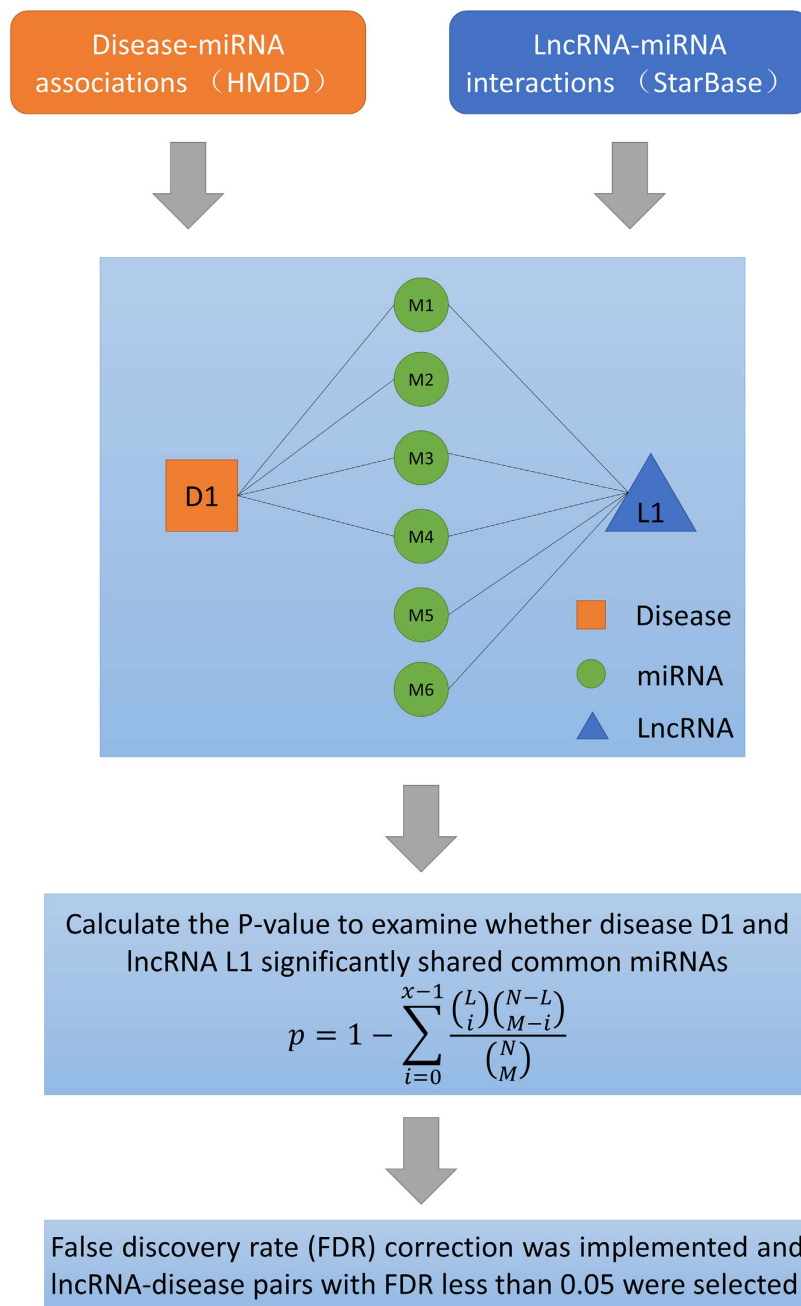
Figure 2. Flowchart of HGLDA, demonstrating the basic ideas of predicting potential disease-related lncRNAs by integrating miRNA-disease associations and lncRNA-miRNA interactions. Firstly, the hypergeometric distribution test was implemented for each lncRNA-disease pair by calculating the P-value to indicate whether this lncRNA and disease significantly shared common miRNAs which can interact with both of them. Then, FDR correction was implemented to all calculated P-values. Finally, those lncRNA-disease pairs with FDR less than 0.05 were selected to be potential lncRNA-disease associations.

correction was implemented to all calculated P-values and those lncRNA-disease pairs with FDR less than 0.05 were considered to be potential lncRNA-disease associations[112].

