Briefings in Bioinformatics, 20(4), 2019, 1403–1419

doi: 10.1093/bib/bby008
Advance Access Publication Date: 1 February 2018
Paper

Inferring and analyzing module-specific lncRNA–mRNA causal regulatory networks in human cancer

Junpeng Zhang, Thuc Duy Le, Lin Liu and Jiuyong Li

Corresponding authors: Junpeng Zhang, School of Engineering, Dali University, Dali, Yunnan 671003, Public Republic of China. Tel.: +86 872 2219799; Fax: +86 872 2219799. E-mail: zhangjunpeng_411@yahoo.com; Jiuyong Li, School of Information Technology and Mathematical Sciences, University of South Australia, Mawson Lakes, 5095, SA, Australia. Tel.: +61 8 830 23898; Fax: +61 8 830 23381. E-mail: jiuyong.li@unisa.edu.au

Abstract

It is known that noncoding RNAs (ncRNAs) cover ~98% of the transcriptome, but do not encode proteins. Among ncRNAs, long noncoding RNAs (lncRNAs) are a large and diverse class of RNA molecules, and are thought to be a gold mine of potential oncogenes, anti-oncogenes and new biomarkers. Although only a minority of lncRNAs is functionally characterized, it is clear that they are important regulators to modulate gene expression and involve in many biological functions. To reveal the functions and regulatory mechanisms of lncRNAs, it is vital to understand how lncRNAs regulate their target genes for implementing specific biological functions. In this article, we review the computational methods for inferring lncRNA–mRNA interactions and the third-party databases of storing lncRNA–mRNA regulatory relationships. We have found that the existing methods are based on statistical correlations between the gene expression levels of lncRNAs and mRNAs, and may not reveal gene regulatory relationships which are causal relationships. Moreover, these methods do not consider the modularity of lncRNA–mRNA regulatory networks, and thus, the networks identified are not module-specific. To address the above two issues, we propose a novel method, MSLCRN, to infer and analyze module-specific lncRNA–mRNA causal regulatory networks. We have applied it into glioblastoma multiforme, lung squamous cell carcinoma, ovarian cancer and prostate cancer, respectively. The experimental results show that MSLCRN, as an expression-based method, could be a useful complementary method to study lncRNA regulations.

Key words: lncRNA; mRNA; lncRNA–mRNA co-expression; lncRNA–mRNA interaction; lncRNA–mRNA causal relationship; human cancer

Junpeng Zhang is an associate professor at the School of Engineering, Dali University. He received his BSc (2009) in Bio-medical Engineering and MSc (2012) in Control Theory and Control Engineering from Kunming University of Science and Technology, Kunming City, China. His research interests include bioinformatics and data mining.

Thuc Duy Le is a research fellow at the University of South Australia (UniSA). He received his BSc (2002) and MSc (2006) in pure Mathematics from the University of Pedagogy, Ho Chi Minh City, Vietnam, and BSc (2010) in Computer Science from UniSA. He received his PhD degree in Computer Science (Bioinformatics) in 2014 at UniSA. His research interests are bioinformatics, data mining and machine learning.

Lin Liu is a senior lecturer at the School of Information Technology and Mathematical Sciences, University of South Australia (UniSA). She received her bachelor and master degrees in Electronic Engineering from Xidian University, China, in 1991 and 1994, respectively, and her PhD degree in computer systems engineering from UniSA in 2006. Her research interests include data mining and bioinformatics, as well as Petri nets and their applications to protocol verification and network security analysis.

Jiuyong Li is a professor at the School of Information Technology and Mathematical Sciences, University of South Australia. He received his PhD degree in computer science from the Griffith University, Australia (2002). His research interests are in the fields of data mining, privacy preserving and bioinformatics. His research has been supported by five prestigious Australian Research Council Discovery grants since 2005 and he has published more than 100 research papers.

Submitted: 13 December 2017; **Received (in revised form):** 8 January 2018

Introduction

Long noncoding RNAs (lncRNAs) are non-protein coding transcripts with >200 nucleotides in length. Unlike small noncoding RNAs (sncRNAs), lncRNAs generally exhibit low sequence conservation. However, owing to rapidly adaptive selection pressures, the low conservation of lncRNAs (such as Air and Xist) does not indicate absence of function [1]. Similar to microRNAs (miRNAs), an important class of sncRNAs, evidence has shown that lncRNAs play important roles in a wide range of biological processes, even in cancers [2, 3]. Despite the importance of lncRNAs, in many physiological and pathological processes, a large number of lncRNAs remain to be functionally characterized. For this reason, the number of studies on lncRNA research has been increased exponentially in the past decade (as shown in Figure 1).

To achieve various biological functions, lncRNAs form gene regulatory networks by interacting with other biological molecules, such as transcription factors, miRNAs, messenger RNAs (mRNAs) and RNA-binding proteins [4]. Among these biological molecules interacting with lncRNAs, mRNAs are the most popular ones. By regulating the transcription and translation of mRNAs, lncRNAs could get involved in several vital biological processes, such as cell differentiation, cell proliferation and cytoprotective programs [5]. Therefore, the identification of lncRNA–mRNA regulatory networks would help to uncover functions and regulatory mechanisms of lncRNAs.

A straightforward method for identifying lncRNA–mRNA regulatory networks is sequence-based complementary base pairing. To predict lncRNA targets, several sequence-based methods, such as GUUGle [6], RNAup [7], RNAplex [8], IntaRNA [9], RactIP [10], LncTar [11] and RIblast [12], have been developed. Owing to the long sequence and complex tertiary structure of each lncRNA, the computational costs of predicting large-scale lncRNA–mRNA regulatory relationships are usually high. Moreover, these sequence-based methods only consider the sequence information of lncRNAs and target mRNAs, and thus, the predicted lncRNA–mRNA regulatory networks are static. However, previous studies [13–15] have shown that lncRNAs exhibit condition-specific expression fashion and dynamic networks of gene regulation. To identify dynamic or condition-specific lncRNA–mRNA regulatory networks, it is necessary to use expression data. Some expression-based methods [16–19] for predicting co-expressed lncRNA–mRNA networks have been proposed. However, as the predictions are based on statistical associations found in gene expression levels only, they may not represent the real 'causal' lncRNA–mRNA regulatory relationships. Furthermore, the existing expression-based methods do not consider the modularity of lncRNA–mRNA regulatory networks, an important feature of gene regulatory networks [20].

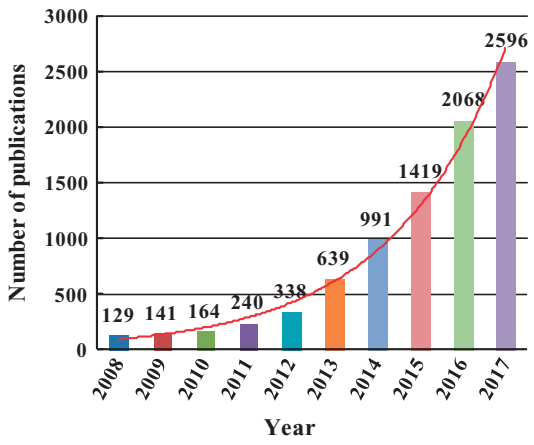


Figure 1. The number of lncRNA-related publications in the past decade. The number of queried publications is obtained from PubMed library with keyword 'lncRNA'.

In this article, we first review the computational methods for inferring lncRNA–mRNA interactions and the public databases for storing lncRNA–mRNA regulatory relationships. Second, we propose a novel method to infer Module-Specific LncRNA–mRNA Causal Regulatory Network (thus the proposed method is called MSLCRN). In the first step, by considering modularity of networks, MSLCRN uses Weighted Gene Co-expression Network Analysis (WGCNA) [21] to identify lncRNA–mRNA co-expression modules. In each module, the lncRNAs and mRNAs are regarded as module-specific genes. In the second step, MSLCRN uses a causal inference method named intervention calculus when the directed acyclic graph (DAG) is absent (IDA) [22, 23] to estimate the causal effects of possible lncRNA–mRNA causal pairs in each module. To speed up the estimation, the parallelized version of IDA [24] is used to calculate the causal effects. For each module, the noncausal lncRNA–mRNA pairs are eliminated, and the retained lncRNA–mRNA causal pairs are further assembled to generate a module-specific lncRNA–mRNA causal network. To obtain a global lncRNA–mRNA causal regulatory network, we further integrate the identified module-specific lncRNA–mRNA causal networks in the third step.

To evaluate MSLCRN, we have applied it into four human cancer data sets, including glioblastoma multiforme (GBM), lung squamous cell carcinoma (LSCC), ovarian cancer (OvCa) and prostate cancer (PrCa) from [25]. The validation, survival and enrichment analysis results show that the proposed method can help with revealing the functions and regulatory mechanisms of lncRNAs. MSLCRN is released under the GPL-3.0 License, and is freely available through GitHub (https://github.com/zhangjunpeng411/MSLCRN).

Computational methods for inferring lncRNA–mRNA interactions

In this section, we review the computational approaches for inferring lncRNA–mRNA interactions. In Table 1, we divide the methods into two categories: (1) sequence-based method, and (2) expression-based method. We will separately review these methods as follows.

Sequence-based method

The common characteristic of the sequence-based methods is that the identification of RNA–RNA interactions depends on RNA binding energy between two RNA molecules. To evaluate the strength of RNA binding energy, a number of energy models [6–12, 26–32] are proposed to predict RNA–RNA interactions.

Gerlach and Giegerich [6] propose a utility program GUUGle for locating potential helical regions under RNA complementary base pairs rules. The method can be effectively used as a filter for noncoding RNA (ncRNA) target prediction. However, the reliable prediction of RNA–RNA binding energies is also important for the identification of RNA–RNA interactions. To study the thermodynamics of RNA–RNA interactions, Mückstein *et al.* [7] present an extension of the standard partition function method called RNAup to RNA secondary structures. By comparing

Table 1. Summary of computational methods or tools for inferring lncRNA–mRNA interactions