LFSCM. LFSCM is composed of the following three steps (See Fig. 3): calculating disease semantic similarity based on the disease MeSH descriptors and their direct acyclic graphs (DAGs); calculating miRNA functional similarity based on disease semantic similarity and disease-miRNA associations; calculating lncRNA functional similarity based on miRNA functional similarity and lncRNA-miRNA interactions. For the disease semantic similarity calculation, the method in the literature[113] was adopted. The semantic similarity between two diseases was calculated based on the nodes shared by their disease
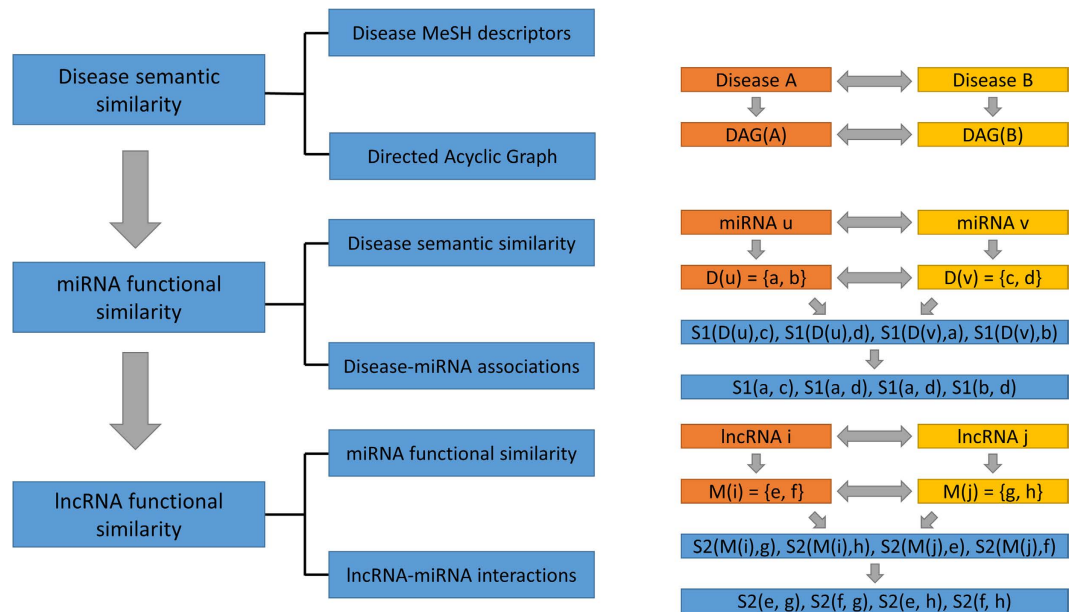
Figure 3. Flowchart of LFSCM, demonstrating the basic ideas of calculating lncRNA functional similarity based on disease semantic similarity, disease-miRNA associations, and lncRNA-miRNA interactions. Firstly, disease semantic similarity among all the diseases investigated in this paper was calculated based on their disease DAGs. Then, disease set associated with each miRNA was identified and the similarity among these disease sets was calculated and considered to be miRNA functional similarity. Finally, lncRNA functional similarity was calculated based on miRNA functional similarity and lncRNA-miRNA interactions.

DAGs. The variable *S1* is denoted as disease semantic similarity matrix, in which the entity *S1(i,j)* in row *i* column *j* represents the semantic similarity between disease *i* and *j*.

For the miRNA functional similarity, the semantic similarity of their associated disease groups was measured. the similarity calculation between miRNA *u* and *v* is taken as an example to demonstrate the procedure, which consisted of three steps: obtaining all the known diseases associated with miRNA *u* and *v*, which are defined as variable *D(u)* and *D(v)* , respectively; calculating the similarity between each disease in one disease groups and the other disease groups; calculating the similarity between two disease groups as the functional similarity between miRNA *u* and *v*. In the second step, taking the similarity calculation between *D(v)* and disease *D1* in the groups of *D(u)* as an example, similarity was defined as follows:

$$S1(D1, D(v)) = \max_{D \in D(v)}(S1(D1, D)) \tag{2}$$

In the third step, the functional similarity between miRNA *u* and *v* was defined

$$S2(u, v) = \frac{\sum_{D \in D(u)} S1(D, D(v)) + \sum_{D \in D(v)} S1(D, D(u))}{|D(u)| + |D(v)|} \tag{3}$$

where *S2* is the miRNA functional similarity matrix and the entity *S2(i,j)* in row *i* column *j* is the functional similarity between miRNA *i* and *j*.