Methods/tools	Categories of methods	Brief descriptions	Available
GUUGle [6]	Sequence-based	Target prediction by locating potential helical regions of RNA–RNA pairs under RNA base pairing rules, which include G-U bases	http://bibiserv2.cebitec.uni-bielefeld.de/guugle
RNAup [7]	Sequence-based	Target prediction by studying thermodynamics of RNA–RNA pairs based on the sum of the energy of binding and hybridization	http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAup.cgi
RNAcofold [26]	Sequence-based	Target prediction by computing the hybridization energy and base pairing pattern of RNA–RNA pairs	http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAcofold.cgi
Alkan *et al.* [27]	Sequence-based	Target prediction by minimizing the joint free energy of RNA–RNA pairs under a number of energy models, including base pair energy model, stacked pair energy model, loop energy model	On request
RNAplex [8]	Sequence-based	Target prediction by finding possible hybridization sites of RNA–RNA pairs	http://www.tbi.univie.ac.at/~htafer/
IntaRNA [9]	Sequence-based	Target prediction by incorporating accessibility of target sites as well as the existence of a user-definable seed	http://rna.informatik.uni-freiburg.de/IntaRNA/Input.jsp
RactIP [10]	Sequence-based	Target prediction by integrating approximate information on an ensemble of equilibrium joint structures into the objective function of integer programming	http://rtips.dna.bio.keio.ac.jp/ractip/
PETcofold [28]	Sequence-based	Target prediction by taking covariance information in intramolecular and intermolecular base pairs into account	http://rth.dk/resources/petcofold
RIsearch [29]	Sequence-based	Target prediction by implementing a simplified Turner energy model for fast computation of hybridization	https://rth.dk/resources/risearch/risearch1.php
RIsearch2 [30]	Sequence-based	An updated version of RIsearch, and predict targets using a single integrated seed-and-extend framework based on suffix arrays	https://rth.dk/resources/risearch
LncTar [11]	Sequence-based	lncRNA target prediction by finding the minimum free energy joint structure of RNA–RNA pairs based on base pairing	http://www.cuilab.cn/lnctar
lncRNATargets [31]	Sequence-based	lncRNA target prediction based on nucleic acid thermodynamics	http://www.herbbol.org: 8001/lrt/
Terai *et al.* [32]	Sequence-based	lncRNA target prediction by developing an integrated pipeline on the K computer, which is one of the fastest super-computers in the world	http://rtools.cbrc.jp/cgi-bin/RNARNA/index.pl
RIblast [12]	Sequence-based	Target prediction based on the seed-and-extension approach	http://github.com/fukunagatsu/RIblast
Liao *et al.* [16]	Expression-based	Identify lncRNA–mRNA interactions by using Pearson method, and the identified lncRNA–mRNA interactions should be co-expressed in the same direction in no less than 3 Mouse microarray data sets.	On request
Guo *et al.* [17]	Expression-based	Identify lncRNA–mRNA interactions by using Pearson method in OvCa malignant progression	On request
Du *et al.* [18]	Expression-based	Identify lncRNA–mRNA interactions by using Pearson method and a power function in thyroid cancer	On request
Liu *et al.* [33]	Expression-based	Identify lncRNA–mRNA interactions by using Pearson method in human colorectal carcinoma	On request
Huang *et al.* [34]	Expression-based	Identify lncRNA–mRNA interactions associated with pneumonia by using Pearson method	On request
Li *et al.* [35]	Expression-based	Identify dynamic lncRNA–mRNA interactions associated with venous congestion by using Pearson method	On request
Wu *et al.* [19]	Expression-based	Identify lncRNA–mRNA interactions by using a generalized linear model to regress mRNA expression on lncRNA expression in breast cancer	On request
Fu *et al.* [36]	Expression-based	Identify lncRNA–mRNA interactions by considering mRNA loci within lncRNA and the Pearson correlation in cartilage	On request
Zhang *et al.* [37]	Expression-based	Identify lncRNA–mRNA interactions by considering mRNA loci within lncRNA and the Pearson correlation in cartilage in peripheral blood mononuclear cells	On request
Iwakiri *et al.* [38]	Expression-based	Identify tissue-specific lncRNA–mRNA interactions by integrating the tissue specificity of lncRNAs and mRNAs into sequence-based prediction of human lncRNA–RNA interactions	On request
Lv *et al.* [39]	Expression-based	Identify tissue-specific lncRNA–mRNA interactions by using Pearson and sequence-based methods and in human intrahepatic cholangiocarcinoma	On request

predicted free energies of binding with RNA interference experimental data, RNAup can produce biologically reasonable results. For genome-wide predictions of ncRNA targets, RNAup is not fast enough. Therefore, it is usually to be combined with other faster RNA–RNA prediction methods.

To extend the standard dynamic programming algorithms for computing RNA secondary structures, Bernhart *et al.* [26] propose a program named RNAcofold to compute the hybridization energy and base pairing pattern of the co-folding of two RNA molecules. However, the method disregards some important interaction structures, and is restricted to dimeric complexes. Moreover, for the RNA–RNA interaction prediction, predicting the joint secondary structure of two interacting RNAs is also important. To solve it, Alkan *et al.* [27] develop several algorithms to minimize the joint free energy between the two RNAs under a number of energy models. Assuming that conserved RNA–RNA interactions imply conserved function, Seemann *et al.* [28] also implement a comparative method called PETcofold to predict the joint secondary structure of two interacting RNAs. As PETcofold considers sequence conservation, an increasing amount of structural covariance can further improve its performance.

RNAup [7] and RNAcofold [26] are too slow for genome-wide search in finding target sites of ncRNAs. To accelerate the speed of RNA–RNA interaction predictions, RNAplex [8] is presented to quickly find possible hybridization sites between two interacting RNAs. To focus on the target search on short highly stable interactions, RNAplex introduces a per nucleotide penalty. Meanwhile, another general and fast approach IntaRNA [9] is proposed to efficiently predict bacterial RNA–RNA interactions. Compared with other existing target prediction methods, IntaRNA considers both the accessibility of target sites and the existence of a user-defined seed. Therefore, it shows a higher accuracy than competing methods. Kato *et al.* [10] also present a fast and accurate prediction method RactIP for comprehensive type of RNA–RNA interactions. In terms of predicting joint secondary structures of two interacting RNAs, RactIP run incomparably faster than competitive programs.

To further achieve a speed improvement of predicting RNA–RNA interactions, Wenzel *et al.* [29] present RIsearch for fast computation of hybridization between two interacting RNAs. They show that the energy model of RIsearch is an accurate approximation of the full energy model for near-complementary RNA–RNA duplexes. Furthermore, RIsearch is faster than RNAplex [8] in RNA–RNA interaction search. Recently, RIsearch2 [30], an updated version of RIsearch [29], is proposed to localize potential near-complementary RNA–RNA interactions between two RNA sequences. The comparison results show that RIsearch2 is much faster than the previous methods, such as GUUGle [6], RNAplex [8], IntaRNA [9] and RIsearch [29].

Although the above RNA–RNA interaction prediction methods can be extended to predict lncRNA–mRNA interactions, none of them are exclusively used for identifying the RNA targets of lncRNAs in a large scale. To efficiently identify lncRNA–mRNA interactions, Li *et al.* [11] propose a tool named LncTar. LncTar explores lncRNA–mRNA interactions by finding the minimum free energy joint structure of two interacting RNAs based on base pairing. As LncTar runs fast and does not have a limit to RNA size, it can be used for large-scale identification of the RNA targets for all RNAs. Another web-based platform lncRNATargets [31] is also provided for lncRNA target prediction. Because there is no limit to RNA size, lncRNATargets can also be used to identify the RNA targets of all RNAs. In a whole human transcriptome, Terai *et al.* [32] develop an integrated pipeline to predict lncRNA–mRNA interactions for the first time. In the pipeline, IntaRNA [9] is used to calculate interaction energy, and RactIP [10] is used to predict joint secondary structure. Recently, to further shorten the running time of predicting lncRNA–mRNA interactions, an ultrafast RNA–RNA interaction prediction method RIblast [12] based on the seed-and-extension method is presented. The comparison results show that RIblast runs faster than RNAplex [8], IntaRNA [9], Terai *et al.* pipeline [32], and thus can be applied to a large scale of lncRNA target identification.

Expression-based method

At the gene expression level, the co-expressed lncRNA–mRNA pairs are regarded as lncRNA–mRNA interactions for the expression-based methods. Among the existing expression-based methods [16–19, 33–39], Pearson correlation method is a key step of most methods to identify co-expressed lncRNA–mRNA pairs.

Liao *et al.* [16] construct a lncRNA–mRNA co-expression network from re-annotated mouse microarray data sets. By using Pearson method, they only keep the lncRNA–mRNA pairs with $P < 0.01$ and Pearson correlation ranked in the top or bottom 0.05 percentile. The study is the first large-scale prediction of lncRNA functions from a lncRNA–mRNA co-expression network. To identify immune-associated lncRNA biomarkers in OvCa, Guo *et al.* [17] make a comprehensive analysis of lncRNA–mRNA co-expression patterns. To identify lncRNA–mRNA co-expression pairs, they calculate Pearson correlation between differentially expressed lncRNAs and mRNAs. They only reserve the lncRNA–mRNA co-expression pairs with Pearson correlation > 0.5 and the corresponding False Discovery Rate (FDR) < 0.01. Liu *et al.* [33] and Huang *et al.* [34] also use Pearson method to study lncRNA–mRNA co-expression networks in human colorectal carcinoma and pneumonia, respectively. The inferred lncRNA–mRNA co-expression networks will help to study lncRNA functions. Recently, Du *et al.* [18] propose a two-step method to conduct a comprehensive analysis of lncRNA–mRNA co-expression patterns in thyroid cancer. First, they use Pearson method to calculate Pearson correlation, and the cutoff of Pearson correlation is 0.5 and the corresponding FDR cutoff is 0.01. Second, the Pearson correlations are transformed into an adjacency matrix.

Owing to dynamic characteristic of gene regulatory networks, Wu *et al.* [19] identify two distinct lncRNA–mRNA co-expression networks in tumor and normal breast tissue. They use a generalized linear model to regress mRNA expression on lncRNA expression in tumor and normal breast tissue, and only focus on dynamic breast lncRNA–mRNA co-expression pairs that differ in tumor and normal breast tissue. Meanwhile, to study the potential role of lncRNAs in venous congestion, Li *et al.* [35] also construct a dynamic lncRNA–mRNA co-expression network. By using Pearson method, they separately calculate Pearson correlations of each lncRNA–mRNA pair in venous congestion and normal samples. The lncRNA–mRNA pairs with Pearson correlation >0.99 or <-0.99 and P-value <0.01 are selected as lncRNA–mRNA co-expression pairs. They construct two types of lncRNA–mRNA co-expression networks: 'lost' network where lncRNA–mRNA co-expression pairs only existed in normal samples, and 'obtained' network where lncRNA–mRNA co-expression pairs only existed in venous congestion samples. The 'lost' and 'obtained' networks are further integrated to obtain a dynamic lncRNA–mRNA co-expression network.

The above methods simply use matched lncRNA and mRNA expression data to identify lncRNA–mRNA co-expression pairs. To identify 'cis-regulated target genes' of lncRNAs, some methods also consider mRNA loci information within lncRNA. For example, Fu *et al.* [36] combine mRNA loci information and matched lncRNA and mRNA expression data to predict lncRNA targets. They identify the mRNAs as targets under two conditions: (i) the mRNA loci are within a 300-kb window up- or downstream of lncRNA, and (ii) lncRNA–mRNA co-expression pairs are significantly positive correlated (Pearson correlation > 0.8 and the corresponding P-value < 0.05). Zhang *et al.* [37] also use a similar method to Fu *et al.* [36] for identifying lncRNA targets. The mRNAs can be regarded as targets when (1) the mRNA loci are within a 10 window up- or downstream of lncRNA, and (2) lncRNA–mRNA co-expression pairs are significantly positive correlated (Pearson correlation > 0.98 and the corresponding P-value < 0.05).

Apart from mRNA loci information within lncRNA, some emerging methods consider predictions from sequence-based methods as putative lncRNA–mRNA interactions. For example, Iwakiri *et al.* [38] integrate tissue-specific lncRNA and mRNA expression data into predictions from a sequence-based method in [32]. They discover that integrating tissue specificity can improve prediction accuracy of lncRNA–mRNA interactions. Lv *et al.* [39] also combine matched lncRNA and mRNA expression data with predictions from a sequence-based method LncTar [11]. They first use Pearson method to identify co-expressed lncRNA–mRNA co-expression pairs with Pearson correlation >0.95 or <-0.95. Then, LncTar is used to further filter the identified lncRNA–mRNA co-expression pairs.

Public databases for storing lncRNA–mRNA regulatory relationships

In this section, we review the public databases of storing lncRNA–mRNA regulatory relationships. Table 2 shows a summary of the third-party public databases, including experimentally validated and computationally predicted databases.

NPInter [40] contains experimentally validated interactions between ncRNAs, especially lncRNAs and miRNAs. The database contains 915 067 interactions in 188 tissues or cell lines from 68 kinds of experimental technologies. There is a classification of the functional interactions based on the functional process that ncRNA is involved in. Moreover, NPInter allows users to search interactions, related publications and other information.

LncRNADisease [41] not only collects experimentally supported lncRNA–disease associations and lncRNA interactions, but also predicts novel lncRNA–disease associations. Recently, the database curates 478 entries of experimentally validated lncRNA interactions. LncRNADisease provides users several ways to search lncRNA-related diseases and interactions.

To study differentially expressed genes after lncRNA knockdown or overexpression, Jiang *et al.* [42] develop a database called LncRNA2Target in human and mouse organisms. The database has a collection of 396 experimentally validated lncRNA–target interactions. In LncRNA2Target, if a gene is differentially expressed after lncRNA knockdown or overexpression, it is regarded as a target of a lncRNA. For convenience, LncRNA2Target allows users to search for the targets of single lncRNA or for the lncRNAs that target a specific gene. Meanwhile, Zhou *et al.* [43] also build a reference resource LncReg for lncRNA-related regulatory networks. The database has 1,081 experimentally validated lncRNA-related regulatory records between 258 nonredundant lncRNAs and 571 nonredundant genes.

IRNdb [44] is a database that focuses on collecting immunologically relevant lncRNA–target, miRNA–target and PIWI-interacting RNA–target interactions. The current version of IRNdb documents 22 453 immunologically relevant lncRNA–target interactions by integrating three databases: LncRNADisease [41], LncRNA2Target [42] and LncReg [43]. The aim is to help researchers study the roles of ncRNAs in the immune system. Recently, a new experimentally validated database named lncRInter [45] was developed to collect reliable and high-quality lncRNA–target interactions. The extracted lncRNA–target interactions are all from published literature, and are supported by certain biological experiments (e.g. luciferase reporter assay, *in vitro* binding assay, RNA pull-down). In total, lncRInter contains 1036 experimentally validated lncRNA–target interactions in 15 organisms.

In addition to the experimentally validated databases presented above, there are several computationally predicted databases for collecting lncRNA–mRNA interactions. For instance, starBase [46] is a comprehensive database of systematically identifying the RNA–RNA and protein–RNA interaction networks from 108 CLIP-Seq (PAR-CLIP, HITS-CLIP, iCLIP, CLASH) data sets. The lncRNA–mRNA interactions can be extracted from protein–RNA interaction networks. InCaNet [47] aims to establish a comprehensive regulatory network between lncRNAs and cancer genes. They identify lncRNA–cancer gene interactions by computing gene co-expression between lncRNAs and cancer genes. BmncRNAdb [48] is a comprehensive database of silkworm lncRNAs and miRNAs. The database provides three online tools for users to predict both lncRNA–target and miRNA–target interactions. lncRNAtor [49] collect expression data from 243 RNA-seq experiments including 5237 samples of various tissues and developmental stages. The lncRNA–mRNA co-expression pairs are identified through co-expression analysis of lncRNAs and mRNAs. lncRNome [50] is a comprehensive knowledgebase of sequence, structure, biological functions, genomic variations and epigenetic modifications on $>17\,000$ lncRNAs in human. For lncRNA–protein interactions, the database incorporates PAR-CLIP experiments and a support vector machine-based prediction method. Co-lncRNA [51] and LncRNA2Function [52] predict co-expressed lncRNA–mRNA interactions from RNA-Seq data, and further annotates the potential functions of human lncRNAs using functional enrichment analysis. lncRNAMap [53] is an integrated and comprehensive database to explore regulatory functions of human lncRNAs. By integrating small RNAs supported by publicly available deep sequencing data, lncRNAMap construct lncRNA-derived siRNA–target interactions.

In summary, for experimentally validated databases, users can select individual database or combine several databases as ground truth to validate the predicted lncRNA–mRNA interactions. As for computationally predicted databases, they can be used as initial structural of sequence-based or expression-based methods to identify lncRNA–mRNA interactions.

Inferring and analyzing MSLCRN networks

Repurposed microarray data across human cancers

We collect the repurposed lncRNA and mRNA expression data of GBM, LSCC, OvCa and PrCa from [25]. A lncRNA or mRNA is eliminated if it does not have a corresponding gene symbol in a data set. By calculating average expression values of replicate

Table 2. Public databases for storing lncRNA–mRNA regulatory relationships

Databases	Types of databases	Brief descriptions	Organisms	Available
NPInter [40]	Validated	A database of experimentally verified functional interactions between ncRNAs (including lncRNAs, miRNAs, etc) and biomolecules (proteins, RNAs and DNAs)	22 organisms	http://www.bioinfo.org/NPInter/
LncRNADisease [41]	Validated	A database of experimentally supported lncRNA–disease association data and lncRNA–target interactions in various levels, including protein, RNA, miRNA and DNA	Human	http://www.cuilab.cn/lncrnadisease
LncRNA2Target [42]	Validated	A database of lncRNA–target regulatory relationships experimentally validated by lncRNA knockdown or overexpression	Human, mouse	http://www.lncrna2target.org/
LncReg [43]	Validated	A database of experimentally validated lncRNA–target interactions from public literature	7 organisms	http://bioinformatics.ustc.edu.cn/lncreg/
IRNdb [44]	Validated	A database of immunologically relevant ncRNAs (miRNAs, lncRNAs and other ncRNAs) and target genes	Human, mouse	http://compbio.massey.ac.nz/apps/irndb
lncRInter [45]	Validated	A database of experimentally validated lncRNA–target interactions extracted from peer-reviewed publications	15 organisms	http://bioinfo.life.hust.edu.cn/lncRInter/
starBase [46]	Predicted	A comprehensive database of systematically identifying the RNA–RNA and protein–RNA interaction networks from 108 CLIP-Seq (PAR-CLIP, HITS-CLIP, iCLIP, CLASH) data sets	Human	http://starbase.sysu.edu.cn/
lnCaNet [47]	Predicted	A database of establishing a comprehensive regulatory network source for lncRNA and cancer genes	Human	http://lncanet.bioinfo-minzhao.org/
BmncRNAdb [48]	Predicted	A comprehensive database of the silkworm lncRNAs and miRNAs, as well as the three online tools for users to predict the target genes of lncRNAs or miRNAs	*Bombyx mori*	http://gene.cqu.edu.cn/BmncRNAdb/index.php
lncRNAtor [49]	Predicted	A comprehensive resource of encompassing annotation, sequence analysis, gene expression, protein binding and phylogenetic conservation	6 organisms	http://lncrnator.ewha.ac.kr/
lncRNome [50]	Predicted	A comprehensive knowledgebase on the types, chromosomal locations, description on the biological functions and disease associations of lncRNAs	Human	http://genome.igib.res.in/lncRNome/
Co-LncRNA [51]	Predicted	A computationally predicted database to identify GO annotations and KEGG pathways affected by co-expressed protein-coding genes of a single or multiple lncRNAs	Human	http://www.bio-bigdata.com/Co-LncRNA/
LncRNA2Function [52]	Predicted	A comprehensive resource of investigating the functions of lncRNAs based on co-expressed lncRNA–mRNA interactions	Human	http://mlg.hit.edu.cn/lncrna2function/
lncRNAMap [53]	Predicted	An integrated and comprehensive database of regulatory functions of lncRNAs and acting as ceRNAs	Human	http://lncRNAMap.mbc.nctu.edu.tw/

lncRNAs and mRNAs, we obtain unique expression value of these replicates. Consequently, we get the matched expression data of 9704 lncRNAs and 18 282 mRNAs in 451 GBM, 113 LSCC, 585 OvCa and 150 PrCa samples.

Pipeline of MSLCRN

As shown in Figure 2, MSLCRN contains the following three steps to infer module-specific lncRNA–mRNA causal regulatory networks.

i. Identification of lncRNA–mRNA co-expression modules. Given the matched lncRNA and mRNA expression data, we use WGCNA to generate gene co-expression modules. A module containing at least two lncRNAs and two mRNAs are regarded as a lncRNA–mRNA co-expression module, and used as the input of the second step.
ii. Identification of module-specific lncRNA–mRNA causal regulatory networks. For each lncRNA–mRNA co-expression module, with each lncRNA–mRNA pair, we apply parallel IDA to estimate the causal effect of the lncRNA on the

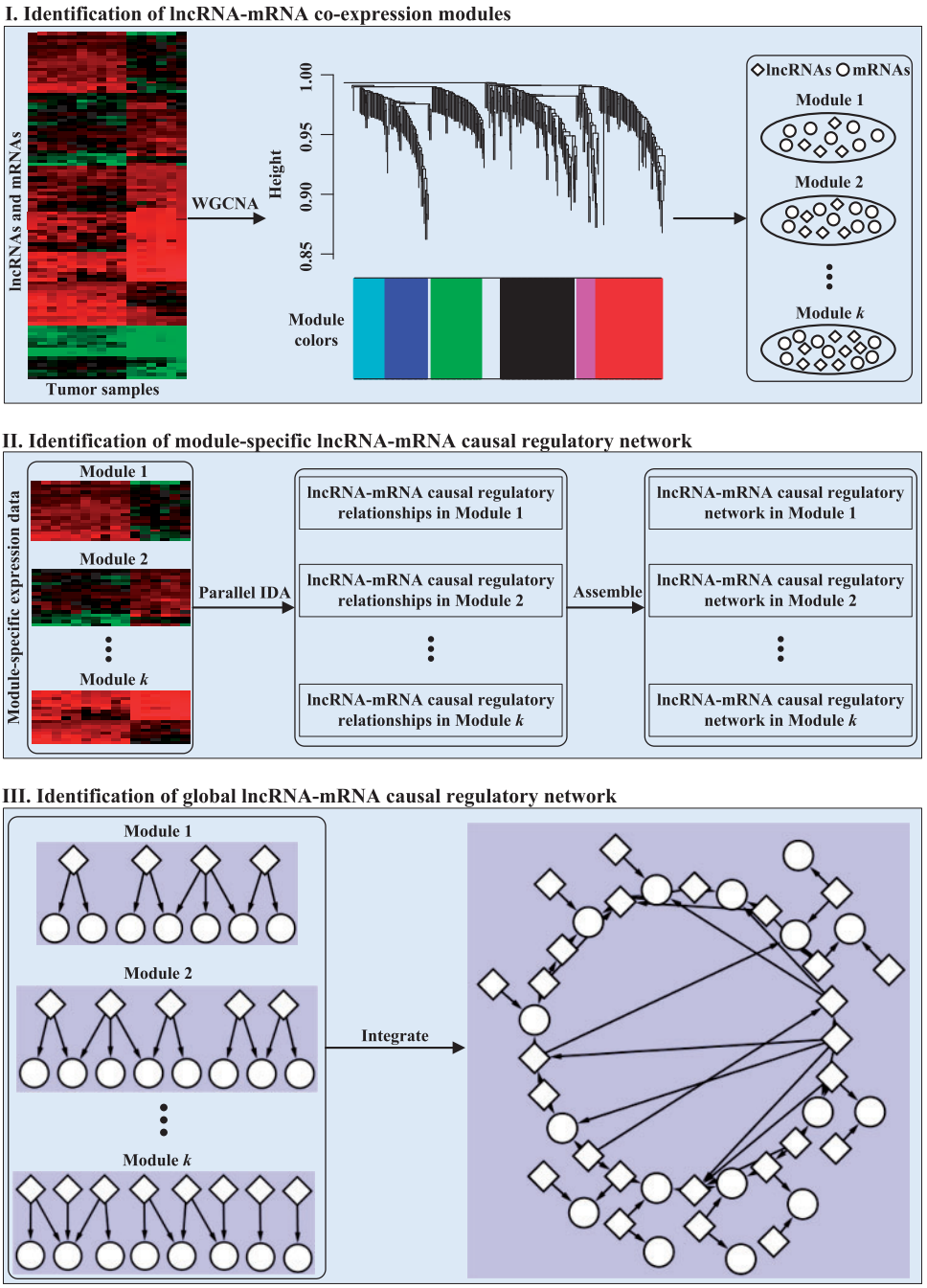


Figure 2. The pipeline of MSLCRN. First, WGCNA is used to identify lncRNA–mRNA co-expression modules from matched lncRNA and mRNA expression data. Second, we infer lncRNA–mRNA causal regulatory relationships in each module by using parallel IDA method. For each module, we assemble the identified lncRNA–mRNA regulatory relationships to obtain a module-specific lncRNA–mRNA causal regulatory network. Third, the module-specific lncRNA–mRNA causal regulatory networks are integrated to form a global lncRNA–mRNA causal regulatory network.

mRNA. We use the absolute value of the causal effect (AVCE) to evaluate the strength of the regulation of the lncRNA on the mRNA, and a higher AVCE indicates a stronger lncRNA regulation. The lncRNA–mRNA pairs with high AVCEs in each module are considered as module-specific lncRNA–mRNA causal regulatory relationships, and we call each module with these relationships identified a module-specific causal regulatory network.

iii. Identification of global lncRNA–mRNA causal regulatory network. We integrate the module-specific lncRNA–mRNA causal regulatory networks to form the global lncRNA–mRNA causal regulatory network.