For the lncRNA functional similarity calculation, similar method as miRNA functional similarity calculation was adopted. Here, lncRNA *i* and *j* is take as an example. Firstly, all the miRNAs interacting with these two lncRNA as miRNA groups are defined as *M(i)* and *M(j)*, respectively. Then, the similarity between miRNA group *M(j)* and miRNA *M1* in the miRNA group *M(i)* was defined as follows:

$$S2(M1, M(j)) = \max_{M \in M(j)}(S2(M1, M)) \tag{4}$$

Finally, the similarity between two miRNA groups was calculated and regarded as the functional similarity between corresponding two lncRNAs.

$$FS(i, j) = \frac{\sum_{M \in M(i)} S2(M, M(j)) + \sum_{M \in M(j)} S2(M, M(i))}{|M(i)| + |M(j)|} \tag{5}$$

where *FS* is the lncRNA functional similarity matrix and the entity *FS(i,j)* in row *i* column *j* is the functional similarity between lncRNA *i* and *j*.

References

1. Crick, F., Barnett, L., Brenner, S. & Watts-Tobin, R. General Nature of the Genetic Code for Proteins. *Nature* **192,** 1227–1232 (1961).
2. Yanofsky, C. Establishing the triplet nature of the genetic code. *Cell* **128,** 815–818 (2007).
3. Bertone, P. *et al.* Global identification of human transcribed sequences with genome tiling arrays. *Science* **306,** 2242–2246 (2004).
4. Birney, E. *et al.* Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447,** 799–816 (2007).
5. Carninci, P. *et al.* Genome-wide analysis of mammalian promoter architecture and evolution. *Nat Genet* **38,** 626–635 (2006).
6. Claverie, J. M. Fewer genes, more noncoding RNA. *Science* **309,** 1529–1530 (2005).
7. Core, L. J., Waterfall, J. J. & Lis, J. T. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322,** 1845–1848 (2008).
8. Kapranov, P. *et al.* RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science* **316,** 1484–1488 (2007).
9. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409,** 860–921 (2001).
10. Kapranov, P., Willingham, A. T. & Gingeras, T. R. Genome-wide transcription and the implications for genomic organization. *Nat Rev Genet* **8,** 413–423 (2007).
11. Esteller, M. Non-coding RNAs in human disease. *Nat Rev Genet* **12,** 861–874 (2011).
12. Mercer, T. R., Dinger, M. E. & Mattick, J. S. Long non-coding RNAs: insights into functions. *Nat Rev Genet* **10,** 155–159 (2009).
13. Guttman, M., Russell, P., Ingolia, N. T., Weissman, J. S. & Lander, E. S. Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins. *Cell* **154,** 240–251 (2013).
14. Cabili, M. N. *et al.* Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev* **25,** 1915–1927 (2011).
15. Harrow, J. *et al.* GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res* **22,** 1760–1774 (2012).
16. Guttman, M. *et al.* Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. *Nat Biotechnol* **28,** 503–510 (2010).
17. Mercer, T. R., Dinger, M. E., Sunkin, S. M., Mehler, M. F. & Mattick, J. S. Specific expression of long noncoding RNAs in the mouse brain. *Proc Natl Acad Sci USA* **105,** 716–721 (2008).
18. Ponting, C. P., Oliver, P. L. & Reik, W. Evolution and functions of long noncoding RNAs. *Cell* **136,** 629–641 (2009).
19. Borsani, G. *et al.* Characterization of a murine gene expressed from the inactive X chromosome. *Nature* **351,** 325–329 (1991).
20. Brannan, C. I., Dees, E. C., Ingram, R. S. & Tilghman, S. M. The product of the H19 gene may function as an RNA. *Mol. Cell. Biol.* **10,** 28–36 (1990).