Identification of lncRNA–mRNA co-expression modules

In systems biology, WGCNA [21] is a popular method for finding the correlation patterns among genes across samples, and can be used to identify clusters or modules of highly co-expressed genes. Therefore, we use WGCNA to first infer lncRNA–mRNA co-expression modules.

Specifically, the matched lncRNA and mRNA expression data are used as the input of WGCNA. For each pair of genes i and j, the gene co-expression similarity s_{ij} of the pair is defined as:

$$s_{ij} = |cor(i, j)| \quad (1)$$

where $|cor(i, j)|$ is the absolute value of the Pearson correlation between genes i and j. The gene co-expression similarity matrix is denoted by $S = [s_{ij}]$.

To pick an appropriate soft-thresholding power for transforming the similarity matrix S into an adjacency matrix A, we use the scale-free topology criterion for soft-thresholding and the minimum scale free topology fitting index R^2 is set as 0.9. Then, the topological overlap matrix (TOM) $W = [w_{ij}]$ is generated based on the adjacency matrix $A = [a_{ij}]$. The TOM similarity w_{ij} between genes i and j is defined:

$$w_{ij} = \frac{\sum_u a_{iu} a_{uj} + a_{ij}}{\min\{\sum_u a_{iu}, \sum_u a_{uj}\} + 1 - a_{ij}} \quad (2)$$

where u denotes all genes of the matched lncRNA and mRNA expression data. The TOM dissimilarity between genes i and j is denoted by $d_{ij} = 1 - w_{ij}$. To identify gene co-expression modules, the TOM dissimilarity matrix $D = [d_{ij}]$ is clustered using optimal hierarchical clustering method [54]. Here, the identified gene co-expression modules are groups of lncRNAs and mRNAs with high topological overlap. The lncRNAs and mRNAs of each lncRNA–mRNA co-expression module are considered for possible lncRNA–mRNA causal relationships in the next step.

Identification of module-specific lncRNA–mRNA causal regulatory networks

After the identification of lncRNA–mRNA co-expression modules, we use the parallel IDA method [24] to estimate causal effects of possible lncRNA–mRNA causal pairs in each module. The application of parallel IDA method to matched lncRNA and mRNA expression data for estimating causal effects includes two steps: (i) learning the causal structure from expression data using the parallel-PC algorithm [24], and (ii) estimating the causal effects of lncRNAs on mRNAs by applying do-calculus [55].

In step (i), $\mathbf{V} = \{L_1, \ldots, L_m, T_1, \ldots, T_n\}$ is a set of random variables denoting m lncRNAs and n mRNAs. The causal structure is in the form of a DAG, where a node denotes a lncRNA L_i or mRNA T_j and an edge between two nodes represents a causal relationship between them. We use the parallel-PC algorithm, a parallel version of the PC algorithm [56], to learn the causal structures (the DAGs) from expression data. Starting with a fully connected undirected graph, the parallel-PC algorithm determines if an edge is retained or removed in the graph by conducting conditional independence tests in parallel. Then, to get a DAG, the directions of edges in the obtained graph are oriented. As different DAGs may represent the same conditional independence, the parallel-PC algorithm uses a completed partially directed acyclic graph (CPDAG) to uniquely describe an equivalence class of DAGs. In this work, we use the R-package *ParallelPC* [57] to implement the parallel-PC algorithm and set the significant level of the conditional independence tests $\alpha = 0.01$.

In step (ii), we are only interested in estimating the causal effect of the directed edge $L_i \rightarrow T_j$, where vertex is L_i a parent of vertex T_j. As described above, a CPDAG may generate a class of DAGs. For the causal effect of $L_i \rightarrow T_j$ in a CPDAG, we use do-calculus [55] to estimate the causal effects of L_i on T_j in a class of DAGs. Then, we use the minimum absolute value of all possible causal effects as a final causal effect of $L_i \rightarrow T_j$. As for the details of how the parallel IDA method is applied to estimate causal relationships from expression data, the readers can refer to [24].

The estimated causal effects can be positive or negative, reflecting the up or down regulation by the lncRNAs on the mRNAs. For the purpose of constructing the regulatory networks, we use the absolute values of the causal effects (AVCEs) to evaluate the strengths of the regulation and thus to confirm the regulatory relationships.

We set different AVCE cutoffs from 0.10 to 0.60 with a step of 0.05, to generate MSLCRN networks in GBM, LSCC, OvCa and PrCa, respectively. For each cutoff, we merge the identified MSLCRN networks to obtain global lncRNA–mRNA causal regulatory networks in the four human cancers, respectively. As shown in Table 3, a higher cutoff selection causes a smaller global lncRNA–mRNA causal regulatory network but better goodness of fit. To make a trade-off between the size of the global lncRNA–mRNA causal regulatory networks and goodness of fit, we set a compromised AVCE cutoff with a value of 0.45. If the AVCE of a lncRNA on a mRNA is 0.45 or above, we consider there is a causal regulatory relationship between the lncRNA–mRNA pair. Under the compromise cutoff, we have a moderate size of the global lncRNA–mRNA causal regulatory networks in GBM, LSCC, OvCa and PrCa. Meanwhile, the node degree distributions of four global lncRNA–mRNA causal regulatory networks also follow power law distribution (the fitted power curve is in the form of $y = ax^b$) well with $R^2 > 0.8$.

Validation, survival and enrichment analysis

Previous studies have demonstrated that about 20% of the nodes in a biological network are essential and are regarded as hub genes [58, 59]. Therefore, when analyzing a global lncRNA–mRNA causal network, we select the 20% of lncRNAs with the highest degrees in the network as hub lncRNAs. The degree of a lncRNA node in the global network is the number of mRNAs connected with it.

To validate the predicted module-specific lncRNA–mRNA causal regulatory relationships, we obtain the experimentally validated lncRNA–mRNA regulatory relationships from the three widely used databases, NPInter v3.0 [40], LncRNADisease v2017 [41] and LncRNA2Target v1.2 [42]. Furthermore, we retain experimentally validated lncRNA–mRNA regulatory relationships associated with the four human cancer data sets as ground truth.

We perform survival analysis using the R-package survival [60]. A multivariate Cox model is used to predict the risk score of each tumor sample. Then, all tumor samples in each cancer data set are equally divided into high- and low-risk groups according to their risk scores. Moreover, we calculate the Hazard Ratio between the high- and the low-risk groups and perform the Log-rank test.

To further investigate the underlying biological processes and pathways related to each of the MSLCRN networks, we use the R-package clusterProfiler [61] to conduct functional enrichment analysis on the networks, respectively. The Gene Ontology (GO) [62] biological processes and Kyoto Encyclopedia

Table 3. Degree distributions of global lncRNA–mRNA causal regulatory networks with different cutoffs in GBM, LSCC, OvCa and PrCa

Datasets	Cutoffs	Number of causal regulations	$y=ax^b$	R^2
GBM	0.10	11 847	$y=227.4x^{-0.6893}$	0.4161
	0.15	10 924	$y=249.5x^{-0.7275}$	0.5460
	0.20	9732	$y=274.5x^{-0.767}$	0.6475
	0.25	8461	$y=295.8x^{-0.8074}$	0.6757
	0.30	7176	$y=319.4x^{-0.8319}$	0.6807
	0.35	6041	$y=336.3x^{-0.8703}$	0.7203
	0.40	4997	$y=374.1x^{-0.9348}$	0.7999
	0.45	**4074**	$\mathbf{y=408.2x^{-1.034}}$	**0.8694**
	0.50	3279	$y=419.4x^{-1.18}$	0.9244
	0.55	2583	$y=389.6x^{-1.259}$	0.9463
	0.60	1862	$y=366.6x^{-1.43}$	0.9792
LSCC	0.10	789 172	$y=314.3x^{-0.6071}$	0.4829
	0.15	684 524	$y=347.5x^{-0.6323}$	0.5841
	0.20	569 369	$y=390.5x^{-0.6525}$	0.6578
	0.25	451 346	$y=485.5x^{-0.6928}$	0.7789
	0.30	340 860	$y=634.1x^{-0.7554}$	0.8796
	0.35	244 547	$y=814.7x^{-0.8379}$	0.9504
	0.40	166 593	$y=972.4x^{-0.935}$	0.9848
	0.45	**108 024**	$\mathbf{y=1031x^{-1.018}}$	**0.9933**
	0.50	66 335	$y=942.5x^{-1.068}$	0.9963
	0.55	37 632	$y=780.7x^{-1.089}$	0.9948
	0.60	19 547	$y=656.5x^{-1.169}$	0.9972
OvCa	0.10	333 146	$y=327.2x^{-0.5928}$	0.5042
	0.15	232 794	$y=419.2x^{-0.6262}$	0.6531
	0.20	159 872	$y=639.8x^{-0.7216}$	0.8247
	0.25	112 792	$y=881.6x^{-0.8356}$	0.9120
	0.30	80 808	$y=1008x^{-0.9472}$	0.9551
	0.35	57 099	$y=954.5x^{-1.014}$	0.9744
	0.40	38 517	$y=819.8x^{-1.066}$	0.9748
	0.45	**24 439**	$\mathbf{y=657.5x^{-1.066}}$	**0.9697**
	0.50	14 435	$y=540x^{-1.079}$	0.9551
	0.55	7973	$y=436.8x^{-1.107}$	0.9319
	0.60	4026	$y=328.5x^{-1.107}$	0.9460
PrCa	0.10	1 894 322	$y=308.9x^{-0.6245}$	0.2750
	0.15	1 749 595	$y=358.6x^{-0.6787}$	0.3582
	0.20	1 594 744	$y=401.3x^{-0.7169}$	0.4316
	0.25	1 429 858	$y=427.1x^{-0.732}$	0.4919
	0.30	1 260 968	$y=438.9x^{-0.7244}$	0.5616
	0.35	1 097 654	$y=440.6x^{-0.702}$	0.6470
	0.40	946 439	$y=448.5x^{-0.6816}$	0.7338
	0.45	**812 687**	$\mathbf{y=517.5x^{-0.7005}}$	**0.8206**
	0.50	694 558	$y=667x^{-0.7588}$	0.8823
	0.55	584 834	$y=883.3x^{-0.8469}$	0.9332
	0.60	474 654	$y=1113x^{-0.9684}$	0.9503

Note. The AVCE cutoffs range from 0.10 to 0.60 with a step of 0.05.
The bold values are the degree distributions of global lncRNA-mRNA causal regulatory networks with a compromised AVCE cutoff (0.45) in four human cancers.

of Genes and Genomes (KEGG) [63] pathways with adjusted P-value <0.05 [adjusted by Benjamini-Hochberg (BH) method] are regarded as functional categories for the MSLCRN networks.

We also collect a list of lncRNAs and mRNAs that are associated with GBM, LSCC, OvCa and PrCa to study disease enrichment of each of the MSLCRN networks. The list of disease-associated lncRNAs is obtained from LncRNADisease v2017 [41], Lnc2Cancer v2016 [64] and MNDR v2.0 [65]. The list of disease-associated mRNAs is from DisGeNET v5.0 [66]. To evaluate whether a MSLCRN network is significantly enriched in a specific disease, we use a hyper-geometric distribution test as follows:

$$p = 1 - F(x|B,N,M) = 1 - \sum_{i=0}^{x-1} \frac{\binom{N}{i}\binom{B-N}{M-i}}{\binom{B}{M}} \quad (3)$$

In the formula, B is the number of all genes in the expression data set, N denotes the number of all genes associated with a specific disease in the expression data set, M is the number of genes in a MSLCRN network and x is the number of genes associated with a specific disease in a MSLCRN network. A MSLCRN network is significantly enriched in a specific disease if the P-value < 0.05.

Network analysis, validation and comparison on MSLCRN networks

lncRNAs exhibit dynamic positive gene regulation across cancers

By following the first step of the MSLCRN method, we have identified 23, 38, 45 and 32 lncRNA–mRNA co-expression modules in GBM, LSCC, OvCa and PrCa, respectively. In the second step of the MSLCRN method, we eliminate the noncausal lncRNA–mRNA pairs in lncRNA–mRNA co-expression modules. As a result, we generate 23, 38, 45 and 32 module-specific lncRNA–mRNA causal regulatory networks in GBM, LSCC, OvCa and PrCa, respectively. After merging the module-specific lncRNA–mRNA causal regulatory networks for each data set, we obtain the four global lncRNA–mRNA regulatory networks in GBM, LSCC, OvCa and PrCa, respectively.

To understand the overlap and difference of module-specific genes, module-specific lncRNA–mRNA causal regulatory relationships and module-specific hub lncRNAs in the four human cancers, we generate three set intersection plots using the R-package *UpSetR* [67]. As shown in Figure 3, we find that the majority of module-specific genes (∼57.52%), module-specific lncRNA–mRNA causal regulatory relationships (∼99.02%) and module-specific hub lncRNAs (∼89.22%) tend to be cancer-specific. Only a small portion of module-specific genes (396) and module-specific lncRNA–mRNA causal regulatory relationships (6) are shared by the four cancers. Especially, none of the module-specific hub lncRNAs are common between the four cancers. In addition, the causal effects are positive for 99.56%, 96.72%, 99.93% and 78.63% of the causal regulatory relationships identified in GBM, LSCC, OvCa and PrCa, respectively. These results indicate that lncRNAs are more likely to exhibit dynamic positive gene regulation across cancers. The results are also consistent with the proposition that the positive gene regulation by lncRNAs would be desired in specific situations [68].

Differential network analysis uncovers cancer-specific lncRNA–mRNA causal networks

In this section, we focus on studying cancer-specific lncRNA–mRNA causal networks using differential network analysis. Thus, the GBM-specific, LSCC-specific, OvCa-specific and PrCa-specific lncRNA–mRNA causal networks are identified. As shown in Figure 4A, the distributions of node degrees in these four cancer-specific lncRNA–mRNA causal networks follow power law distributions well, with $R^2 = 0.9774$, 0.9923, 0.9723 and 0.8310, respectively. Thus, these four cancer-specific lncRNA–mRNA causal networks are scale free, indicating that most mRNAs are regulated by a small number of lncRNAs.

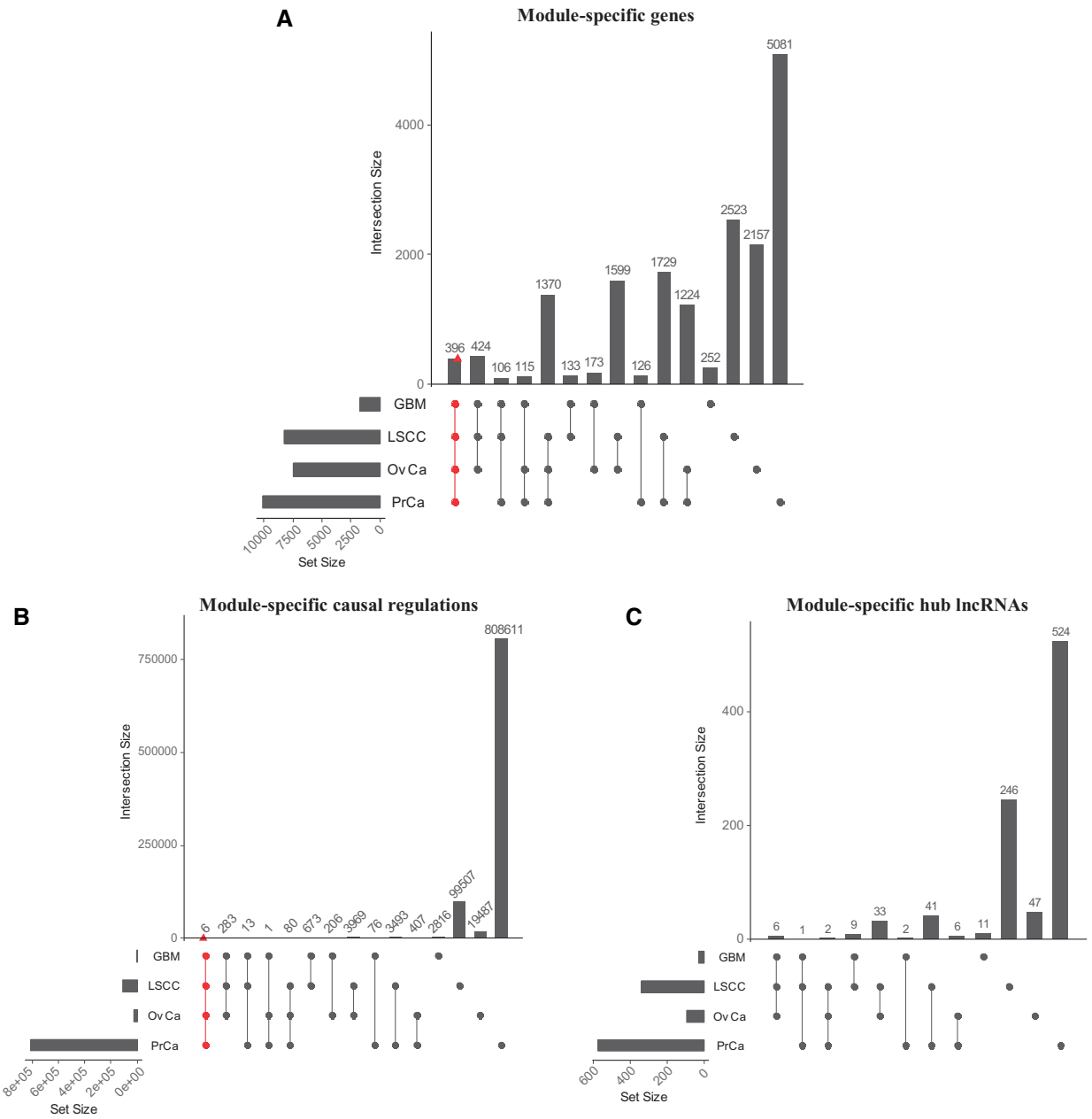
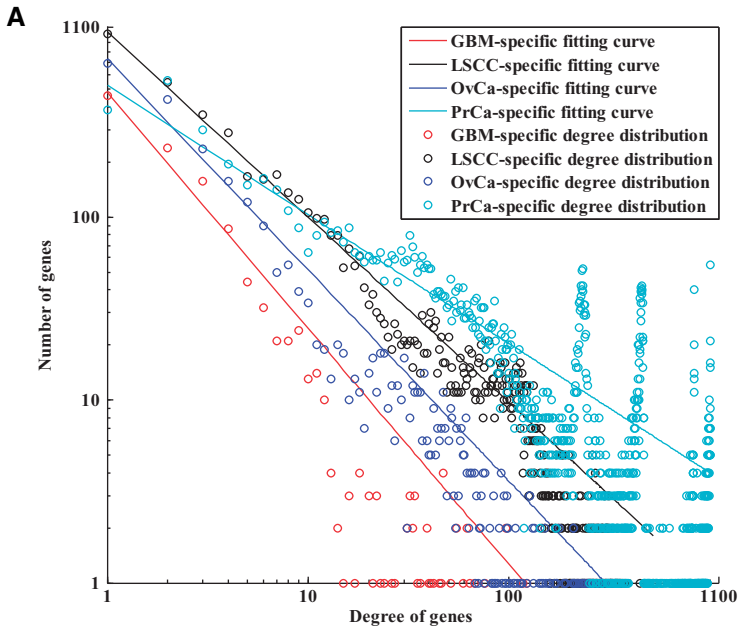


Figure 3. Overlap and difference of module-specific genes, module-specific causal regulations and module-specific hub lncRNAs across GBM, LSCC, OvCa and PrCa. (**A**) Module-specific genes (both lncRNAs and mRNAs) intersection plot. (**B**) Module-specific causal regulations intersection plot. (**C**) Module-specific hub lncRNAs intersection plot. The red lines denote common genes and causal regulations across GBM, LSCC, OvCa and PrCa.

Next, we use four lists of lncRNAs and mRNAs associated with GBM, LSCC, OvCa and PrCa, to discover lncRNA–mRNA causal networks that are associated with the four human cancers. We define that cancer-related lncRNA–mRNA causal regulatory relationships are those in which at least one regulatory party is cancer-related lncRNA or mRNA. As a result, we have extracted GBM-related, LSCC-related, OvCa-related and PrCa-related lncRNA–mRNA causal networks from the four cancer-specific lncRNA–mRNA causal networks (details in Supplementary File S1). To understand the potential biological processes and pathways of the four cancer-related lncRNA–mRNA causal networks, we identify significant GO biological processes and KEGG pathways using functional enrichment analysis. In Figure 4B, several top GO biological processes and KEGG pathways, such as cytokine activity [69], G-protein coupled receptor binding [70], TNF signaling pathway [71], cAMP signaling pathway [72], pathways in cancer, are closely associated with the occurrence and development of cancer. This result suggests that the identified cancer-related lncRNA–mRNA causal networks may be involved in the occurrence and development of human cancer.

Conservative network analysis highlights a core lncRNA–mRNA causal regulatory network across human cancers

Although most of the lncRNA–mRNA causal regulatory relationships are cancer-specific, there are still a number of common



Cancer-specific networks	#Causal regulations	$y=ax^b$	R^2
GBM-specific	2816	$y=481.2x^{-1.292}$	0.9774
LSCC-specific	99507	$y=1034x^{-1.014}$	0.9923
OvCa-specific	19487	$y=736.6x^{-1.156}$	0.9723
PrCa-specific	808611	$y=524.3x^{-0.7055}$	0.8310