21. Brockdorff, N. *et al.* The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. *Cell* **71,** 515–526 (1992).
22. Khalil, A. M. *et al.* Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc Natl Acad Sci USA* **106,** 11667–11672 (2009).
23. Guttman, M. *et al.* Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* **458,** 223–227 (2009).
24. Rinn, J. L. *et al.* Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* **129,** 1311–1323 (2007).
25. Wang, K. C. *et al.* A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* **472,** 120–124 (2011).
26. Wapinski, O. & Chang, H. Y. Long noncoding RNAs and human disease. *Trends Cell Biol* **21,** 354–361 (2011).
27. Wilusz, J. E., Sunwoo, H. & Spector, D. L. Long noncoding RNAs: functional surprises from the RNA world. *Genes Dev* **23,** 1494–1504 (2009).
28. Amaral, P. P., Clark, M. B., Gascoigne, D. K., Dinger, M. E. & Mattick, J. S. lncRNAdb: a reference database for long noncoding RNAs. *Nucleic Acids Res* **39,** D146–D151 (2011).
29. Chen, G. *et al.* LncRNADisease: a database for long-non-coding RNA-associated diseases. *Nucleic Acids Res* **41,** D983–D986 (2013).
30. Mohanty, V., Gökmen-Polar, Y., Badve, S. & Janga, S. C. Role of lncRNAs in health and disease—size and shape matter. *Brief Funct Genomics* **14,** 115–129 (2014).
31. Wang, K. C. & Chang, H. Y. Molecular mechanisms of long noncoding RNAs. *Mol. Cell* **43,** 904–914 (2011).
32. Congrains, A. *et al.* Genetic variants at the 9p21 locus contribute to atherosclerosis through modulation of ANRIL and CDKN2A/B. *Atherosclerosis* **220,** 449–455 (2012).
33. Johnson, R. Long non-coding RNAs in Huntington's disease neurodegeneration. *Neurobiol Dis* **46,** 245–254 (2012).
34. Pasmant, E., Sabbagh, A., Vidaud, M. & Bièche, I. ANRIL, a long, noncoding RNA, is an unexpected major hotspot in GWAS. *FASEB J* **25,** 444–448 (2011).
35. Zhang, Q., Chen, C.-Y., Yedavalli, V. S. & Jeang, K.-T. NEAT1 Long Noncoding RNA and Paraspeckle Bodies Modulate HIV-1 Posttranscriptional Expression. *MBio* **4,** e00596–00512 (2013).
36. Barsyte-Lovejoy, D. *et al.* The c-Myc oncogene directly induces the H19 noncoding RNA by allele-specific binding to potentiate tumorigenesis. *Cancer Res* **66,** 5330–5337 (2006).
37. Gupta, R. A. *et al.* Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* **464,** 1071–1076 (2010).
38. Calin, G. A. *et al.* Ultraconserved regions encoding ncRNAs are altered in human leukemias and carcinomas. *Cancer Cell* **12,** 215–229 (2007).
39. Panzitt, K. *et al.* Characterization of HULC, a novel gene with striking up-regulation in hepatocellular carcinoma, as noncoding RNA. *Gastroenterology* **132,** 330–342 (2007).
40. de Kok, J. B. *et al.* DD3PCA3, a very sensitive and specific marker to detect prostate tumors. *Cancer Res* **62,** 2695–2698 (2002).
41. Széll, M., Bata-Csörgő, Z. & Kemény, L. The enigmatic world of mRNA-like ncRNAs: their role in human evolution and in human diseases. *Semin Cancer Biol* **18,** 141–148 (2008).
42. Zhang, X. *et al.* A pituitary-derived MEG3 isoform functions as a growth suppressor in tumor cells. *J Clin Endocrinol Metab* **88,** 5119–5126 (2003).
43. Ji, P. *et al.* MALAT-1, a novel noncoding RNA, and thymosin β4 predict metastasis and survival in early-stage non-small cell lung cancer. *Oncogene* **22,** 8031–8041 (2003).

44. Hung, T. & Chang, H. Y. Long noncoding RNA in genome regulation: Prospects and mechanisms. *RNA Biol* **7,** 582–585 (2010).
45. Maass, P. G., Luft, F. C. & Bähring, S. Long non-coding RNA in health and disease. *J Mol Med (Berl)* **92,** 337–346 (2014).
46. Kogo, R. *et al.* Long noncoding RNA HOTAIR regulates polycomb-dependent chromatin modification and is associated with poor prognosis in colorectal cancers. *Cancer Res* **71,** 6320–6326 (2011).
47. Li, G. *et al.* Long Noncoding RNA Plays a Key Role in Metastasis and Prognosis of Hepatocellular Carcinoma. *Biomed Res Int* **2014,** 780521 (2014).
48. Liu, X. *et al.* Lnc RNA HOTAIR functions as a competing endogenous RNA to regulate HER2 expression by sponging miR-331-3p in gastric cancer. *Mol Cancer* **13,** 92 (2014).
49. Hrdlickova, B., de Almeida, R. C., Borek, Z. & Withoff, S. Genetic variation in the non-coding genome: Involvement of micro-RNAs and long non-coding RNAs in disease. *Biochim. Biophys. Acta.* **1842,** 1910–1922 (2014).
50. Ariel, I. *et al.* The imprinted H19 gene is a marker of early recurrence in human bladder carcinoma. *Mol Pathol* **53,** 320–323 (2000).
51. Godinho, M., Meijer, D., Setyono-Han, B., Dorssers, L. C. & van Agthoven, T. Characterization of BCAR4, a novel oncogene causing endocrine resistance in human breast cancer cells. *J. Cell Physiol.* **226,** 1741–1749 (2011).
52. Godinho, M. *et al.* Relevance of BCAR4 in tamoxifen resistance and tumour aggressiveness of human breast cancer. *Br J Cancer* **103,** 1284–1291 (2010).
53. Godinho, M. F. *et al.* BCAR4 induces antioestrogen resistance but sensitises breast cancer to lapatinib. *Br J Cancer* **107,** 947–955 (2012).
54. Chen, X. & Yan, G.-Y. Novel human lncRNA–disease association inference based on lncRNA expression profiles. *Bioinformatics* **29,** 2617–2624 (2013).
55. Sun, J. *et al.* Inferring novel lncRNA–disease associations based on a random walk model of a lncRNA functional similarity network. *Mol Biosyst* **10,** 2074–2081 (2014).
56. Li, J. *et al.* A bioinformatics method for predicting long noncoding RNAs associated with vascular disease. *Sci China Life Sci* **57,** 852–857 (2014).
57. Yang, X. *et al.* A network based method for analysis of lncRNA-disease associations and prediction of lncRNAs implicated in diseases. *PLoS One* **9,** e87797 (2014).
58. Zhao, T. *et al.* Identification of cancer-related lncRNAs through integrating genome, regulome and transcriptome features. *Mol Biosyst* **11,** 126–136 (2015).
59. Zhou, M. *et al.* Prioritizing candidate disease-related long non-coding RNAs by walking on the heterogeneous lncRNA and disease network. *Mol Biosyst* **11,** 760–769 (2015).
60. Dinger, M. E. *et al.* NRED: a database of long noncoding RNA expression. *Nucleic Acids Res* **37,** D122–D126 (2009).
61. Bu, D. *et al.* NONCODE v3. 0: integrative annotation of long noncoding RNAs. *Nucleic Acids Res* **40,** D210–D215 (2012).
62. 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* **9,** e84408 (2014).
63. Lu, M. *et al.* An analysis of human microRNA and disease associations. *PLoS One* **3,** e3420 (2008).
64. Jiang, Q. *et al.* miR2Disease: a manually curated database for microRNA deregulation in human disease. *Nucleic Acids Res* **37,** D98–D104 (2009).
65. Yang, Z. *et al.* dbDEMC: a database of differentially expressed miRNAs in human cancers. *BMC genomics* **11,** S5 (2010).
66. Wang, Y. *et al.* Mammalian ncRNA-disease repository: a global view of ncRNA-mediated disease network. *Cell Death Dis* **4,** e765 (2013).
67. Donahue, H. J. & Genetos, D. C. Genomic approaches in breast cancer research. *Brief Funct Genomics* **12,** 391–396 (2013).