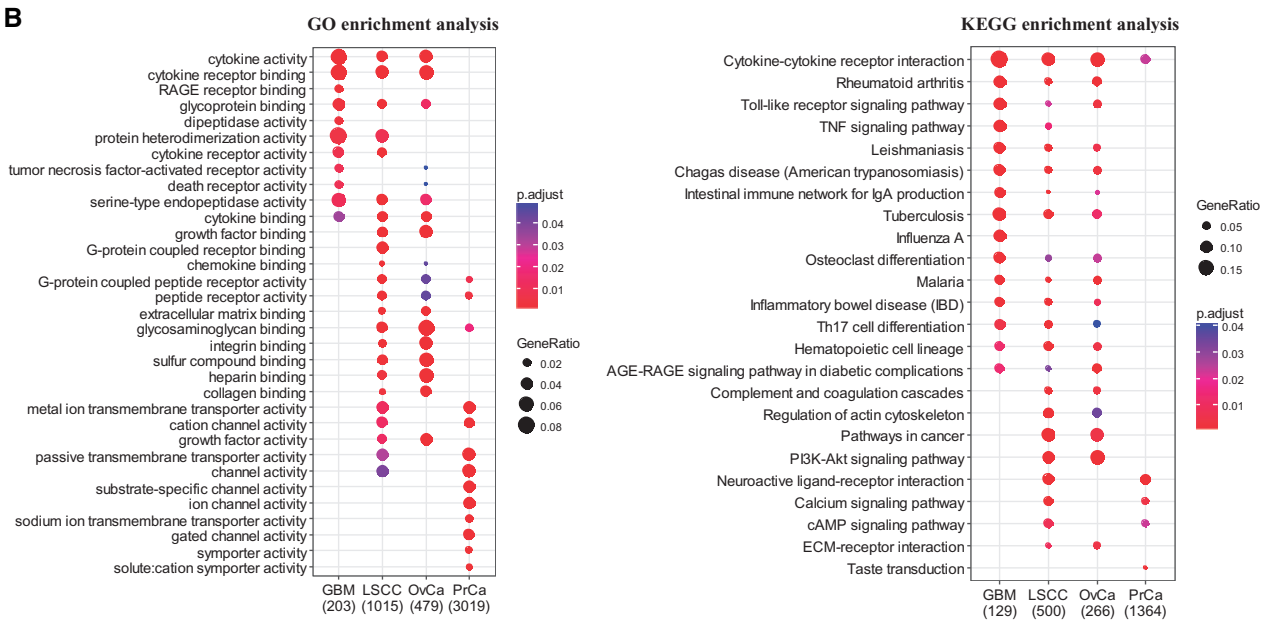


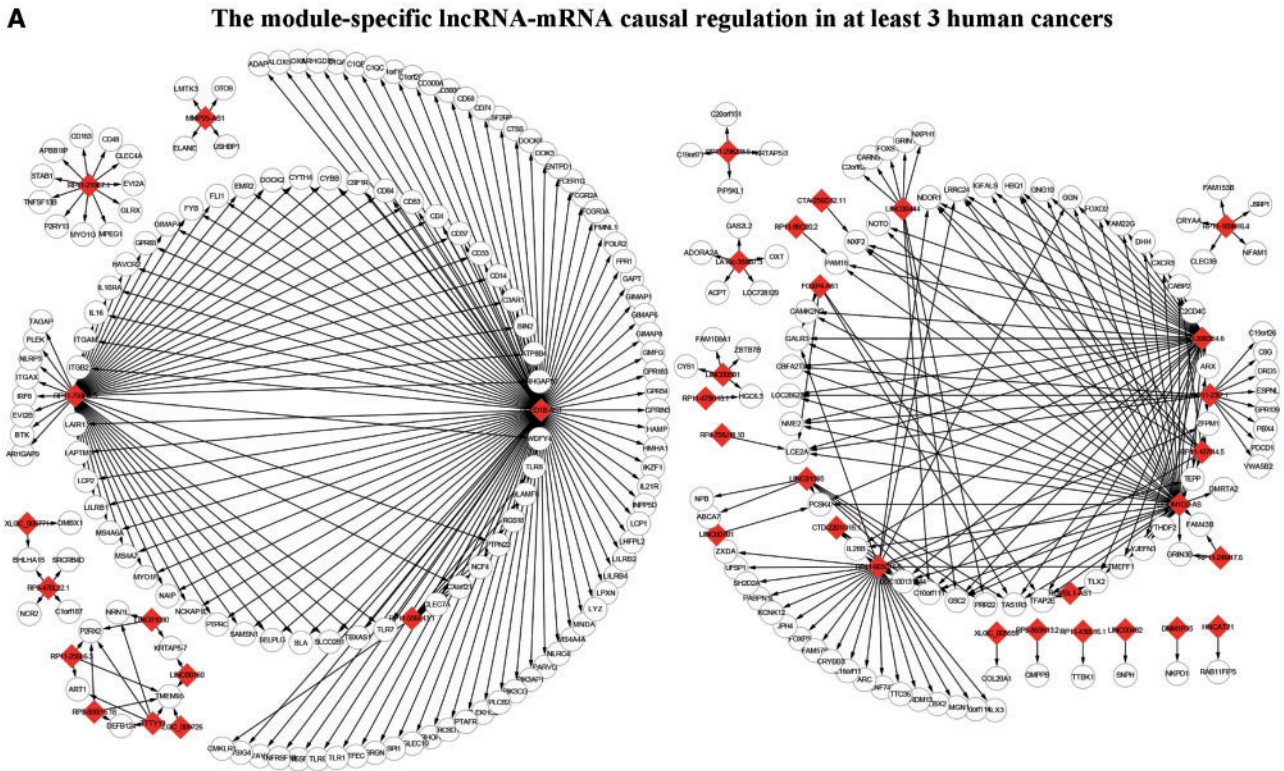
Figure 4. Differential network analysis of global lncRNA–mRNA causal networks across GBM, LSCC, OvCa and PrCa. (**A**) Degree distribution of cancer-specific lncRNA–mRNA causal networks in GBM, LSCC, OvCa and PrCa. (**B**) Functional enrichment analysis of cancer-related lncRNA–mRNA causal networks in GBM, LSCC, OvCa and PrCa.

causal regulatory relationships between the four global networks. To evaluate whether there is a common core of lncRNA–mRNA causal regulatory relationships in the global regulatory networks across human cancers, we concentrate on the conserved lncRNA–mRNA causal regulatory relationships that existed in at least three human cancers.

As shown in Figure 5A, the majority of the conserved lncRNA–mRNA causal regulatory relationships form a closely connected community. This finding indicates that the conserved lncRNA–mRNA causal regulatory network may be a core network across human cancers.

The survival analysis shows that the lncRNAs and mRNAs in the core network can significantly distinguish the metastasis risks between the high- and low-risk groups in GBM, OvCa and PrCa data sets (Figure 5B). This result suggests that the core network may act as a common network biomarker of GBM, OvCa

A



B

Datasets	Chi-square	Log-rank *p*-value	HR	Lower 95% CI of HR	Upper 95% CI of HR	#Cancer genes
GBM	309.06	0	4.44	3.45	5.73	34
LSCC	0.70	0.40	1.30	0.70	2.42	26
OvCa	254.32	0	4.99	3.89	6.39	30
PrCa	12.80	3.46E-04	5.95	1.66	21.36	38

Figure 5. Conservative network analysis of global lncRNA–mRNA causal networks across GBM, LSCC, OvCa and PrCa. (**A**) The core lncRNA–mRNA causal network that occurred in at least three human cancers. The red diamond nodes and white circle nodes denote lncRNAs and mRNAs, respectively. (**B**) Survival analysis of the core lncRNA–mRNA causal network.

and PrCa. In Figure 5B, we also find that the core network contains several cancer genes (34, 26, 30 and 38 cancer genes associated with GBM, LSCC, OvCa and PrCa, respectively).

By conducting GO and KEGG enrichment analysis, we find that the core network is significantly enriched in 399 GO biological processes and 3 KEGG pathways (details in Supplementary File S2). Of the 399 GO biological processes, 2 GO terms, including negative regulation of cell adhesion (GO: 0007162) and cytokine production in immune response (GO: 0002367), are involved in three cancer hallmarks: Tissue Invasion and Metastasis, Tumor Promoting Inflammation and Evading Immune Detection [73]. This observation implies that the core network may control these cancer-related hallmarks.

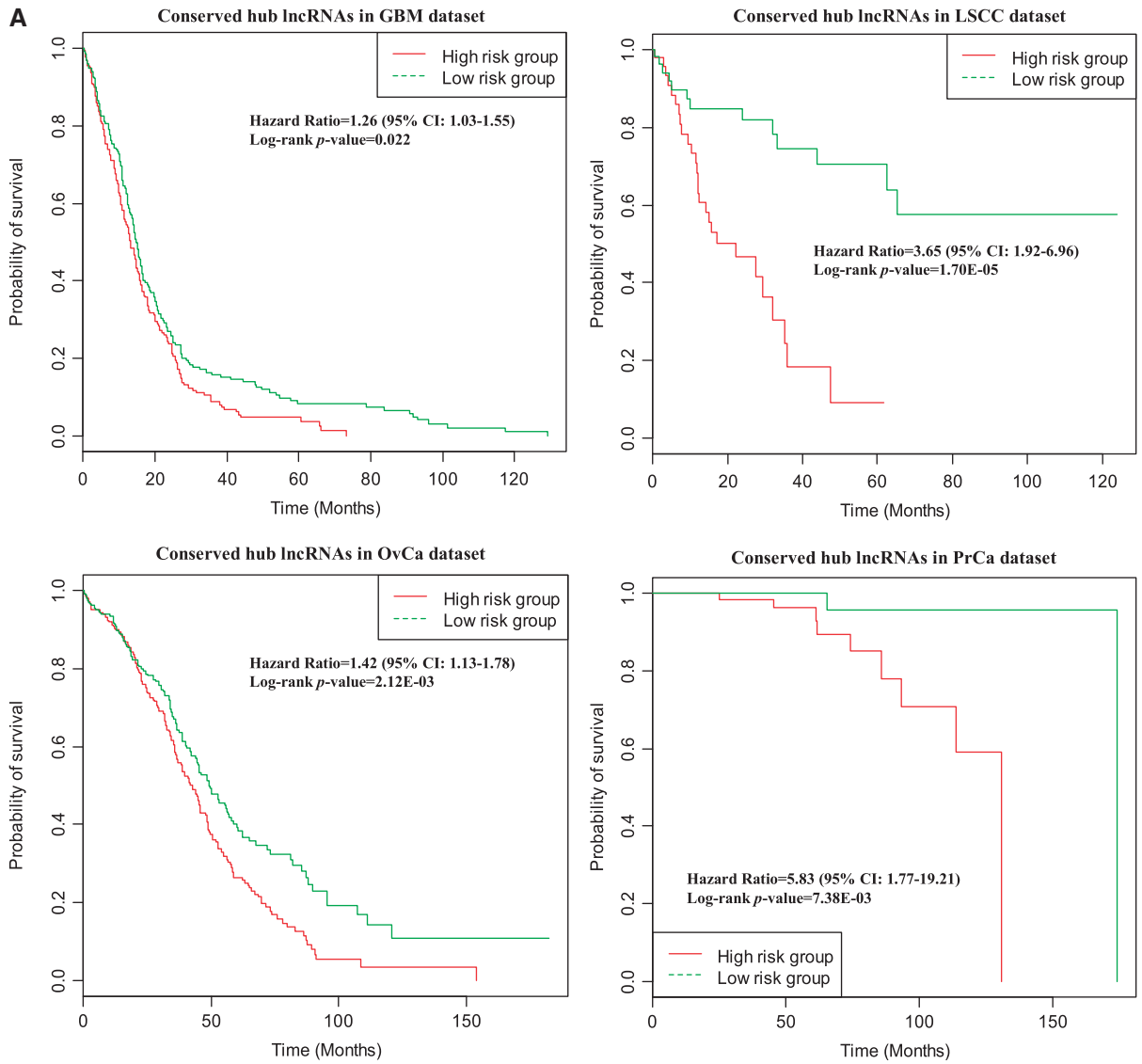
Hub lncRNAs are discriminative and can distinguish metastasis risks of human cancers

We divide the hub lncRNAs into two categories: (1) conserved hub lncRNAs, which exist in at least three human cancers; and (2) cancer-specific hub lncRNAs, which only exist in single human cancer. As a result, we obtain 9 conserved hub lncRNAs and 828 cancer-specific hub lncRNAs (include 11 GBM-specific, 246 LSCC-specific, 47 OvCa-specific and 524 PrCa-specific hub lncRNAs).

To evaluate whether the hub lncRNAs can distinguish metastasis risks of human cancers, we use them to predict metastasis risks for tumor samples in GBM, LSCC, OvCa and PrCa. As shown in Figure 6A, the conserved hub lncRNAs can discriminate the metastasis risks of tumor samples significantly (Log-rank P-value < 0.05) in four human cancers. In Figure 6B, excepting LSCC-specific hub lncRNAs owing to failing to fit a Cox regression model, GBM-specific, OvCa-specific and PrCa-specific hub lncRNAs can discriminate the metastasis risks of tumor samples significantly in GBM, OvCa and PrCa, respectively (Log-rank P-value < 0.05). These results suggest that the hub lncRNAs are discriminative and can act as biomarkers to distinguish between high- and low-risk tumor samples.