68. Karagoz, K., Sinha, R. & Arga, K. Y. Triple Negative Breast Cancer: A Multi-Omics Network Discovery Strategy for Candidate Targets and Driving Pathways. *OMICS* **19,** 115–130 (2015).
69. Meng, J., Li, P., Zhang, Q., Yang, Z. & Fu, S. A four-long non-coding RNA signature in predicting breast cancer survival. *J Exp Clin Cancer Res* **33,** 84 (2014).
70. Xu, N., Wang, F., Lv, M. & Cheng, L. Microarray expression profile analysis of long non-coding RNAs in human breast cancer: A study of Chinese women. *Biomed Pharmacother* **69,** 221–227 (2015).
71. Vincent-Salomon, A. *et al.* X Inactive–Specific Transcript RNA Coating and Genetic Instability of the X Chromosome in BRCA1 Breast Tumors. *Cancer Res* **67,** 5134–5140 (2007).
72. Rodriguez, B. A. *et al.* Estrogen-mediated epigenetic repression of the imprinted gene cyclin dependent kinase inhibitor 1C in breast cancer cells. *Carcinogenesis* **32,** 812–821 (2011).
73. Bavarva, J. H., Tae, H., Settlage, R. E. & Garner, H. R. Characterizing the genetic basis for nicotine induced cancer development: a transcriptome sequencing study. *PLoS One* **8,** e67252 (2013).
74. Choudhry, H. *et al.* Tumor hypoxia induces nuclear paraspeckle formation through HIF-2α dependent transcriptional activation of NEAT1 leading to cancer cell survival. *Oncogene*, doi: 10.1038/onc.2014.378 (2014).
75. White, N. M. *et al.* Transcriptome sequencing reveals altered long intergenic non-coding RNAs in lung cancer. *Genome. Biol* **15,** 429 (2014).
76. Liu, J. *et al.* Genome and transcriptome sequencing of lung cancers reveal diverse mutational and splicing events. *Genome Res* **22,** 2315–2327 (2012).
77. Jemal, A., Siegel, R., Xu, J. & Ward, E. Cancer statistics, 2010. *CA Cancer J Clin* **60,** 277–300 (2010).
78. Brambilla, E., Travis, W. D., Colby, T., Corrin, B. & Shimosato, Y. The new World Health Organization classification of lung tumours. *Eur Respir J* **18,** 1059–1068 (2001).
79. Scott, W. J., Howington, J., Feigenberg, S., Movsas, B. & Pisters, K. Treatment of non-small cell lung cancer stage I and stage II: ACCP evidence-based clinical practice guidelines. *Chest* **132,** 234S–242S (2007).
80. van Zandwijk, N. Neoadjuvant strategies for non-small cell lung cancer. *Lung Cancer* **34,** S145–S150 (2001).
81. Prensner, J. R. & Chinnaiyan, A. M. The emergence of lncRNAs in cancer biology. *Cancer Discov* **1,** 391–407 (2011).
82. Gutschner, T. & Diederichs, S. The hallmarks of cancer: a long non-coding RNA point of view. *RNA Biol* **9,** 703–719 (2012).
83. Qi, P. & Du, X. The long non-coding RNAs, a new cancer diagnostic and therapeutic gold mine. *Mod Pathol* **26,** 155–165 (2013).
84. Gutschner, T. *et al.* The noncoding RNA MALAT1 is a critical regulator of the metastasis phenotype of lung cancer cells. *Cancer Res* **73,** 1180–1189 (2013).
85. Jiang, Y. J. & Bikle, D. D. LncRNA: a new player in 1α, 25 (OH) 2 vitamin D3/VDR protection against skin cancer formation. *Exp Dermatol* **23,** 147–150 (2014).
86. Ji, P., Diederichss & Wang, W. MALAT-1, a novel noncoding RNA, and thymosin beta4 predict metastasis and survival in early-stage non-small cell lung cancer. *Oncogene* **22,** 8031–8041 (2003).
87. Tano, K. *et al.* MALAT-1 enhances cell motility of lung adenocarcinoma cells by influencing the expression of motility-related genes. *FEBS Lett* **584,** 4575–4580 (2010).
88. Zhang, E. *et al.* P53-regulated long non-coding RNA TUG1 affects cell proliferation in human non-small cell lung cancer, partly through epigenetically regulating HOXB7 expression. *Cell Death Dis* **5,** e1243 (2014).