Experimentally validated lncRNA–mRNA regulations are mostly bad hits for LncTar

Using a collection of experimentally validated lncRNA–mRNA regulatory relationships (details in Supplementary File S3) as the ground truth, the numbers of experimentally confirmed lncRNA–mRNA causal regulations are 17, 14, 20 and 42 in GBM, LSCC, OvCa and PrCa, respectively (details in Supplementary File S4).



B

Hub lncRNAs	Chi-square	Log-rank *p*-value	HR	Lower 95% CI of HR	Upper 95% CI of HR
GBM-specific	13.27	2.70E-04	1.44	1.17	1.78
LSCC-specific	/	/	/	/	/
OvCa-specific	36.07	1.91E-09	1.97	1.56	2.47
PrCa-specific	21.21	4.11E-06	8.25	1.74	39.07

Figure 6. Survival analysis of hub lncRNAs. (**A**) Conserved hub lncRNAs in GBM, LSCC, OvCa and PrCa datasets. (**B**) Survival analysis of cancer-specific hub lncRNAs.

We further apply a representative sequence-based method called LncTar [11] to the experimentally validated lncRNA–mRNA causal regulatory relationships discovered by MSLCRN. There are two main reasons for choosing LncTar. First, LncTar does not have a limit to input RNA size. Second, LncTar uses a quantitative standard rather than expert knowledge to determine whether lncRNAs interact with mRNAs. Similar to LncTar, we also set -0.1 as normalized binding free energy (ndG) cutoff to determine whether lncRNA–mRNA pairs interact with each other. In other words, the lncRNA–mRNA pairs with $ndG \leq -0.1$ are regarded as lncRNA–mRNA regulatory relationships. Among the experimentally confirmed lncRNA–mRNA causal regulatory relationships that are discovered by MSLCRN, the numbers of successfully predicted lncRNA–mRNA regulations using LncTar are 0, 0, 1 and 1 in GBM, LSCC, OvCa and PrCa, respectively (details in Supplementary File S4). The result indicates that our experimentally confirmed lncRNA–mRNA causal regulations are mostly bad hits for LncTar. Meanwhile, this result also suggests that expression-based and sequence-based methods may be complementary with each other in predicting lncRNA–mRNA regulations.

MSLCRN networks are biologically meaningful

In this section, we conduct GO and KEGG enrichment analysis to check whether the MSLCRN networks are associated with some biological processes and pathways significantly. Enrichment analysis uncovers that 15 of the 23 (∼65.22%) MSLCRN networks in GBM, 29 of the 38 (∼76.32%) MSLCRN networks in LSCC, 30 of the 45 (∼66.67%) MSLCRN networks in OvCa and 20 of the 32 (∼62.50%) MSLCRN networks in PrCa are significantly enriched in at least one GO biological process or KEGG pathway, respectively (details in Supplementary File S5). This result implies that most of the MSLCRN networks in each cancer are functional networks.

We further investigate whether the MSLCRN networks are significantly enriched in GBM, LSCC, OvCa and PrCa diseases, respectively. We discover that 5 of the 23 MSLCRN networks, 7 of the 38 MSLCRN networks, 6 of the 45 MSLCRN networks and 6 of the 32 MSLCRN networks are significantly enriched in GBM, LSCC, OvCa and PrCa diseases, respectively (details in Supplementary File S5). This result indicates that several MSLCRN networks are closely associated with GBM, LSCC, OvCa and PrCa diseases.

Altogether, functional and disease enrichment analysis results show that MSLCRN networks are biologically meaningful.

Comparison with other PC-based network inference methods

Based on a parallel version of the PC algorithm [56], the parallel IDA method in the second step of MSLCRN learns the causal structure from expression data. Owing to the popularity of the PC algorithm in causal structure learning, some other network inference methods, including PCA-CMI [74], PCA-PMI [75] and CMI2NI [76], have also successfully applied it for network inference. Different from the three methods using conditional or partial mutual information to infer lncRNA–mRNA regulations, our method estimates causal effects to identify lncRNA–mRNA regulations. For comparisons, we also use the PCA-CMI, PCA-PMI and CMI2NI methods, to infer module-specific lncRNA–mRNA regulatory relationships. Similar to our method (which uses the parallel IDA method), the strength cutoff of lncRNA–mRNA regulatory relationships in PCA-CMI, PCA-PMI and CMI2NI methods is also set to 0.45.

We evaluate the performance of each method in terms of finding experimentally validated lncRNA–mRNA regulatory relationships, functional MSLCRN networks and disease-associated MSLCRN networks. As shown in Table 4, in terms of the three criteria, MSLCRN performs the best in GBM, LSCC, OvCa and PrCa data sets. This result suggests that MSLCRN is a useful method to infer module-specific lncRNA–mRNA regulatory network in human cancers.

Table 4. Comparison results in terms of experimentally validated lncRNA–mRNA regulatory relationships, functional MSLCRN networks and disease-associated MSLCRN networks

Methods	GBM (a, b, c)	LSCC (a, b, c)	OvCa (a, b, c)	PrCa (a, b, c)
MSLCRN	(17, 15, 5)	(14, 29, 7)	(20, 30, 6)	(42, 20, 6)
PCA-CMI	(2, 13, 0)	(0, 11, 0)	(0, 7, 1)	(0, 20, 2)
PCA-PMI	(2, 15, 1)	(0, 11, 0)	(0, 8, 2)	(1, 18, 1)
CMI2NI	(2, 15, 0)	(0, 11, 0)	(0, 7, 1)	(0, 19, 1)

Note. a = number of experimentally validated lncRNA–mRNA regulatory relationships; b = number of functional MSLCRN networks; c = number of disease-associated MSLCRN networks.

Conclusions and discussion

Notwithstanding lncRNAs do not encode proteins directly, they engage in a wide range of biological processes including cancer developments through their interactions with other biological macromolecules, e.g. DNA, RNA and protein. Therefore, to uncover the functions and regulatory mechanisms of lncRNAs, it is necessary to investigate lncRNA–target regulatory network across different types of biological conditions.

As a biological network, the lncRNA–target regulatory network exhibits a high degree of modularity. Each functional module is responsible for implementing specific biological functions. Moreover, modularity is an important feature of human cancer development and progression. Thus, from a network community point of view, it is necessary to investigate module-specific lncRNA–mRNA regulatory networks.

Until now, several statistical correlation or association measures, e.g. Pearson, Mutual Information and Conditional Mutual Information, have been used to infer gene regulatory networks. However, these methods tend to identify indirect regulatory relationships between genes. The identified gene regulatory networks cannot reflect real 'causal' regulatory relationships. To better understand lncRNA regulatory mechanism, it is vital to investigate how lncRNAs causally influence the expression levels of their target mRNAs.

In this work, the computational methods for inferring lncRNA–mRNA interactions and the publicly available databases of lncRNA–mRNA regulatory relationships are first reviewed. Then, to address the above two issues, we propose a novel computational method, MSLCRN, to study module-specific lncRNA–mRNA causal regulatory networks across GBM, LSCC, OvCa and PrCa diseases. In contrast to other approaches (expression-based and sequence-based methods), MSLCRN has two unique features. First, MSLCRN considers the modularity of lncRNA–mRNA regulatory networks. Instead of studying global regulatory relationships between lncRNAs and mRNAs, we focus on investigating the regulatory behavior of lncRNAs in the modules of interest. Second, considering the restrictions with conducting gene knockout experiments, MSLCRN uses the causal inference method, IDA, to infer causal relationships between lncRNAs and mRNAs based on expression data. The promising results suggest that exploiting modularity of gene regulatory network and causality-based method could provide another effective approach to elucidating lncRNA functions and regulatory mechanisms of human cancers.

Despite the advantages of MSLCRN, there is still room to improve it. First, the WGCNA method only allows clustering genes across all samples from the matched lncRNA and mRNA expression data. In fact, a class of genes may exhibit similar expression patterns across a subset of samples. An alternative solution of this problem is to use a bi-clustering method to identify lncRNA–mRNA co-expression modules. Second, it is still time-consuming to estimate causal effects from large expression data sets. When constructing the module-specific lncRNA–mRNA causal regulatory networks, the running time of parallel IDA is still high on estimating the causal effects of lncRNAs on mRNAs. In future, more efficient parallel IDA method is needed to explore lncRNA–mRNA causal regulatory relationships in large-scale expression data. Third, previous research [38] has shown that the prediction accuracy of lncRNA–mRNA interactions can be improved by integrating both sequence data and

expression data. To improve the accuracy of the predicted lncRNA–mRNA regulatory relationships, it is necessary to develop an ensemble method (fusing sequence-based and expression-based methods) to infer lncRNA–mRNA regulatory network. Finally, recent studies [77] show that lncRNAs can act as competing endogenous RNAs (ceRNAs) or miRNA sponges to attract miRNAs for bindings by competing with mRNAs. Therefore, some predicted lncRNA–mRNA regulatory relationships are lncRNA-related ceRNA–ceRNA interactions. To further improve the prediction of lncRNA–mRNA regulatory relationships, it is necessary to remove the crosstalk relationships between lncRNAs and mRNAs.

Key Points

- Among ncRNAs, lncRNAs are a large and diverse class of RNA molecules, and are thought to be a gold mine of potential oncogenes, anti-oncogenes and new biomarkers.
- lncRNAs exhibit dynamic positive gene regulation across human cancers.
- Hub lncRNAs are discriminative and can distinguish metastasis risks of human cancers.
- There is still a lack of ground truth for validating predicted lncRNA–mRNA regulatory relationships.
- There is still room to develop reliable methods for elucidating lncRNA regulatory mechanisms.

Supplementary Data

Supplementary data are available online at https://academic.oup.com/bib.

Funding

The National Natural Science Foundation of China (No: 61702069); the Applied Basic Research Foundation of Science and Technology of Yunnan Province (No: 2017FB099); the NHMRC Grant (No: 1123042); and the Australian Research Council Discovery Grant (No: DP140103617).