89. Shi, X. *et al.* A critical role for the long non-coding RNA GAS5 in proliferation and apoptosis in non-small-cell lung cancer. *Mol Carcinog* **54**, E1–E12 (2013).

90. Han, D. *et al.* Long noncoding RNAs: novel players in colorectal cancer. *Cancer lett* **361**, 13–21 (2015).

91. Jiao, S. *et al.* Estimating the heritability of colorectal cancer. *Hum Mol Genet* **23**, 3898–3905 (2014).

92. Parkin, D. M., Bray, F., Ferlay, J. & Pisani, P. Global cancer statistics, 2002. *CA Cancer J Clin* **55**, 74–108 (2005).

93. Wood, L. D. *et al.* The genomic landscapes of human breast and colorectal cancers. *Science* **318**, 1108–1113 (2007).

94. Brim, H. *et al.* Genomic aberrations in an African American colorectal cancer cohort reveals a MSI-specific profile and chromosome X amplification in male patients. *PLoS One* **7**, e40392 (2012).

95. Lassmann, S. *et al.* Array CGH identifies distinct DNA copy number profiles of oncogenes and tumor suppressor genes in chromosomal-and microsatellite-unstable sporadic colorectal carcinomas. *J Mol Med (Berl)* **85**, 293–304 (2007).

96. Xu, C., Yang, M., Tian, J., Wang, X. & Li, Z. MALAT-1: a long non-coding RNA and its important 3′end functional motif in colorectal cancer metastasis. *Int J Oncol* **39**, 169–175 (2011).

97. Yang, M.-H. *et al.* MALAT1 promotes colorectal cancer cell proliferation/migration/invasion via PRKA kinase anchor protein 9. *Biochim Biophys Acta* **1852**, 166–174 (2015).

98. Zheng, H.-T. *et al.* High expression of lncRNA MALAT1 suggests a biomarker of poor prognosis in colorectal cancer. *Int J Clin Exp Pathol* **7**, 3174–3181 (2014).

99. Ji, Q. *et al.* Long non-coding RNA MALAT1 promotes tumour growth and metastasis in colorectal cancer through binding to SFPQ and releasing oncogene PTBP2 from SFPQ/PTBP2 complex. *Br J Cancer* **111**, 736–748 (2014).

100. Ji, Q. *et al.* Resveratrol inhibits invasion and metastasis of colorectal cancer cells via MALAT1 mediated Wnt/β-catenin signal pathway. *PLoS One* **8**, e78700 (2013).

101. Yoshimizu, T. *et al.* The H19 locus acts *in vivo* as a tumor suppressor. *Proc Natl Acad Sci USA* **105**, 12417–12422 (2008).

102. Miroglio, A. *et al.* Specific hypomethylated CpGs at the IGF2 locus act as an epigenetic biomarker for familial adenomatous polyposis colorectal cancer. *Epigenomics* **2**, 365–375 (2010).

103. Tian, F. *et al.* Loss of imprinting of IGF2 correlates with hypomethylation of the H19 differentially methylated region in the tumor tissue of colorectal cancer patients. *Mol Med Rep* **5**, 1536–1540 (2012).

104. Cui, H. *et al.* Loss of imprinting in colorectal cancer linked to hypomethylation of H19 and IGF2. *Cancer Res* **62**, 6442–6446 (2002).

105. Nakagawa, H. *et al.* Loss of imprinting of the insulin-like growth factor II gene occurs by biallelic methylation in a core region of H19-associated CTCF-binding sites in colorectal cancer. *Proc Natl Acad Sci USA* **98**, 591–596 (2001).

106. Tsang, W. P. *et al.* Oncofetal H19-derived miR-675 regulates tumor suppressor RB in human colorectal cancer. *Carcinogenesis* **31**, 350–358 (2010).

107. Tanaka, K. *et al.* Loss of imprinting of long QT intronic transcript 1 in colorectal cancer. *Oncology* **60**, 268–273 (2000).

108. Wang, E. *et al.* Predictive genomics: A cancer hallmark network framework for predicting tumor clinical phenotypes using genome sequencing data. *Semin Cancer Biol* **30**, 4–12 (2015).

109. Wang, E. Understanding genomic alterations in cancer genomes using an integrative network approach. *Cancer letters* **340**, 261–269 (2013).

110. Li, Y. *et al.* HMDD v2.0: a database for experimentally supported human microRNA and disease associations. *Nucleic Acids Res* **42**, D1070–D1074 (2014).

111. 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* **42**, D92–D97 (2013).

112. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* **57**, 289–300 (1995).

113. 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* **26**, 1644–1650 (2010).

Acknowledgements

Author Contributions

X.C. conceived the project, developed the prediction method, designed and implemented the experiments, analyzed the result, and wrote the paper.

Additional Information

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Chen, X. Predicting lncRNA-disease associations and constructing lncRNA functional similarity network based on the information of miRNA. *Sci. Rep.* **5**, 13186; doi: 10.1038/srep13186 (2015).