References

1. Pang KC, Frith MC, Mattick JS. Rapid evolution of noncoding RNAs: lack of conservation does not mean lack of function. *Trends Genet* 2006;**22**(1):1–5.
2. Kung JT, Colognori D, Lee JT. Long noncoding RNAs: past, present, and future. *Genetics* 2013;**193**(3):651–69.
3. Schmitt AM, Chang HY. Long noncoding RNAs in cancer pathways. *Cancer Cell* 2016;**29**(4):452–63.
4. Zhang Y, Tao Y, Liao Q. Long noncoding RNA: a crosslink in biological regulatory network. *Brief Bioinform* 2017. doi: 10.1093/bib/bbx042.
5. Yoon JH, Abdelmohsen K, Gorospe M. Posttranscriptional gene regulation by long noncoding RNA. *J Mol Biol* 2013;**425**(19):3723–30.
6. Gerlach W, Giegerich R. GUUGle: a utility for fast exact matching under RNA complementary rules including G-U base pairing. *Bioinformatics* 2006;**22**(6):762–4.
7. Mückstein U, Tafer H, Hackermüller J, *et al*. Thermodynamics of RNA-RNA binding. *Bioinformatics* 2006;**22**(10):1177–82.
8. Tafer H, Hofacker IL. RNAplex: a fast tool for RNA-RNA interaction search. *Bioinformatics* 2008;**24**(22):2657–63.
9. Busch A, Richter AS, Backofen R. IntaRNA: efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions. *Bioinformatics* 2008;**24**(24):2849–56.
10. Kato Y, Sato K, Hamada M, *et al*. RactIP: fast and accurate prediction of RNA-RNA interaction using integer programming. *Bioinformatics* 2010;**26**(18):i460–6.
11. Li J, Ma W, Zeng P, *et al*. LncTar: a tool for predicting the RNA targets of long noncoding RNAs. *Brief Bioinform* 2015;**16**(5):806–12.
12. Fukunaga T, Hamada M. RIblast: an ultrafast RNA-RNA interaction prediction system based on a seed-and-extension approach. *Bioinformatics* 2017;**33**(17):2666–74.
13. Derrien T, Johnson R, Bussotti G, *et al*. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res* 2012;**22**(9):1775–89.
14. Gloss BS, Dinger ME. The specificity of long noncoding RNA expression. *Biochim Biophys Acta* 2016;**1859**(1):16–22.
15. Munshi A, Mohan V, Ahuja YR. Non-coding RNAs: a dynamic and complex network of gene regulation. *J Pharmacogenomics Pharmacoproteomics* 2016;**7**:156.
16. Liao Q, Liu C, Yuan X, *et al*. Large-scale prediction of long noncoding RNA functions in a coding-non-coding gene co-expression network. *Nucleic Acids Res* 2011;**39**(9):3864–78.
17. Guo Q, Cheng Y, Liang T, *et al*. Comprehensive analysis of lncRNA-mRNA co-expression patterns identifies immune-associated lncRNA biomarkers in ovarian cancer malignant progression. *Sci Rep* 2015;**5**(1):17683.
18. Du Y, Xia W, Zhang J, *et al*. Comprehensive analysis of long noncoding RNA-mRNA co-expression patterns in thyroid cancer. *Mol Biosyst* 2017;**13**(10):2107–15.
19. Wu W, Wagner EK, Hao Y, *et al*. Tissue-specific co-expression of long non-coding and coding RNAs associated with breast Cancer. *Sci Rep* 2016;**6**:32731.
20. Barabási AL, Oltvai ZN. Network biology: understanding the cell's functional organization. *Nat Rev Genet* 2004;**5**(2):101–13.
21. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 2008;**9**:559.
22. Maathuis HM, Kalisch M, Bühlmann P. Estimating high-dimensional intervention effects from observational data. *Ann Stat* 2009;**37**(6A):3133–64.
23. Maathuis HM, Colombo D, Kalisch M, *et al*. Predicting causal effects in large-scale systems from observational data. *Nat Methods* 2010;**7**(4):247–8.
24. Le T, Hoang T, Li J, *et al*. A fast PC algorithm for high dimensional causal discovery with multi-core PCs. *IEEE/ACM Trans Comput Biol Bioinform* 2016. doi: 10.1109/TCBB.2016.2591526.
25. Du Z, Fei T, Verhaak RG, *et al*. Integrative genomic analyses reveal clinically relevant long noncoding RNAs in human cancer. *Nat Struct Mol Biol* 2013;**20**(7):908–13.
26. Bernhart SH, Tafer H, Mückstein U, *et al*. Partition function and base pairing probabilities of RNA heterodimers. *Algorithms Mol Biol* 2006;**1**(1):3.
27. Alkan C, Karakoç E, Nadeau JH, *et al*. RNA-RNA interaction prediction and antisense RNA target search. *J Comput Biol* 2006;**13**(2):267–82.
28. Seemann SE, Richter AS, Gesell T, *et al*. PETcofold: predicting conserved interactions and structures of two multiple alignments of RNA sequences. *Bioinformatics* 2011;**27**(2):211–19.
29. Wenzel A, Akbasli E, Gorodkin J. RIsearch: fast RNA-RNA interaction search using a simplified nearest-neighbor energy model. *Bioinformatics* 2012;**28**(21):2738–46.

30. Alkan F, Wenzel A, Palasca O, *et al*. RIsearch2: suffix array-based large-scale prediction of RNA-RNA interactions and siRNA off-targets. *Nucleic Acids Res* 2017;**45**:e60.
31. Hu R, Sun X. lncRNATargets: a platform for lncRNA target prediction based on nucleic acid thermodynamics. *J Bioinform Comput Biol* 2016;**14**(4):1650016.
32. Terai G, Iwakiri J, Kameda T, *et al*. Comprehensive prediction of lncRNA-RNA interactions in human transcriptome. *BMC Genomics* 2016;**17(Suppl 1)**:12.
33. Liu J, Wu S, Li M, *et al*. LncRNA expression profiles reveal the co-expression network in human colorectal carcinoma. *Int J Clin Exp Pathol* 2016;**9**:1885–1892.
34. Huang S, Feng C, Chen L, *et al*. Identification of potential key long non-coding RNAs and target genes associated with pneumonia using long non-coding RNA sequencing (lncRNA-Seq): a preliminary study. *Med Sci Monit* 2016;**22**:3394–408.
35. Li J, Xu Y, Xu J, *et al*. Dynamic co-expression network analysis of lncRNAs and mRNAs associated with venous congestion. *Mol Med Rep* 2016;**14**(3):2045–51.
36. Fu M, Huang G, Zhang Z, *et al*. Expression profile of long non-coding RNAs in cartilage from knee osteoarthritis patients. *Osteoarthritis Cartilage* 2015;**23**(3):423–32.
37. Zhang F, Gao C, Ma XF, *et al*. Expression profile of long non-coding RNAs in peripheral blood mononuclear cells from multiple sclerosis patients. *CNS Neurosci Ther* 2016;**22**(4): 298–305.
38. Iwakiri J, Terai G, Hamada M. Computational prediction of lncRNA-mRNA interactionsby integrating tissue specificity in human transcriptome. *Biol Direct* 2017;**12**(1):15.
39. Lv L, Wei M, Lin P, *et al*. Integrated mRNA and lncRNA expression profiling for exploring metastatic biomarkers of human intrahepatic cholangiocarcinoma. *Am J Cancer Res* 2017;**7**: 688–99.
40. Hao Y, Wu W, Li H, *et al*. NPInter v3.0: an upgraded database of noncoding RNA-associated interactions. *Database* 2016; **2016**:baw057.
41. Chen G, Wang Z, Wang D, *et al*. LncRNADisease: a database for long-non-coding RNA-associated diseases. *Nucleic Acids Res* 2013;**41**:D983–6.
42. Jiang Q, Wang J, Wu X, *et al*. LncRNA2Target: a database for differentially expressed genes after lncRNA knockdown or overexpression. *Nucleic Acids Res* 2015;**43**:D193–6.
43. Zhou Z, Shen Y, Khan MR, *et al*. LncReg: a reference resource for lncRNA-associated regulatory networks. *Database* 2015; **2015**:bav083.
44. Denisenko E, Ho D, Tamgue O, *et al*. IRNdb: the database of immunologically relevant non-coding RNAs. *Database* 2016; **2016**:baw138.
45. Liu CJ, Gao C, Ma Z, *et al*. lncRInter: a database of experimentally validated long non-coding RNA interaction. *J Genet Genomics* 2017;**44**(5):265–8.
46. Li JH, Liu S, Zhou H, *et al*. starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res* 2014; **42**(D1):D92–7.
47. Liu Y, Zhao M. lnCaNet: pan-cancer co-expression network for human lncRNA and cancer genes. *Bioinformatics* 2016; **32**(10):1595–7.
48. Zhou QZ, Zhang B, Yu QY, *et al*. BmncRNAdb: a comprehensive database of non-coding RNAs in the silkworm, Bombyx mori. *BMC Bioinformatics* 2016;**17**(1):370.
49. Park C, Yu N, Choi I, *et al*. lncRNAtor: a comprehensive resource for functional investigation of long non-coding RNAs. *Bioinformatics* 2014;**30**(17):2480–5.
50. Bhartiya D, Pal K, Ghosh S, *et al*. lncRNome: a comprehensive knowledgebase of human long noncoding RNAs. *Database* 2013;**2013**:bat034.
51. Zhao Z, Bai J, Wu A, *et al*. Co-LncRNA: investigating the lncRNA combinatorial effects in GO annotations and KEGG pathways based on human RNA-Seq data. *Database* 2015; **2015**:bav082.
52. Jiang Q, Ma R, Wang J, *et al*. LncRNA2Function: a comprehensive resource for functional investigation of human lncRNAs based on RNA-seq data. *BMC Genomics* 2015; **16**(Suppl 3):S2.
53. Chan WL, Huang HD, Chang JG. lncRNAMap: a map of putative regulatory functions in the long non-coding transcriptome. *Comput Biol Chem* 2014;**50**:41–9.
54. Langfelder P, Horvath S. Fast R functions for robust correlations and hierarchical clustering. *J Stat Softw* 2012;**46**:1–17.
55. Judea P. *Causality: Models, Reasoning, and Inference*. New York, NY: Cambridge University Press, 2000.
56. Spirtes P, Glymour C, Scheines R. *Causation, Prediction, and Search*, 2nd edn. Cambridge: MIT Press, 2000.
57. Le T, Hoang T, Li J, *et al*. ParallelPC: an R package for efficient constraint based causal exploration. *arXiv prepring* 2015. arXiv:1510.03042v1
58. Hahn MW, Kern AD. Comparative genomics of centrality and essentiality in three eukaryotic protein-interaction networks. *Mol Biol Evol* 2005;**22**(4):803–6.
59. Song J, Singh M, Roth FP. From hub proteins to hub modules: the relationship between essentiality and centrality in the yeast interactome at different scales of organization. *PLoS Comput Biol* 2013;**9**(2):e1002910.
60. Therneau TM, Grambsch PM, *Modeling Survival Data: Extending the Cox Model*. New York: Springer Press, 2000.
61. Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 2012;**16**(5):284–7.
62. Ashburner M, Ball CA, Blake JA, *et al*. Gene ontology: tool for the unification of biology. *Nat Genet* 2000;**25**(1):25–9.
63. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* 2000;**28**(1):27–30.
64. Ning S, Zhang J, Wang P, *et al*. Lnc2Cancer: a manually curated database of experimentally supported lncRNAs associated with various human cancers. *Nucleic Acids Res* 2016;**44**(D1): D980–5.
65. Wang Y, Chen L, Chen B, *et al*. Mammalian ncRNA-disease repository: a global view of ncRNA-mediated disease network. *Cell Death Dis* 2013;**4**:e765.
66. Piñero J, Bravo À, Queralt-Rosinach N, *et al*. DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res* 2017; **45**(D1):D833–9.
67. Conway JR, Lex A, Gehlenborg N. UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics* 2017;**33**(18):2938–40.
68. Wahlestedt C. Targeting long non-coding RNA to therapeutically upregulate gene expression. *Nat Rev Drug Discov* 2013; **12**(6):433–46.
69. Mantovani G, Macciò A, Lai P, *et al*. Cytokine activity in cancer-related anorexia/cachexia: role of megestrol acetate and medroxyprogesterone acetate. *Semin Oncol* 1998;**25**:45–52.
70. Dorsam RT, Gutkind JS. G-protein-coupled receptors and cancer. *Nat Rev Cancer* 2007;**7**(2):79–94.
71. Wang X, Lin Y. Tumor necrosis factor and cancer, buddies or foes? *Acta Pharmacol Sin* 2008;**29**(11):1275–88.

72. Fajardo AM, Piazza GA, Tinsley HN. The role of cyclic nucleotide signaling pathways in cancer: targets for prevention and treatment. *Cancers* 2014;**6**(1):436–58.
73. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell* 2011;**144**(5):646–74.
74. Zhang X, Zhao XM, He K, *et al*. Inferring gene regulatory networks from gene expression data by path consistency algorithm based on conditional mutual information. *Bioinformatics* 2012;**28**(1):98–104.
75. Zhao J, Zhou Y, Zhang X, *et al*. Part mutual information for quantifying direct associations in networks. *Proc Natl Acad Sci USA* 2016;**113**(18):5130–5.
76. Zhang X, Zhao J, Hao JK, *et al*. Conditional mutual inclusive information enables accurate quantification of associations in gene regulatory networks. *Nucleic Acids Re* 2015;**43**(5):e31.
77. Le TD, Zhang J, Liu L, *et al*. Computational methods for identifying miRNA sponge interactions. *Brief Bioinform* 2017;**18**(4):577–90.